  Bull. Chem. Soc. Jpn., 60, 2611—2614 (1987)

Approximate Formulas for Hosoya's Topological Index

Ivan GUTMAN* and Zoran MARKOVIĆ
Faculty of Science, University of Kragujevac, P.O.Box 60,
YU-34000 Kragujevac, Yugoslavia
(Received June 30, 1986)

Approximate formulas for Hosoya's topological index (Z) are reported for the first time. These formulas reproduce ln Z with an error of some 1—2% and require the knowledge of only a few easily obtainable invariants of the molecular graph.

The topological index Z was introduced by Hosoya[1] in 1971. Since then numerous scientific results on this quantity have accumulated[1-11] revealing the importance of Z in various branches of theoretical chemistry and theoretical physics. In particular, the relations between Z and several physicochemical properties of alkanes are nowadays well established.[3,5,7,10] Hosoya also showed[2] that Z and the closely related nonadjacent numbers play a significant role in the molecular orbital theory of conjugated π-electron systems.

This later resulted in the introduction of the modified topological index,[12] topological bond order,[13,14] matching polynomial,[15-17] a novel theory of aromaticity,[18,19] novel approach to total π-electron energy,[12,20] π-electron charge distribution[21] etc.

A detailed account of various applications of the index Z and especially of its mathematical properties can be found in Chapter 11 of the book.[11]

In spite of such a great interest towards Hosoya's topological index, no attempts seem to have been made to find approximate methods for its calculation. This is even more surprising if one knows that the evaluation of Hosoya's index is a very hard combinatorial task, especially in the case of large polycyclic molecules. In the present paper we wish to fill this gap and to provide a class of reliable approximate formulas for Z, enabling its estimation from a number of easily obtainable invariants of the molecular graph.

Preliminaries

Hosoya's topological index is defined as[1]

$$Z = Z(G) = \sum_{k=0}^{h} m(G,k),$$

where $m(G,k)$ is the number of k-matchings of the molecular graph G. This means that one can select k independent edges in G in $m(G,k)$ distinct ways; therefore $m(G,k)$ is often called the k-th nonincident number of G.

In what follows the molecular graph considered will be assumed to possess n vertices and m edges, whereas h will denote the size of the maximal matching (i.e. $m(G,h)\neq 0$, $m(G,h+1)=0$). Note that $m(G,1)=m$ and $h\leqslant n/2$. Besides, by definition, $m(G,0)=1$ for all graphs G.

The Z-counting polynomial

$$Q(x) = \sum_{k=0}^{h} m(G,k)x^k$$

is an auxiliary quantity, introduced within the theory of Hosoya's index,[1] with the immediate property

$$Q(1) = Z.$$

A less obvious property of $Q(x)$ is that all its zeros are real and negative numbers.[17] Therefore we can write the Z-counting polynomial as

$$Q(x) = \prod_{i=1}^{h} (1+p_i x)$$

where $p_i>0$ and $-1/p_i$ is a root of $Q(x)=0$, $i=1,2,\cdots,h$. Consequently,

$$Z = \prod_{i=1}^{h} (1+p_i).$$

The Approximation

The basic idea of the present consideration is to approximate the Z-counting polynomial $Q(x)$ by means of another polynomial $Q^*(x)$,

$$Q^*(x) = (1+Ax)^t(1+Bx)^{h-t} \quad (1)$$

and to adjust the parameters A and B so that the first three coefficients of $Q^*(x)$ and $Q(x)$ coincide:

$$Q^*(x) = 1 + m(G,1)x + m(G,2)x^2 + \cdots . \quad (2)$$

$Q^*(1)$ would then provide an approximate expression for Z.

Combining Eqs. 1 and 2 we arrive at

$$tA + (h-t)B = m(G,1) \quad (3a)$$

$$\binom{t}{2}A^2 + \binom{h-t}{2}B^2 + t(h-t)AB = m(G,2). \quad (3b)$$

It is easily seen that the approximation $Q(x)\approx Q^*(x)$ is equivalent to the assumption that among the zeros of $Q(x)$, t of them are mutually equal and the remaining $h-t$ zeros are also mutually equal. The problems arising in connection with the choice of the third parameter in Eq. 1, namely t, will be discussed in the subsequent section.

Bearing in mind that

$$m(G, 1) = m,$$

and

$$m(G, 2) = (m^2+m-d)/2,$$

where d denotes the sum of the squares of the vertex degrees of the graph G, we can transform the Eqs. 3a and 3b into the form

$$tA + (h-t)B = m \tag{4a}$$

$$tA^2 + (h-t)B^2 = d-m. \tag{4b}$$

The solution of the system (4) in which we are interested in is

$$A = (m+RS)/h, \tag{5}$$

$$B = (m-R/S)/h, \tag{6}$$

where

$$R = \sqrt{hd-m^2-hm}, \tag{7}$$

$$S = \sqrt{(h-t)/t}\,.$$

We show now that R is necessarily a real quantity. In order to do this note that the right-hand sides of Eqs. 4a and 4b are equal to the sum of the numbers $p_1, p_2, \cdots, p_h$ and $p_1^2, p_2^2, \cdots, p_h^2$, respectively. Since the inequality

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 \leqslant \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

holds for arbitrary numbers $x_1, x_2, \cdots, x_n$, we conclude that

$$\left(\frac{1}{h}\sum_{i=1}^{h} p_i\right)^2 \leqslant \frac{1}{h}\sum_{i=1}^{n} p_i^2$$

i.e.

$$\left(\frac{1}{h}m\right)^2 \leqslant \frac{1}{h}(d-m)$$

i.e.

$$hd - m^2 - hm \geqslant 0$$

which is a sufficient condition for the reality of R.

It is now immediate to conclude that A is positive for all $1 \leqslant t \leqslant h-1$ whereas B is positive for $1 \leqslant t \leqslant T$; where T stands for the integer part of $m^2/(d-m)$.

Substituting Eqs. 5 and 6 back into Eq. 1 and setting $x=1$, we arrive at

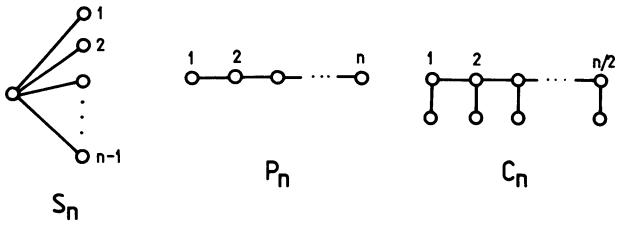
$$Z^x = Z^x(t) = [1+(m+RS)/h]^t[1+(m-R/S)/h]^{h-t} \tag{8}$$

which is the required algebraic expression for our approximate formulas for Hosoya's index Z.

As a matter of fact, it can be shown[23] that $Z^x(t)$ is a monotonously decreasing function of the parameter t and that for $t=1$, $Z^x(t)$ is an upper bound for Z.

The Function Z^x of Some Series of Graphs

In order to illustrate Eq. 8 we have examined the function Z^x for three classes of graphs, namely for the star S_n, the path P_n, and the comb C_n:



Both S_n, P_n, and C_n have n vertices and $n-1$ edges. Their topological indices are given as

$$Z(S_n) = n$$

$$Z(P_n) = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right]$$

$$Z(C_n) = \frac{1}{\sqrt{8}}[(1+\sqrt{2})^{n/2+1} - (1-\sqrt{2})^{n/2+1}].$$

For the star S_n we have $h=1$ and $d=(n-1)\cdot 1^2+1\cdot(n-1)^2=n(n-1)$ and therefore $R=0$. Consequently, $Z^x(S_n)=Z(S_n)$ for all values of n.

In the case of the path P_n, $h=n/2$ if n is even and $h=(n-1)/2$ if n is odd. Furthermore, $d=2\cdot 1^2+(n-2)\cdot 2^2=4n-6$. Substituting this back into Eqs. 7 and 8 we obtain an expression for $Z^x(P_n)$ whose form considerably differs from that of $Z(P_n)$. However, for large values of n,

$$Z^x(P_n) \approx t^{-t/2}(4n/9)^{t/2}(\sqrt{3})^n$$

whereas

$$Z(P_n) \approx \frac{1+\sqrt{5}}{2\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n.$$

This means that for large n, the logarithms of $Z(P_n)$ and $Z^x(P_n)$ are linearly proportional and, in particular,

$$\lim_{n\to\infty} \ln Z(P_n)/\ln Z^x(P_n) = [\ln(1+\sqrt{5}) - \ln 2]/[(\ln 3)/2]$$

$$= 0.88$$

The above limit is independent of the parameter t and is not very different from the value 0.91, estimated for alkanes by least squares fitting (see later).

The case of the comb C_n is similar. Here $h=n/2$, $d=5n-10$ and for large values of n,

$$Z^x(C_n) \approx t^{-t/2}(5n/9)^{t/2}3^{n/2}$$

whereas

$$Z(C_n) \approx \frac{1+\sqrt{2}}{\sqrt{8}}(1+\sqrt{2})^{n/2}.$$

Therefore also the logarithms of $Z(C_n)$ and $Z^x(C_n)$ are linearly proportional and

$$\lim_{n\to\infty} \ln Z(C_n)/\ln Z^x(C_n) = \ln(1+\sqrt{2})/\ln 3 = 0.80.$$

The above examples can be generalized. Let G_1, G_2, $\cdots$, G_n, $\cdots$ be a series of graphs, such that G_{i+1} has

more vertices than G_i. Suppose that the parameter $d(G_n)$ is a linear function of n, say

$$d(G_n) = pn + q. \quad (9)$$

Suppose further that the limit L,

$$L = \lim_{n\to\infty} m(G_n)/h(G_n) \quad (10)$$

exists. Then for large values of n, the function $Z^x(G_n)$ behaves asymptotically as

$$Z^x(G_n) \approx t^{-t/2}\left(\frac{\sqrt{p}}{1+L}\right)^t (1+L)^{h(G_n)}.$$

Note that the conditions (9) and (10) are fulfilled for practically all molecular graphs. Furthermore, in almost all cases of chemical interest, the limit L equals to 2 and $h(G_n)\sim n(G_n)/2$.

Numerical Work

As already indicated in the previous section, a good linear correlation exists between ln Z and ln Z^x. We have examined this correlation on two samples of molecules—the hexanes, heptanes, octanes, and decanes from the book[24] (a total of 141 alkanes) and the benzenoid hydrocarbons from the paper[25] (a total of 116 benzenoid molecules). The results obtained are presented in Tables 1 and 2. Two approximate formulas have been investigated, viz.

$$\ln Z = a_0 \ln Z^x \quad (11)$$

$$\ln Z = a_1 \ln Z^x + b_1 \quad (12)$$

and the coefficients a_0, a_1, and b_1 calculated by least squares fitting.

The parameter t in Eq. 8 needs not be necessarily considered as an integer and we have examined the correlations of the type (11) and (12) for both integer and noninteger values of t, $0<t\leqslant T$. Our calculations show, however, that the accuracy and the reliability of the approximations (11) and (12) are not very sensitive to the choice of t. If t is set to be a constant, then $t=1$ in the case of alkanes, and $t=2$ in the case of benzenoids gave the best results. For benzenoids a somewhat better correlation between Z and Z^x is found when t is adjusted to be equal to $T/2$.

The choice $t=T$ gave the worst correlation in both cases.

Table 1. Correlation between Hosoya's Index and $Z^x(t)$ in the Case of Alkanes with 6, 7, 8, 9, and 10 Carbon Atoms[24]

t	Formula	a_0 or a_1	b_1	Correlation coefficient	Mean error %	Max. error observed %
1	11	0.9707	—	0.995	1.4	3.0
1	12	0.9145	0.2224	0.995	1.1	3.0
$T/2$	11	0.9698	—	0.993	1.4	3.6
$T/2$	12	0.9507	0.0757	0.993	1.3	3.9
T	11	1.0200	—	0.984	2.0	6.1
T	12	1.0144	0.0212	0.984	2.0	6.2

Discussion

Our final conclusion is that in the case of alkanes

$$\ln Z = 0.914 \ln Z^x(t) + 0.222;\ t = 1,$$

whereas in the case of benzenoid hydrocarbons (as well as other polycyclic conjugated molecules)

$$\ln Z = 0.951 \ln Z^x(t) + 0.082;\ t = T/2$$

should be used for the approximate calculation of Hosoya's topological index. These formulas estimate ln Z with an error of some 1—2%, which is an acceptable accuracy for all practical applications of Hosoya's index.

The above formulas have been designed so to reproduce the topological indices of medium sized molecular graphs. However, in view of the three examples examined in a previous section, one may expect that formulas of the type (11) are capable to estimate the (logarithms of the) Hosoya indices of very large molecular graphs and infinite networks.

The right-hand sides of the above approximate expressions depend on the following invariants of the molecular graph: Number of edges (m), vertex degree sequence and size of the maximal matchings (h). The number of vertices and the number of maximal matchings play no (explicit) role in the present approach.

All the required invariants are easily obtained by inspection of the molecular graph. In particular, in the case of acyclic molecular graphs (i.e. molecular graphs of alkanes), $h=(n-n_0)/2$, where n is the number of vertices and n_0 is the number of isolated vertices which are obtained from the molecular graph by deleting from it a vertex of degree one and its first neighbor, and by repeating this deletion procedure

Table 2. Correlation between Hosoya's Index and $Z^x(t)$ in the Case of Benzenoid Hydrocarbons[25]

t	Formula	a_0 or a_1	b_1	Correlation coefficient	Mean error %	Max. error observed %
1	11	0.9166	—	0.9992	0.7	5.0
1	12	0.8917	0.3355	0.9992	0.6	4.0
2	11	0.9359	—	0.9995	0.8	10.2
2	12	0.9015	0.4554	0.9995	0.4	2.2
$T/2$	11	0.9574	—	0.9996	0.4	1.4
$T/2$	12	0.9511	0.0818	0.9996	0.4	1.4
T	11	1.0410	—	0.997	0.9	3.6
T	12	1.0282	0.1569	0.997	1.0	4.7

as long as it is possible. For practically all benzenoid hydrocarbons, $h=n/2$ if n is even and $h=(n-1)/2$ if n is odd.

References

1) H. Hosoya, *Bull. Chem. Soc. Jpn.*, **44**, 2332 (1971).
2) H. Hosoya, *Theor. Chim. Acta*, **25**, 215 (1972).
3) H. Hosoya, K. Kawasaki, and K. Mizutani, *Bull. Chem. Soc. Jpn.*, **45**, 3415 (1972).
4) H. Hosoya, *Fibonacci Quart.*, **11**, 255 (1973).
5) H. Narumi and H. Hosoya, *Bull. Chem. Soc. Jpn.*, **53**, 1228 (1980).
6) I. Gutman, *Croat. Chem. Acta*, **54**, 81 (1981).
7) H. Narumi and H. Hosoya, *Bull. Chem. Soc. Jpn.*, **58**, 1778 (1985).
8) H. Hosoya and A. Motoyama, *J. Math. Phys.*, **26**, 157 (1985).
9) H. Hosoya, *Comput. Math.*, **12B**, 271 (1986).
10) I. Gutman, *Z. Phys. Chem. (Leipzig)*, **267**, 1152 (1986).
11) I. Gutman and O. E. Polansky, "Mathematical Concepts in Organic Chemistry," Springer-Verlag, Berlin (1986).
12) H. Hosoya, K. Hosoi, and I. Gutman, *Theor. Chim. Acta*, **38**, 37 (1975).
13) H. Hosoya and M. Murakami, *Bull. Chem. Soc. Jpn.*, **48**, 3512 (1975).
14) H. Hosoya and K. Hosoi, *J. Chem. Phys.*, **64**, 1065 (1976).
15) I. Gutman, *Publ. Inst. Math. (Beograd)*, **22**, 63 (1977).
16) I. Gutman and H. Hosoya, *Theor. Chim. Acta*, **48**, 279 (1978).
17) C. D. Godsil and I. Gutman, *J. Graph Theory*, **5**, 137 (1981).
18) J. Aihara, *J. Am. Chem. Soc.*, **98**, 2750 (1976).
19) I. Gutman, M. Milun, and N. Trinajstić, *J. Am. Chem. Soc.*, **99**, 1692 (1977).
20) J. Aihara, *J. Org. Chem.*, **41**, 2488 (1976).
21) I. Gutman, T. Yamaguchi, and H. Hosoya, *Bull. Chem. Soc. Jpn.*, **49**, 1811 (1976).
22) I. Gutman, *Croat. Chem. Acta*, **46**, 209 (1974).
23) I. Gutman and J. Cioslowski, *Z. Naturforsch.*, **42a**, 438 (1987).
24) R. R. Dreisbach, "Physical Properties of Chemical Compounds," Am. Chem. Soc., Washington (1959), Vol. I.
25) T. Yamaguchi, M. Suzuki, and H. Hosoya, *Nat. Sci. Rept. Ochanomizu Univ.*, **26**, 39 (1975).