# The Discrimination Ability of Some Topological and Information Distance Indices for Graphs of Unbranched Hexagonal Systems

Elena V. Konstantinova†

Institute of Mathematics, Siberian Branch of Russian Academy of Science, Novosibirsk 630090, Russia

Received July 7, 1995⊗

Some topological and information distance indices are considered. The discrimination ability of distance indices for graphs of unbranched hexagonal systems is examined. Two among them are little known information indices. One of them has the high discriminating ability for considered graphs.

## 1. INTRODUCTION

Topological indices of molecular structures are widely used in computational and mathematical chemistry.[1−9] The construction and investigation of topological indices which could be used to describe molecular structures is one of the important directions of mathematical chemistry. Topological indices are designed basically by transforming a molecular graph into a number. The first mathematical index reflecting the topological structure of a molecular graph was proposed in 1947 by Wiener.[10] This index is based on distances within a graph. There are a number of distance indices available in the literature.[9−11,12] In 1975 Randić[13] presented the molecular connectivity index. This index has been developed into a series of connectivity indices,[14] and later Randić proposed identification numbers.[15] The investigation of new topological indices continues,[16−18] and some attempt to the classification and generalization of them has been proposed.[19−21]

The present work is devoted to compare distance indices. Interest in distance indices as well as in other indices is stimulated by their use in quantitative structure−property and structure−activity relationships.[4,8,12] Well-known topological distance indices will be discussed here. Also we will consider the information distance indices.[1] Two among them are little known in the literature topological indices.[22] However we not consider the connectivity indices here. The molecular connectivity index was investigated by the author in ref 22 for graphs embedded into a hexagonal lattice. Now we will consider graphs of unbranched hexagonal systems which are embedded into a hexagonal lattice and are not embedded.

We should examine the discrimination ability of topological and information distance indices. One of the basic characteristics of topological index is its sensitivity in the process of molecular structure classification.[23,24] The sensitivity of a topological index $I$ is the measure of its ability to distinguish nonisomorphic graphs. The theoretical evaluation of the sensitivity $I$ on the set of all graphs is difficult. Therefore the evaluation $S$ of sensitivity $I$ on a smaller fixed set $M$ of nonisomorphic graphs is achieved by the formula

$$S = (N - N_I)/N \qquad (1)$$

where $N = |M|$ and $N_I$ is a number of graphs that cannot be distinguished by topological index $I$ on the set $M$.

† E-mail: e- konsta@math.nsk.

⊗ Abstract published in *Advance ACS Abstracts,* January 1, 1996.

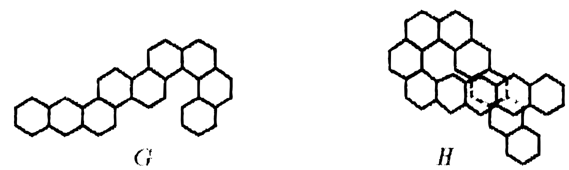


**Figure 1.** The graphs of the class $UB(h)$.

There is the assumption in the problem of compound property prediction that molecules with similar structures (or values of the topological index as a measure of similarity) have similar properties. Thus the study of topological index sensitivity is important for the investigation of topological indices.

## 2. GRAPHS OF UNBRANCHED HEXAGONAL SYSTEMS

We introduce the definition of graphs from ref 25.

Let $G$ be a finite connected graph without loops and multiple edges; $V(G)$ is the set of vertices of graph $G$ with cardinality $p = |V(G)|$, and $E(G)$ is the set of edges of graph $G$, $q = |E(G)|$. Define a class of graphs where all internal faces on a plane are hexagonal, and two arbitrary faces either have only a common edge (i.e., they are adjacent) or have no common vertices. Each face is adjacent to no more than two other faces. Hexagonal faces together with their boundary are called the rings of the graph. By placing each hexagonal ring in correspondence with a new vertex and then joining them (if the corresponding rings are adjacent), we obtain the characteristic (or dualist) graph of the initial one. A set of graphs consisting of $h$ rings for which their characteristic graph is isomorphic to a simple path is denoted $UB(h)$. Graphs of this class model molecular structures of unbranched cata-condensed benzenoid hydrocarbons.[26,27] Graphs $G$ and $H$ (see Figure 1) belong to the class $UB(h)$. The set of graphs $UB(h)$ contains the graphs which are embedded (see graph $G$ in Figure 1) and are not embedded (graph $H$ in Figure 1) into a hexagonal lattice.

## 3. TOPOLOGICAL AND INFORMATION DISTANCE INDICES OF GRAPHS

We consider eight topological and information indices based on distances within a graph. The distance $d(u,v)$ between vertices $u,v$ in graph $G$ is the length of a simple path which joins the vertices $u$ and $v$ in the graph $G$ and contains the minimal number of edges.

0095-2338/96/1636-0054$12.00/0 

**Table 1.** Indices Sensitivities for Graphs $UB(h)$, $h = 1...10^a$

| index | $h = 3$, $n = 2$ | $h = 4$, $n = 4$ | $h = 5$, $n = 10$ | $h = 6$, $n = 25$ | $h = 7$, $n = 70$ | $h = 8$, $n = 196$ | $h = 9$, $n = 574$ | $h = 10$, $n = 1681$ |
|---|---|---|---|---|---|---|---|---|
| $H_D$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $J$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.999 |
| $I_D^W$ | 1 | 1 | 1 | 1 | 1 | 1 | 0.990 | 0.992 |
| $H_\lambda$ | 1 | 1 | 1 | 1 | 1 | 1 | 0.990 | 0.989 |
| $\Delta D$ | 1 | 1 | 1 | 1 | 0.971 | 0.867 | 0.772 | 0.550 |
| $GDI$ | 1 | 1 | 0.8 | 0.68 | 0.571 | 0.474 | 0.392 | 0.323 |
| $D^2$ | 1 | 1 | 0.8 | 0.68 | 0.571 | 0.429 | 0.308 | 0.220 |
| $W$ | 1 | 1 | 0.8 | 0.48 | 0.200 | 0.112 | 0.028 | 0.014 |

[a] In the table $n$ is the number of graphs in the $h$-class.

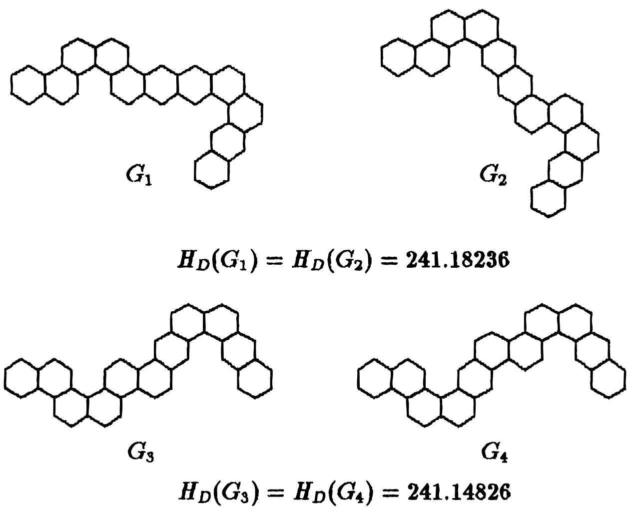


**Figure 2.** The graphs with degenerate values of the $H_D$ index; $h = 11$.

The formulas of well-known topological indices are listed below.

1. The Wiener index[10,28] of the graph $G$ is determined as

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

2. The average distance sum connectivity was introduced by Balaban[29] and is defined as

$$J(G) = \frac{2}{\mu + 1} \sum_{edges} (d(u)d(v))^{-1/2} \tag{3}$$

where $\mu = q - p + 1$ is the cyclomatic number of $G$ (the number of rings in the graph) and $d(v) = \sum_{i=1}^{p} d(i,v)$ is the distance (or distance sum) of a vertex.[30]

3. The mean square distance index was defined by Balaban[29] from the formula

$$D^2(G) = \left( \sum_{i=1}^{\max} g_i i^2 / \sum_{i=1}^{\max} g_i \right)^{1/2} \tag{4}$$

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

4. The graph distance index[11] was defined as

$$GDI(G) = \sum_{i=1}^{\max} (g_i)^2 \tag{5}$$

where $g_i$ has its above significance.

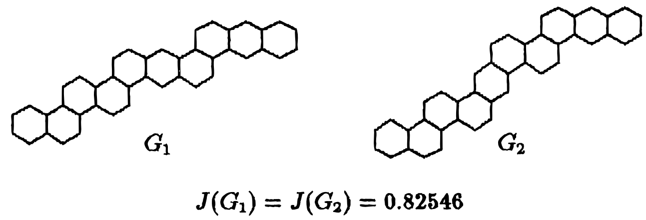


**Figure 3.** The graphs with degenerate values of the $J$ index; $h = 10$.

5. The mean distance deviation[31] is defined as follows

$$\Delta D(G) = \frac{1}{p} \sum_{v \epsilon V(G)} \left| d(v) - \frac{2W(G)}{p} \right| \tag{6}$$

where symbols have their above significance.

We also consider three information indices. Two of them are little known.

The following well-known principle[32] is generally used for the construction of information indices.

Let $X$ be a set consisting of $n$ elements. Let us assume that by some equivalence criterion the elements are divided into $N$ equivalence classes $X_i$, where $n = \sum_{i=1}^{N} n_i$ and $n_i$ is the number of elements in subset $X_i$. Then $p_i = n_i/n$ is the probability for a single element to belong to the $i$th subset; to estimate quantitatively the information that corresponds to one element of the set, one can use the distribution entropy for the set elements defined by the following formula of Shannon[33]

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i$$

Information indices of molecular graphs were constructed for various matrices of graphs and also for some topological indices.[11,32] The formulas of information indices are listed below.

6. The information Wiener index[34] is the information analog of the Wiener index and it takes the form

$$I_D^W(G) = \frac{1}{2} \sum_{i=1}^{\max} \frac{g_i i}{W(G)} \log_2 \frac{g_i i}{W(G)} \tag{7}$$

7. The information distance index of graph vertices was introduced in ref 22 and is defined on the basis of the distance matrix $D = ||d_{ij}||$, $i$, $j = 1...p$, where $d_{ij}$ is the distance between vertices $i$ and $j$ in graph $G$. Let us define the

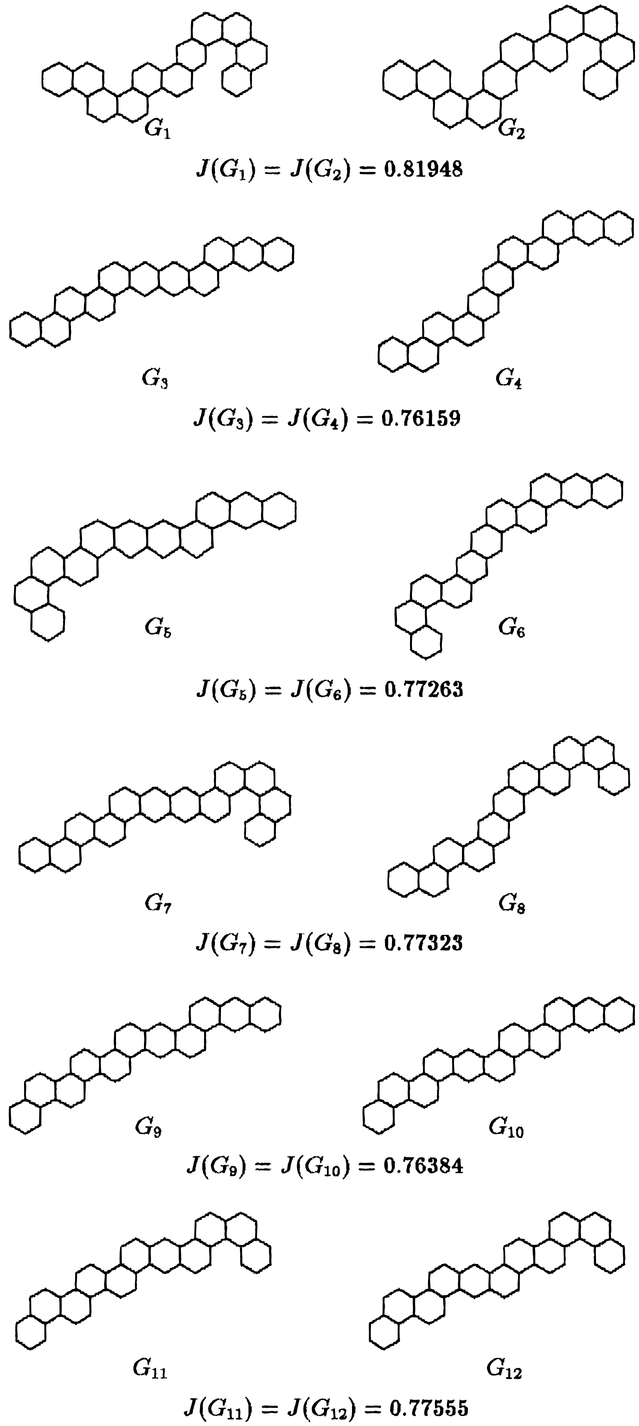


**Figure 4.** The graphs with degenerate values of the $J$ index; $h =$ 11.

information distance index of vertex $i$ as follows

$$H_D(i) = -\sum_{j=1}^{p} \frac{d_{ij}}{d(i)} \log_2 \frac{d_{ij}}{d(i)}$$

where $d(i) = \sum_{j=1}^{p} d_{ij}$ and $p_{ij} = d_{ij}/d(i)$ is the probability for an arbitrarily chosen vertex to be at a distance $d_{ij}$ from the vertex $i$. Then the information distance index of graph vertices take the form

$$H_D = \sum_{i=1}^{p} H_D(i) \tag{8}$$

8. The information layer index of graph vertices was also introduced in ref 22 and it was constructed for the layer matrix. The layer matrix of a graph $G$[31] is a matrix $\lambda(G) = ||\lambda_{ij}||$, $i = 1...p$, $j = 1...d(G)$, where $\lambda_{ij}$ is equal to the number of vertices located at a distance $j$ from vertex $i$, $d(G)$ is the diameter of graph $G$. Let the set $X$ of the vertex $i$ be defined by the nonzero elements of corresponding row of the layer matrix. Then $|X| = p$, if one takes into account the matrix properties. We define the information layer index of vertex $i$ as follows

$$H_\lambda(i) = -\sum_{j=0}^{e(i)} \frac{\lambda_{ij}}{p} \log_2 \frac{\lambda_{ij}}{p}$$

where $p_{ij} = \lambda_{ij}/p$ is the probability for an arbitrarily chosen vertex to find itself in the $j$th layer of the vertex $i$, and $e(i) = \max_{v\epsilon V(G)} d(i,v)$ is the vertex eccentricity. Then the information layer index of graph vertices takes the form

$$H_\lambda = \sum_{i=1}^{p} H_\lambda(i) \tag{9}$$

The sensitivity of the considered topological and information indices for graphs of unbranched hexagonal systems will be investigated. The results of the analysis are presented below. To obtain the graphs, we used the algorithm of fast generation of graphs of unbranched hexagonal systems.[35]

## 4. DISCUSSION OF RESULTS

The graphs of the unbranched hexagonal systems with $h$ rings, $h = 1...11$, were obtained. The sensitivities of indices considered above were calculated for each class of these graphs in accordance with formula (1). The calculation accuracy for all the indices is $10^{-9}$. The results are given in Table 1. The data in the table show that the information distance index of graph vertices $H_D$ is more sensitive (discriminating) than the other indices for all classes of considered graphs. The index $H_D$ degenerates on the graphs of unbranched hexagonal systems with $h = 11$. The number $n$ of graphs in this class is $n = 5002$. There are two pairs of graphs for which the $H_D$ index degenerates (see Figure 2). The $J(G)$ index degenerates on the graphs with $h = 10$ presented in the Figure 3 and degenerates on the six pairs of graphs with $h = 11$ presented in the Figure 4.

Balaban in ref 29 indicated that "the $J$ index is the least degenerate single topological index proposed till now". The above results show that the $H_D$ index has now the lowest degeneracy of all indices considered in this article for graphs with $h$ rings, $1 \leq h \leq 11$.

## ACKNOWLEDGMENT

This work was supported by the Russian Fund of Fundamental Investigations through Grant No. 93-03-18657. The author is grateful to Andrey A. Dobrynin for the program of the generation of graphs of unbranched hexagonal systems.[35] Discussions with V. A. Skorobogatov, Yu. S. Nekrasov, and E. E. Tepfer are acknowledged. The author is grateful to the referees for their constructive comments.

## REFERENCES AND NOTES

(1) Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structure*; Research Studies Press: Chichester, 1983.

(2) Balaban, A. T. Applications of Graph Theory in Chemistry. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 334−343.

(3) Gutman, I.; Polansky, O. *Mathematical Concepts in Organic Chemistry*; Springer: Berlin, 1986.
(4) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Research Studies Press: Letchworth, Herts, 1986.
(5) Hansen, P. J.; Jurs, P. C. Chemical Application of Graph Theory. Part 1. Fundamentals and Topological Indices. *J. Chem. Educ.* **1988**, *65*, 574−580.
(6) Randić, M.; Hansen, P. J.; Jurs, P. C. Search for Useful Graph Theoretical Invariants of Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 60−68.
(7) Rouvray, D. H. The Limits of Applicability of Topological Indices. *J. Mol. Struct. (Theochem)*, **1989**, *185*, 187−201.
(8) Randić, M. Resolution of Ambiguities in Structure-Property Studes by Use of Ortogonal Descriptors. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311−320.
(9) Trinajstić, N. *Chemical Graph Theory*, 2nd revised ed; CRC Press: Boca Raton, FL, 1992.
(10) Wiener, H. J. Structural Determination of Parrafin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.
(11) Rouvray, D. H. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Ed.; Elsevier: Amsterdam, 1983; pp 159−177.
(12) Mihalic, Z.; Trinajstić, N. Graph Theoretical Approach to Structure−Property Relationships. *J. Chem. Educ.* **1992**, *69*, 701−712.
(13) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *69*, 6609−6615.
(14) Kier, L. B.; Hall, L. H. Derivation and Significance of Valence Molecular Connectivity. *J. Pharm. Sci.* **1981**, *70*, 583−589.
(15) Randić, M. On Molecular Identification Numbers. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164−175.
(16) Diudea, M. V.; Minailiuc, O.; Balaban, A. T. Molecular Topology. IV. Regressive Vertex Degree (New Graph Invariants) and Derived Topological Indices. *J. Comput. Chem.* **1991**, *12*, 527−535.
(17) Yao, Y. Y.; Xu, L.; Yang, Y. Q.; Yaun, X. S. Study on Structure−Activity Relationships of Organic Compounds. 3. New Topological Indices and Their Application. *J. Chem. Inf. Somput. Sci.* **1993**, *33*, 478−483.
(18) Estrada, E. Edge Adjacency Relationships in Molecular Graphs Containing Heteroatoms: A New Topological Index Related to Molar Volume. *J. Chem. Inf. Copmut. Sci.* **1995**, *35*, 701−707.
(19) Randić, M. Generalized Molecular Descriptors. *J. Math. Chem.* **1991**, *7*, 155−168.
(20) Mihalic, Z.; Nikolic, S.; Trinajstić, N. Comparative Study of Molecular Descriptors from the Distance Matrix. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 28−37.
(21) Konstantinova, E. V.; Skorobogatov, V. A. Structural and Numerical Invariants of Ordinary and Molecular Graphs. *Vychisl. Systemy* **1991**, *140*, 87−129 (in Russian).
(22) Konstantinova, E. V.; Paleev, A. A. Sensitivity of Topological Indices of Polycyclic Graphs. *Vychisl. Sistemy* **1990**, *136*, 38−48 (in Russian).
(23) Bonchev, D.; Mekenyan, O.; Trinajstić, N. Isomer Discrimination by Topological Information Approach. *J. Comput. Chem.* **1981**, *2*, 127−148.
(24) Razinger, M.; Chrétien, J. R.; Dubois, J. K. Structural Selectivity of Topological Indices in Alkane Series. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 23−27.
(25) Dobrynin, A. A. Graphs and Unbranched Hexagonal Systems with Equal Values of the Wiener Index and Different Numbers of Rings. *J. Math. Chem.* **1992**, *9*, 239−252.
(26) Balaban, A. T.; Harary, F. Chemical Graphs. V. Enumeration and Proposed Nomenclature of Benzenoid Cata-condensed Polycyclic Aromatic Hydrocarbons. *Tetrahedron* **1968**, *24*, 2505−2516.
(27) Trinajstić, N. On the Classification of Polyhex Hydrocarbons. *J. Math. Chem.* **1990**, *5*, 171−175.
(28) Hosoya, H. Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332−2339.
(29) Balaban, A. T. Topological Indices based on Topological Distances in Molecular Graphs. *Pure Appl. Chem.* **1983**, *55*, 199−206.
(30) Entriger, R. C.; Jackson, D. E.; Snyder, D. A. Distance in Graphs. *Czech. Math. J.* **1976**, *26*, 283−296.
(31) Skorobogatov, V. A.; Dobrynin, A. A. Metric Analysis of Graphs. *Math. Chem. (MATCH)* **1988**, *23*, 105−151.
(32) Bonchev, D. Information Indices for Atoms and Molecules. *Math. Chem. (MATCH)* **1979**, *7*, 65−113.
(33) Shannon, C. E.; Weaver, W. *Mathematical Theory of Communications*; University of Illinois: Urbana, 1949.
(34) Bonchev, D.; Trinajstić, N. Information Theory, Distance Matrix and Molecular Branching. *J. Chem. Phys.* **1977**, *67*, 4517−4533.
(35) Dobrynin, A. A. The Effective Algorithm of Generation for Graphs of Unbranched Hexagonal Systems. *Vychisl. Sistemy* **1983**, *130*, 3−38 (in Russian).


CI9502461