

Topological Index and Some Counting Polynomials for Characterizing the Topological Structure and Properties of Molecular Graphs

Haruo HOSOYA

*Department of Chemistry, Ochanomizu University,
Bunkyo-ku, Tokyo 112 Japan*

Abstract. To characterize the topological features of the carbon atom skeleton, or a graph (G), of a hydrocarbon molecule, a number of “topological indices” have been proposed by chemists. The Z -index, Z_G , proposed by the present author was defined in terms of the non-adjacent number, $p(G, k)$, and is closely related to the characteristic polynomial, $P_G(x)$, whose zeroes give the stability of conjugated hydrocarbons. Matching polynomial, $M_G(x)$, was also defined in terms of $p(G, k)$ and its relation with $P_G(x)$ has been analyzed. Contrary to other topological indices including Wiener’s w and p numbers, Z_G and $M_G(x)$ are shown to have a number of interesting mathematical properties. The series of Z_G and $M_G(x)$ for certain series of graphs are found to be transformed into typical orthogonal polynomials, e.g., Hermite, Laguerre, and Chebyshev polynomials. Relation between topological indices and thermodynamic quantities of saturated and unsaturated hydrocarbon molecules was discussed.

1. Introduction

More than ten millions of chemical substances have hitherto been known. Most of them contain carbon atom(s) and are called organic compounds only with a few exceptions (e.g. CO, CO₂, HCN). Hydrocarbon molecules, composed of carbon (C) and hydrogen (H) atoms and depicted as C_mH_n, occupy the central part of the whole family of organic compounds. A saturated hydrocarbon is a hydrocarbon that

can no longer bind with more hydrogen atoms without destroying its carbon atom skeleton. The others called as unsaturated hydrocarbons, such as benzene (see Fig. 1), carry free electrons, and show interesting electronic and sometimes magnetic properties.

The topological structure of a molecule is conventionally expressed by the structural formula composed of the atomic symbols for the components and connecting lines for the chemical bonds, or adjacency relations between the component atoms. In a structural formula, saturated bonds were expressed by single lines while unsaturated bonds were sometimes depicted by multiple lines depending on the degree of unsaturation.

If one substitutes all the atomic symbols and multiple lines of the structural formula into points (vertices) and single lines (edges), respectively, one gets nothing else but what is called a connected undirected graph defined in the graph theory (Harary, 1969). One may call this graph a molecular graph (Balaban, 1976; Trinajstić, 1983; Gutman and Polansky, 1986). Further, in many cases, all the vertices representing hydrogen atoms and the incident lines are suppressed to give smaller graphs. In Fig. 1 are given the molecular formulas, structural formulas, and molecular graphs (of the carbon atom skeletons) of saturated and unsaturated hydrocarbon molecules, butane and benzene, respectively. Incidentally, the two graphs are respectively the members of the path graph (S_N) and cycle graph (C_N).

A graph G composed of N points can be represented by an adjacency matrix A ($N \times N$), whose element A_{ij} is either 1 or 0 depending on the adjacency relation

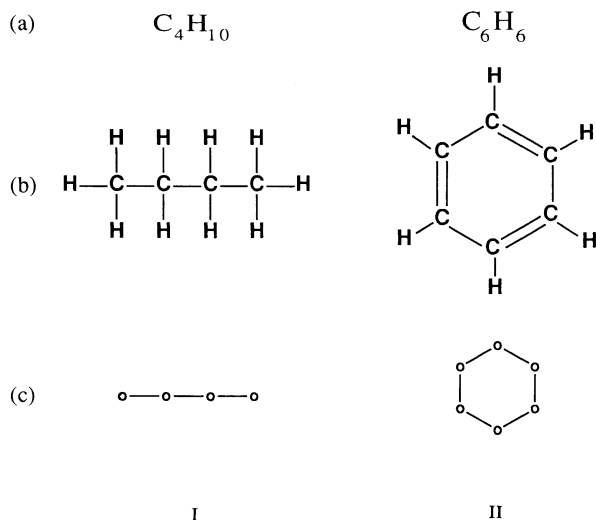


Fig. 1. Molecular formula (a), structural formula (b), and molecular graph (c) of butane (I) and benzene (II).

of points i and j (Fig. 2). The distance matrix \mathbf{D} can also be defined with the element D_{ij} representing the number of the shortest walk between points i and j . The three mathematical objects \mathbf{G} , \mathbf{A} , and \mathbf{D} are equivalent. Namely, the two matrices have the same information content to reproduce a given graph \mathbf{G} . In an attempt to get more compact quantity, such as a set of numbers characterizing the topological structure of a graph, the characteristic polynomial $P_G(x)$ has been defined as

$$P_G(x) = (-1)^N \det(\mathbf{A} - x\mathbf{E}) \quad (1)$$

where \mathbf{E} is the unit matrix of the order N (Cvetković *et al.*, 1979). The characteristic polynomials of the two graphs given in Fig. 1 are, respectively,

$$x^4 - 3x^2 + 1 \quad \text{and} \quad x^6 - 6x^4 + 9x^2 - 4$$

Although for smaller N , say smaller than six, $P_G(x)$ uniquely corresponds to the respective graph, a chance that more than two different graphs have the same $P_G(x)$ rapidly increases with N . The set of graphs with the same $P_G(x)$ are called isospectral or cospectral with each other. Algorithms for constructing as many isospectral graphs have been proposed (Herndon, 1974). The next problem is to seek effective indices with smaller bits of information for characterizing the topological nature of a molecular graph.

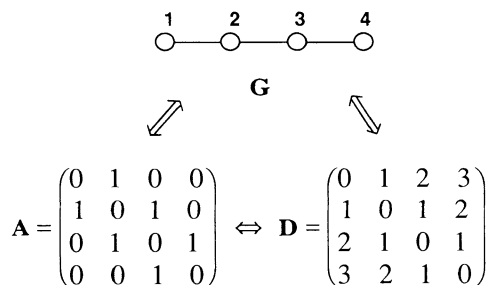


Fig. 2. Graph (\mathbf{G}), adjacency matrix (\mathbf{A}), and distance matrix (\mathbf{D}) of the carbon atom skeleton of butane. Note that butadiene C_4H_6 has the same \mathbf{G} , \mathbf{A} , and \mathbf{D} as butane.

2. Topological Index

In 1971 the present author proposed to define the so-called topological index, Z_G , for characterizing the topological nature of saturated hydrocarbon molecules, as

$$Z_G = \sum_{k=0}^m p(G, k) \quad (2)$$

where $p(G, k)$, the non-adjacent number, is the number of ways for choosing k ($\leq m = [N/2]$) disjoint lines from a given graph G , $p(G, 0)$ being defined to be unity for all the graphs (Hosoya, 1971, 1973).

Later, so many different versions of topological indices have been proposed mostly by chemists that nowadays the term "topological index" is used as the general name for those indices (Balaban, 1976; Trinajstić, 1983; Gutman and Polansky, 1986). Thus, in this paper, we call the Z_G of Eq. (2) the Z-index.

It is convenient to define the following Z-counting polynomial $Q_G(x)$ as

$$Q_G(x) = \sum_{k=0}^m p(G, k)x^k \quad (3)$$

with which Z_G is obtained to be $Q_G(1)$.

For the two graphs, I and II, given in Fig. 1 one can easily obtain $Q_G(x)$ and Z_G , respectively, as

$$\text{I} \quad Q_G(x) = 1 + 3x + x^2 \quad Z_G = 5$$

$$\text{II} \quad Q_G(x) = 1 + 6x + 9x^2 + 2x^3 \quad Z_G = 18$$

As shown in Tables 1 and 2, the Z-indices of the path graphs and cycle graphs are known to form, respectively, the Fibonacci and Lucas numbers, namely,

$$Z(S_N) = F_N \quad (4)$$

$$Z(C_N) = L_N \quad (5)$$

with

$$F_N = F_{N-1} + F_{N-2} \quad (N \geq 2) \quad (\text{Fibonacci numbers}) \quad (6)$$

$$F_0 = F_1 = 1$$

and

$$L_N = L_{N-1} + L_{N-2} \quad (N \geq 2)$$

Table 1. Non-adjacent numbers $p(G, k)$'s and Z-index of path graphs S_N 's.

N	S_N	$p(G,k)$					Z_G
		k=0	1	2	3	4	
1		1					1
2		1	1				2
3		1	2				3
4		1	3	1			5
5		1	4	3			8
6		1	5	6	1		13
7		1	6	10	4		21
8		1	7	15	10	1	34

(Lucas numbers) (7)

$$L_0 = 2 \quad L_1 = 1$$

Useful recursion relations are found for $p(G, k)$, $Q_G(x)$, and Z-index by using the inclusion-exclusion principle as follows (Hosoya, 1971, 1973):






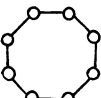
$$p(G, k) = p(G-l, k) + p(G \ominus l, k-1) \tag{8}$$

$$Q_G(x) = Q_{G-l}(x) + xQ_{G \ominus l}(x) \tag{9}$$

$$Z_G = Z_{G-l} + Z_{G \ominus l} \tag{10}$$

where $G-l$ means the subgraph of G obtained by deleting the edge l , and $G \ominus l$ the one by deleting l and all the lines incident to l (Hosoya and Hosoi, 1976). Note that

Table 2. Non-adjacent numbers $p(G, k)$'s and Z-index of cycle graphs C_N 's.

N	C_N	$p(G, k)$					Z_G
		$k=0$	1	2	3	4	
3		1	3				4
4		1	4	2			7
5		1	5	5			11
6		1	6	9	2		18
7		1	7	14	7		29
8		1	8	20	16	2	47

the second term in the right-hand-side of Eq. (9) is multiplied by x .

A connected graph without any ring is called a tree graph, while the one with at least a ring a non-tree graph. The characteristic polynomial $P_G(x)$ already defined in Eq. (1) for a tree graph can be expressed in terms of $p(G, k)$'s as follows (Hosoya, 1971, 1972a):

$$P_G(x) = \sum_{k=0}^m (-1)^k p(G, k) x^{N-2k} \quad (G \in \text{tree}) \quad (11)$$

while for a non-tree graph, one can show the following relation between $P_G(x)$ and $p(G, k)$'s for G and all the set of the subgraphs of G obtained by deleting the component ring R_i in G :

$$+ \sum_i^{\text{rings}} (-2)^{r_i} \sum_{k=0}^m (-1)^k p(G \ominus R_i, k) x^{N-n_i-k}. \quad (12)$$

The first summation in the second term of the right-hand-side of Eq. (12) runs over

$\{R_i\}$'s), which are not only all the component rings but also all the possible combinations of disjoint rings in G . The symbols n_i and r_i , respectively, stand for the numbers of points and components of R_i .

The characteristic polynomial of a graph turns out to be equivalent to the secular determinant, $\det(H - S\varepsilon)$, used in the Hückel molecular orbital (HMO) method for calculating the π -electronic structure of conjugated hydrocarbon molecules, through the transformation, $x = (\alpha - \varepsilon)/\beta$, where α and β are the elements of the matrix H , respectively, for the component atoms (diagonal) and bonds (off-diagonal) of the molecular graph G , and S is assumed to be diagonal ($S_{ij} = \delta_{ij}$) (Hückel, 1931).

The Z -index and $P_G(x)$, as well as other topological indices, are known to be useful not only for characterizing the topological nature of molecules but also for explaining and predicting various properties of molecules and substances (Balaban, 1976; Trinajstić, 1983; Gutman and Polansky, 1986; Hosoya, 1971, 1972b; Hosoya *et al.*, 1972; Gao and Hosoya, 1988).

In this respect, the modified Z -index, \tilde{Z}_G , has been defined through $P_G(x)$ but not through $p(G, k)$ numbers as

$$\tilde{Z}_G = (-i)^N P_G(i). \quad (13)$$

For tree graphs, Z_G is identical with \tilde{Z}_G , whereas for non-tree graphs, their difference $\Delta Z_G = \tilde{Z}_G - Z_G$ for the graphs of conjugated unsaturated hydrocarbon molecules, such as butadiene and benzene families, plays an important role for predicting the stability of molecules (Hosoya *et al.*, 1975).

3. Matching Polynomial and Orthogonal Polynomials

Independently by many groups of researchers, the matching polynomial has been proposed to be defined in terms of the $p(G, k)$ numbers as follows (Aihara, 1976; Gutman *et al.*, 1977; Farrell, 1979):

$$M_G(x) = \sum_{k=0}^m (-1)^k p(G, k) x^{N-2k}. \quad (14)$$

As evident from Eq. (11) $M_G(x)$ is identical to $P_G(x)$ for tree graphs. The difference between $P_G(x)$ and $M_G(x)$ for non-tree graphs is expressed by the second term of Eq. (12) in terms of the $p(G, k)$ contributions from the component rings.

It was shown that the matching polynomials of several series of graphs have close relationships with the typical orthogonal polynomials (Heilman and Lieb, 1970; Gutman *et al.*, 1977; Farrell, 1979). The $M_G(x)$'s of the path graphs (S_N) and cycle graphs (C_N) given in Tables 1 and 2 can be transformed, respectively, into the

Chebyshev polynomials of the second and first kinds, $U_N(\cos\theta)$ and $T_N(\cos\theta)$. Namely, if they are defined as

$$U_N(\cos\theta) = \sin(N+1)\theta / \sin\theta \quad (15)$$

$$T_N(\cos\theta) = \cos N\theta, \quad (16)$$

one gets the following relations

$$M_{S_N}(x) = U_N(x/2) \quad (17)$$

$$M_{C_N}(x) = 2T_N(x/2). \quad (18)$$

The complete graph K_N is defined as the graph in which all the N vertices are connected to each other (Harary, 1969). The matching polynomial of K_N can be transformed into the Hermite polynomial H_N as follows (Table 3):

$$M_{K_N}(x) = 2^{-N/2} H_N(x/\sqrt{2}), \quad (19)$$

where H_N is defined as follows:

Table 3. Relation between the Hermite polynomial and the matching polynomial of complete graphs K_N 's.








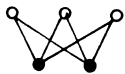
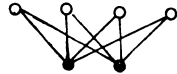


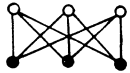
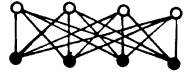
N	K_N	$M_{K_N}(x)$	$H_N(x)$
1		x	2x
2		$x^2 - 1$	$4x^2 - 2$
3		$x^3 - 3x$	$8x^3 - 12x$
4		$x^4 - 6x^2 + 3$	$16x^4 - 48x^2 + 12$
5		$x^5 - 10x^3 + 15x$	$32x^5 - 160x^3 + 120x$

Table 4. Relation between the associated Laguerre polynomial and the matching polynomial of complete bipartite graphs $K_{M,N}$'s.

M	N	$K_{M,N}$	$M_{K_{M,N}}(x)$	$L_N^{M-N}(x)$
2	1		$x^3 - 2x$	$x - 2$
3	1		$x^4 - 3x^2$	$-x + 3$
3	2		$x^5 - 6x^3 + 6x$	$\frac{1}{2}x^2 - 3x + 3$
4	2		$x^6 - 8x^4 + 12x^2$	$\frac{1}{2}x^2 - 4x + 6$

$$M_{K_{M,N}}(x) = \left\{ (-1)^N N! x^{M-N} \right\} L_N^{M-N}(x^2) \quad (M \geq N)$$

Table 5. Relation between the Laguerre polynomial and the matching polynomial of complete bipartite graphs $K_{N,N}$'s.

N	$K_{N,N}$	$M_{K_{N,N}}(x) = (-1)^N N! L_N(x^2)$
0	\emptyset	1
1		$x^2 - 1$
2		$x^4 - 4x^2 + 2$
3		$x^6 - 9x^4 + 18x^2 - 6$
4		$x^8 - 16x^6 + 72x^4 - 96x^2 + 24$

$$H_N(x) = (-1)^N \exp(x^2) d^N / dx^N \{ \exp(-x^2) \}. \quad (20)$$

A bipartite graph is defined as a connected graph with two classes of vertices, say M white and N black ones, such that no two vertices in the same class are connected. A complete bipartite graph denoted as $K_{M,N}$ is a bipartite graph in which all the M and N vertices in the Two classes are connected with each other.

The matching polynomial of the complete bipartite graph $K_{M,N}$ is shown to be identical with the associated Laguerre polynomial $L_M^N(x^2)$ as follows (Table 4):

$$M_{K_{M,N}} = (-1)^N N! x^{M-N} / N! L_N^{M-N}(x^2). \quad (21)$$

$$L_M^N(x) = \sum_{k=0}^M (-1)^k \binom{M+N}{M-k} \frac{x^k}{k!} \quad (22)$$

For a special case with $M = N$ the matching polynomial exactly corresponds to the Laguerre polynomial L_N as in Table 5.

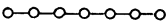
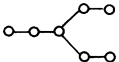
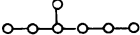
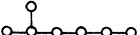
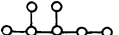
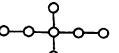
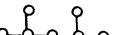
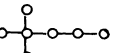
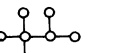
4. Structural Dependency of Molecular Properties

Based on a tremendously large amount of knowledge of experimental facts on the properties of chemical substances, chemists have abstracted various empirical relationships between thermodynamic properties and the topological structure of molecules, especially of organic molecules. Although this problem by itself is very important in the understanding of the nature, prediction of a given thermodynamic property of a substance from the first principle is a challenging but extremely difficult task, since it is an outcome of the cooperative phenomena of randomly moving $10^{10} \sim 10^{20}$ particles under the combination of various physical conditions, say, temperature, pressure, and volume. Recently, from some pragmatic motivations, study on this type of correlation if refocused by drug-designing chemists and is called quantitative structure-activity-relationship, or simply as QSAR.

Our standpoint is a little different from either of both the sides explained above. By analyzing the correlation between the physico-chemical properties of molecules and mathematical properties abstracted from the topological structures of molecules, one would be able to find a breakthrough toward the goal.

In Table 6, the relation between the boiling points and the Z -indices of the nine isomers of heptane, C_7H_{16} , are given (Hosoya, 1971; Hosoya *et al.*, 1972; Narumi and Hosoya, 1985). It is interesting to observe that the boiling point decreases with the degree of branching of the molecular skeleton in proportion to the value of the Z -index, and nine distinct integers from 13 to 21 were assigned to all the family of isomers. The entropy, ΔS° , of these hydrocarbons is found to be also in parallel with Z_G as seen in Fig. 3, where other thermodynamic properties of these isomers are also

Table 6. Relation between the thermodynamic quantities (boiling point, bp, and density of liquid, d) and the two topological indices (Z_G and p) of heptane isomers.

	Graph	$p(G,k)$				Z_G	bp(°C)	p	d(20°C)
		k=0	1	2	3				
1		1	6	10	4	21	98.4	4	0.6838
2		1	6	9	4	20	93.4	6	0.6982
3		1	6	9	3	19	91.9	5	0.6871
4		1	6	9	2	18	90.0	4	0.6786
5		1	6	8	2	17	89.7	6	0.6951
6		1	6	7	2	16	86.0	6	0.6933
7		1	6	8	0	15	80.5	4	0.6727
8		1	6	7	0	14	79.2	4	0.6739
9		1	6	6	0	13	80.9	6	0.6901

plotted and compared with two other topological indices, Wiener number, w , and path number, p , or p_3 .

The indices, w and p , were proposed by Harold Wiener in his pioneering papers of QSAR on thermodynamic properties of saturated hydrocarbons in 1947 (Wiener, 1947), just in the same year as Nobert Wiener introduced his cybernetics. The half-sum of the off-diagonal elements of D for G is w (Hosoya, 1971), while p is half the number of such matrix elements with $D_{ij} = 3$, namely the number of pairs of vertices whose shortest distance is three. Correlation between Z_G and w is fair but not so good, whereas p has no correlation with them at all.

It is clear from Figs. 3a and b that there are two distinct types of thermodynamic properties of these hydrocarbons, i.e., A-type being correlated well with Z_G or w , and B-type with p (Narumi and Hosoya, 1985). We can give a qualitative explanation for the good correlation between Z_G and the A-type properties, i.e.,

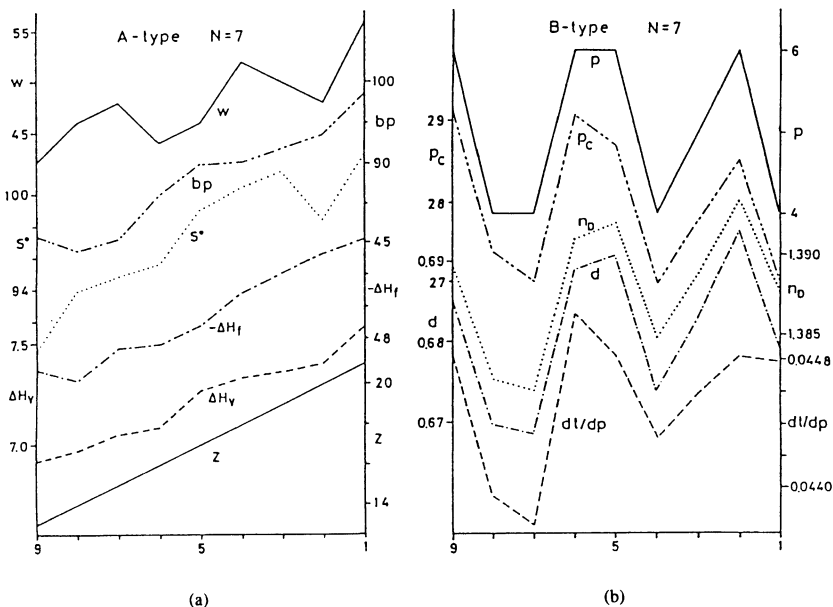


Fig. 3. Correlation between various thermodynamic quantities and topological indices of heptane isomers. (a) A-type properties dependent on Z and (b) B-type on p (See text).

boiling point (bp), entropy (ΔS°), heat of formation (ΔH_f), and heat of vaporization (ΔH_v), as follows. With the increment of branching, the degree of freedom of the internal rotations around each CC bond decreases, lowers the specific heat, and also lowers the entropy. Parallelism between the entropy and boiling point is easily understood (Narumi and Hosoya, 1980). The A-type properties were interpreted to be governed by dynamical features of the topological structure of molecular skeletons.

On the other hand, the B-type properties, with a good correlation with the index, e.g., the density of liquid, d , refractive index, n_D , critical pressure, p_C , etc., are largely governed by the static features of the topological structure of molecules. Namely, the value of these properties reflects some degree of close-packing of molecules whose behavior looks something like that of rigidly moving particles in a vessel of a given volume.

As has been mentioned before, the Z-index for large conjugated hydrocarbons such as polyenes and polycyclic aromatic hydrocarbons is shown to have a good correlation with their π -electronic properties as the bond order and π -electronic energy (Hosoya and Hosoi, 1976; Hosoya *et al.*, 1975).

REFERENCES

- Aihara, J. (1976), A new definition of Dewar-type resonance energy, *J. Am. Chem. Soc.*, **98**, 2750–2758.
- Balaban, A. T. ed. (1976), *Chemical Applications of Graph Theory*, Academic Press, London.
- Cvetković, D., Doob, M., and Sachs, H. (1979), *Spectra of Graphs. Theory and Applications*, Academic Press, Berlin.
- Farrell, E. J. (1979), An introduction to matching polynomial, *J. Combin. Theory*, **B27**, 75–86.
- Gao, Y.-D. and Hosoya, H. (1988), Topological index and thermodynamic properties. IV. Size dependency of the structure-activity correlation of alkanes, *Bull. Chem. Soc. Jpn.*, **61**, 3093–3102.
- Gutman, I. (1979), The matching polynomial, *MATCH* **6**, 75–91.
- Gutman, I., Milun, M., and Trinajstić, N. (1977), Non-parametric resonance energies of arbitrary conjugated systems, *J. Am. Chem. Soc.*, **99**, 1692–1704.
- Gutman, I. and Polansky, O. E. (1986), *Mathematical Concepts in Organic Chemistry*, Springer-Verlag, Berlin.
- Harary, F. (1969), *Graph Theory*, Addison-Wesley, Reading, MA.
- Heilman, O. E. and Lieb, E. H. (1970), Monomers and dimers, *Phys. Rev. Lett.*, **24**, 1412–1414.
- Herndon, W. C. (1974), Isospectral molecules, *Tetrahedron Lett.*, 671–674.
- Hosoya, H. (1971), Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons, *Bull. Chem. Soc. Jpn.*, **44**, 2331–2339.
- Hosoya, H. (1972a), Graphical enumeration of the coefficients of the secular polynomials of the Hückel molecular orbitals, *Theor. Chim. Acta. (Berl.)*, **25**, 215–222.
- Hosoya, H. (1972b), Topological index as a sorting device for coding chemical structures, *J. Chem. Doc.*, **12**, 181–183.
- Hosoya, H. (1973), Topological index and Fibonacci numbers with relation to chemistry, *Fibonacci Quart.*, **11**, 255–266.
- Hosoya, H. (1981), Graphical and combinatorial aspects of some orthogonal polynomials, *Natur. Sci. Rept. Ochanomizu Univ.*, **32**, 127–138.
- Hosoya, H. and Hosoi, K. (1976), Topological index as applied to π -electronic systems. III. Mathematical relations among various bond orders, *J. Chem. Phys.*, **64**, 1065–1073.
- Hosoya, H., Hosoi, K., and Gutman, I. (1975), A topological index for the total π -electron energy. Proof of a generalised Hückel rule for an arbitrary network, *Theor. Chim. Acta. (Berl.)*, **38**, 37–47.
- Hosoya, H., Kawasaki, K., and Mizutani, K. (1972), Topological index and thermodynamic properties. I. empirical rules on the boiling point of saturated hydrocarbons, *Bull. Chem. Soc. Jpn.*, **45**, 3415–3421.
- Hückel, E. (1931), Quanten theoretische Beiträge zum Benzolproblem. I., *Z. Phys.*, **70**, 204–272.
- Narumi, H. and Hosoya, H. (1980), Topological index and thermodynamic properties. II. Analysis of the topological factors on the absolute entropy of acyclic saturated hydrocarbons, *Bull. Chem. Soc. Jpn.*, **53**, 1228–1237.
- Narumi, H. and Hosoya, H. (1985), Topological index and thermo-dynamic properties. III. Classification of various topological aspects of properties of acyclic saturated hydrocarbons, *Bull. Chem. Soc. Jpn.*, **58**, 1778–1786.
- Trinajstić, N. (1983), *Chemical Graph Theory*, CRC Press, Boca Raton, FL.
- Wiener, H. (1947), Structural determination of paraffin boiling points, *J. Am. Chem. Soc.*, **69**, 17–20.