match no. 23 pp. 95-103 1988

ON THE HOSOYA INDEX OF VERY LARGE MOLECULES

Ivan Gutman

Faculty of Science, University of Kragujevac, P.O.Box 60,
YU-34000 Kragujevac, Yugoslavia

(Received: January 1988)

Abstract

It is shown that in the case of large molecules the Hosoya index is an exponential function of the number of edges of the molecular graph. This result holds irrespective of any other structural detail of the molecule considered. Approximate formulas for the Hosoya index can be tested on the basis of the present findings. Only the approximate formula from Ref. 5 has the required asymptotic behaviour. A mistake in Ref. 5 is corrected.

INTRODUCTION

The Hosoya index Z(G) of a (molecular) graph G is equal to the total number of matchings of G. It has been introduced in 1971 by Haruo Hosoya[1] and eventually found various physicochemical applications. For a detailed exposition of the theory of the Hosoya index see the book[2].

It is a well-known fact[3] that the Hosoya index rapidly increases with the increasing size of the respective graph (i.e. with the increasing number of vertices and/or edges). However, practically nothing is known about the dependence of Z on the structure of the graph when its size is very large (or more precisely: when its size tends to infinity). The aim of the present paper is to fil this gap.

The calculation of the Hosoya index is not easy[4] and is usually performed by means of pertinent recurrence relations[1,2]. Therefore the evaluation of Z(G) for large polycyclic graphs is a fairly difficulat task and exact Z values are known only in a rather limited number of cases[3]. In order to overcome this difficulty, approximate formulas for Z(G) have been designed[5-7], which make the (approximate) calculation of Z of large molecular graphs quite easy. However, the quality of these approximations could be tested only if exact Z values for large molecular graphs were available.

THE MAIN RESULTS

In the present paper we show that the Hosoya index of a molecular graph G increases as an exponential function of m(G), the number of edges of G. In particular, we prove the following two results.

Denote by m = m(G) and D = D(G) the number of edges and the maximal degree of a vertex, respectively, of the graph G.

Theorem 1. If G is a connected graph with more than two vertices, then

$$(4D-4)^{-1} \ln(4D-3)\, m < \ln Z < m . \qquad (1)$$

Let $G_1, G_2, \ldots, G_k, \ldots$ be an infinite series of connected graphs, such that G_{k+1} has more vertices and more edges than the graph G_k.

Theorem 2. If $D(G_k)$ is independent of k (and is equal to D), then for $k \to \infty$ the limit L

$$L = \lim\, m(G_k)^{-1} \ln Z(G_k)$$

exists and

$$(4D-4)^{-1} \ln(4D-3) < L < 1 .$$

This means that for large values of k, $Z(G_k)$ asymptotically behaves as $\exp[L\, m(G_k)]$, i.e. the Hosoya index is an exponential function of the number of edges.

PROOFS

First of all, one should note that Theorem 2 is an immediate consequences of Theorem 1.

In order to prove Theorem 1 we introduce the following notation[2]. The graph G has n vertices, m edges and the maxi-

mal degree of a vertex of G is D. The zeros of the matching polynomial[8-10] of G are $x_1, x_2, \ldots, x_n$ and they are labeled so that $x_1 \geq x_2 \geq \ldots \geq x_n$.

Then according to the well-known property[8] $x_i + x_{n-i+1} = 0$, we have $x_1^2 \geq x_i^2$ for all i=2,3,...,n.

We further recall the identity[6,7,10]

$$\sum_{i=1}^{n} x_i^2 = 2\,m\,. \qquad (2)$$

In Ref. 6 has been shown that the Hosoya index is related to the zeros of the matching polynomials in the following manner:

$$\ln Z = 0.5 \sum_{i=1}^{n} \ln(1 + x_i^2)\,. \qquad (3)$$

In addition to Eqs. (2) and (3) we shall need an auxiliary result, which we formulate as

Lemma 1. Let a be a positive constant and $t = a^{-1} \ln(1+a)$. Then the inequalities

$$tx < \ln(1 + x) < x \qquad (4)$$

hold for all values of x from the interval (0,a).

Proof. From the series

$$\exp(x) = 1 + x + x^2/2! + x^3/3! + \ldots$$

is evident that $\exp(x) < 1 + x$ for all positive values of x. Hence the right-hand side of (4) follows.

Consider now the function

$$f(x) = \exp(sx) - (1 + x)$$

where s is a parameter, $0 < s < 1$. Let b be the (unique) non-zero solution of the equaton $f(x) = 0$. Then

$$f(x) < 0 \quad \text{for} \quad 0 < x < b \, . \tag{5}$$

In order to see this, observe that $f(0) = 0$ and $df(0)/dx = s - 1 < 0$. Whence for positive, near-zero values of x, f(x) is a decreasing function, i.e. its values are less than f(0). Consequently, f(x) must be negative. Since f(x) changes sign at $x = b$, the inequality (5) follows.

The quantities b and s are related as $s = b^{-1} \ln(1 + b)$. Bearing in mind that for all positive values of b, the value of $b^{-1} \ln(1 + b)$ is less than unity, we arrive at the left-hand side of (4) by setting $t = s$ and $a = b$.

Lemma 1 is thus proved.

Let X be an upper bound of x_1, i.e. $x_1 < X$. Then Lemma 1 implies that for $t = X^{-2} \ln(1 + X^2)$,

$$t\, x_i^2 < \ln(1 + x_i^2) < x_i^2$$

holds for $i=1,2,\ldots,n$ i.e. for all zeros of the matching polynomial of the graph G. Combining the above inequalities with the identities (2) and (3) we arrive at

$$t\, m < \ln Z < m . \qquad (6)$$

The statement of Theorem 1 is now obtained from the result[11,12] that $X = 2(D - 1)^{1/2}$ is an upper bound for the greatest zero of the matching polynomial of a connected graph with more than two vertices. Then (1) follows from (6).

DISCUSSION

It is worth noting that the inequality $x_1 < 2(D - 1)^{1/2}$ is strict for all graphs. Nevertheless, for certain graphs, the difference $x_1 - 2(D - 1)^{1/2}$ can become arbitrarily small[12]. Thus $2(D - 1)^{1/2}$ is the best possible upper bound for x_1, depending (solely) on the maximal vertex degree.

In molecular graphs, only three differernt values of D can occur, namely 2, 3 and 4. For D = 2, 3 and 4, $(4D - 4)^{-1} \ln(4D - 3)$ is equal to 0.402, 0.275 and 0.214, respectively.

In the case of bipartite graphs (molecular graphs of alternant hydrocarbons), the upper bound for x_1 can be improved[12].

Let G be a bipartite graph and its vertices divided into two classes, such that adjacent vertices are never in the same class. Let D_1 and D_2 denote the maximal degree of a vertex from the first and second class, respectively. Let

$$F = (D_1 - 1)^{1/2} + (D_2 - 1)^{1/2} .$$

Then[12] F is an upper bound for x_1. On the basis of this result one can improve Theorem 1 in the following manner.

Theorem 3. If G is a connected bipartite graph, then

$$F^{-2} \ln(1 + F^2)\, m < \ln Z < m .$$

Let $Z^*(G)$ be an approximate topological formula for the Hosoya index Z(G). From Theorem 2 we see that Z^* must meet the requirement that when the size of the graph G is sufficiently large, then $\ln Z^*(G)$ is linearly dependent on m(G). Among the approximate formulas for Z, proposed recently by the present author and his coworkers[5-7], only the formula from Ref. 5 has the necessary linear dependence on m. Hence, although the formulas from Refs. 6 and 7 give quite reliable results for medium-sized molecular graphs, their applicability to very large graphs is now seen to be unjustified.

Concerning the approximate formula $Z^*(G)$, put forward in Ref. 5, a certain mistake must be pointed out.

Let $G_1, G_2, \ldots, G_k, \ldots$ be the same infinite sequence of graphs as considered in connection with Theorem 2. The expression for the asymptotic behaviour of $Z^*(G_k)$ (the formula after Eq. (10) in Ref. 5) is incorrect and should read:

$$Z^*(G_k) \approx\approx t^{-t/2}\, n(G_k)^{t/2}\, W^t\, (1 + q/r)^{h(G_k)}$$

where

$$W = (p - q^2/r - q)^{1/2}/(1 + q/r)$$

and where t is an empirical parameter whose numerical value has been determined in Ref. 5. Further,

$$p = \lim d(G_k)/n(G_k)$$

$$q = \lim m(G_k)/n(G_k)$$

$$r = \lim h(G_k)/n(G_k) \ .$$

The above three limits are for $k \to \infty$. In accordance with the notation of Ref. 5, d(G) and h(G) denote the sum of the squares of the vertex degrees and the size of the maximal matching, respectively. For molecular graphs, p, q and r are necessarily finite and non-zero.

Since for large values of k,

$$h(G_k) \approx (r/q)\ m(G_k)$$

we see that

$$\ln Z^*(G_k) \approx (r/q)\ \ln(1 + q/r)\ m(G_k)$$

in agreement with the results of the present paper.

REFERENCES

1. H.Hosoya, Bull.Chem.Soc.Japan 44, 2332 (1971).
2. I.Gutman and O.E.Polansky, Mathematical Concepts in Organic Chemistry, Springer-Verlag, Berlin 1986.
3. For examples of Hosoya indices of large graphs see: H.Hosoya and A.Motoyama, J.Math.Phys. 26, 157 (1985); H.Hosoya, Computers & Math. 12B, 271 (1986).
4. The calculation of the Hosoya index of a molecular graph is an NP-complete problem. This is an immediate cosequence of the results obtained in: L.G.Valiant, SIAM J.Comput. 8, 410 (1979).
5. I.Gutman and Z.Marković, Bull.Chem.Soc.Japan 60, 2611 (1987).
6. I.Gutman, Z.Marković and S.Marković, Chem.Phys.Letters 134, 139 (1987).
7. I.Gutman, V.R.Rosenfeld and Z.Marković, J.Serb.Chem.Soc. 52, 139 (1987).
8. I.Gutman, Match 6, 75 (1979).
9. C.D.Godsil and I.Gutman, J.Graph Theory 5, 137 (1981).
10. D.Cvetković, M.Doob, I.Gutman and A.Torgašev, Recent Results in the Theory of Graph Spectra, Elsevier, Amsterdam 1988.
11. O.J.Heilmann and E.H.Lieb, Commun.Math.Phys. 25, 190 (1972).
12. I.Gutman, Ph.D.Thesis (Mathematics), Univ.Belgrade 1980.