match no. 28 pp. 235-247 1992

Topological Formulas and Upper/Lower Bounds in Chemical Polygonal Graphs, Particularly in Benzenoid Polyhexes

Tetsuo Morikawa* and Alexandru T. Balaban**
*Chemistry Department, Joetsu University of Education, Joetsu, Japan
**Organic Chemistry Department, The Polytechnic, Bucharest, Roumania

(received: January 1991)

(Abstract) A topological formula is presented which states that the summation of the numbers of vertices of polygons in a polygonal (planar) graph is equal to adding together the number of divalent vertices on the periphery, twice the number of divalent vertices in the interior, twice the number of trivalent vertices on the periphery, and treble the number of trivalent vertices in the interior. From this formula and the cyclomatic number in graph theory, a large number of topological formulas are derived. Using some of them we can choose pairs of molecular variables which classify hydrocarbon molecular formulas; the upper/lower bounds for the pairs complete the mathematical control of the classification.

Glossary of Symbols

In a chemical polygonal graph G:

T	vertex having degree 3 (trivalent)
D	vertex having degree 2 (divalent)
n_T	number of T's $(= n_{Tp} + n_{Ti})$
n_D	number of D's $(= n_{Dp} + n_{Di})$
n_{TD}	total number of vertices $(= n_T + n_D)$
n_{Tp}	number of peripheral T's

n_{Ti} number of internal T's
n_{Dp} number of peripheral D's
n_{Di} number of internal D's
n_G number of components of G
n_e total number of edges
n_{fj} number of polygonal faces composed of j vertices
n_{f6} number of hexagonal faces
n_h total number of holes
μ_G cyclomatic number for G
μ_i newly defined number for the internal graph of G
d() increment in the variable ()
k integer parameter
In a hydrocarbon molecular formula:
n_C number of carbon atoms
n_H number of hydrogen atoms
n_R number of D's unattached by hydrogens (radical sites) on the periphery

(The detailed definitions of symbols will be expalined below.)

Introduction

Polygonal graphs in chemistry are formed from tetragons, pentagons, hexagons, and some other polygons. Such a chemical polygonal system represents, for example, a polycyclic hydrocarbon (hydrogen-depleted) structure. Several topological formulas in connection with polygonal graphs are known in the chemical literature[1-6], but some of them and relationships between them are not rigorously founded on mathematics. We prove that a topological formula for polygonal graphs relates the number of polygonal faces to the number of vertices, and show that a combination of this formula and the cyclomatic number in graph theory yields a large number of topological and/or graph-theoretical formulas. A particular case, where graphs are connected and faces are all hexagonal, is discussed in detail.

In the isomer enumeration of a given class of molecular formulas the first step is to separate it so that each different part has similar characteristics. Dias[1] adopted two kinds of topological variable, the number of trivalent vertices in the interior and the number of disconnections, in order to classify hydrocarbon molecular formulas; however, there is no mathematical basis for his procedure. We describe how an inequality relation introduced by Harary and Harborth[7] improves the mathematical background to Dias's tables. A table is presented, where an alternative pair of topological variables classifies hydrocarbon molecular formulas.

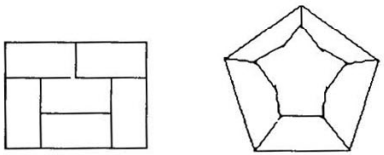

Fig. 1 Examples of Polygonal Graphs (Isomorphic)

Chemical Polygonal Formulas

Let us assume that a polygonal (planar) graph G is constructed by the supplement of one polygonal face composed of j vertices to another polygonal (planar) graph G', and that the j-gonal face is formed by the attachment of p paths to G'. The j-gonal face is composed of three kinds of parts: (i) The a vertices having degree 2 that constitute the p paths, where the endpoints are not counted. (ii) The b vertices having degree 3 on the periphery of G'. (iii) The 2p + c vertices having degree 2 of G',

where none of the c vertices are connected with the p paths.

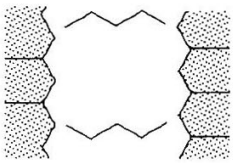

Fig. 2 Attachment of p paths

By the attachment of the p paths, the decrease in the number of peripheral vertices having degree 2 occurs in only (iii), and the c vertices of (iii) become internal; hence, $d(n_{Dp}) = a - (2p + c)$, and $d(n_{Di}) = c$. After the formation of the j-gonal face, every vertex of (ii) becomes internal, and there is no change in internal T's other than (ii); namely, $d(n_{Ti}) = b$. The number of the peripheral T's increases in (iii) , and decreases in (ii); namely, $d(n_{Tp}) = 2p - b$. Clearly $d(n_{fj}) = 1$ in the j-gon formation. Hence,

$$d(n_{Dp}) + 2d(n_{Di}) + 2d(n_{Tp}) + 3d(n_{Ti}) - jd(n_{fj}) = a + b + c + 2p - j = 0$$

which means that the value, $n_{Dp} + 2n_{Di} + 2n_{Tp} + 3n_{Ti} - j \times n_{fj}$, is a constant, i.e., an invariant in the j-gon formation. The constant becomes zero, even though any kind of polygon ($j \geq 3$) is chosen as the starting (initial) molecule that consists of only D's; e.g., $6 + 2\times0 + 2\times0 + 3\times0 - 6\times1 = 0$ for benzene.

Summing all the j-gonal faces for n_G components of a polygonal graph G up yields the same constant zero. Therefore we conclude that

$$n_{Dp} + 2n_{Di} + 2n_{Tp} + 3n_{Ti} = \sum_{j\ (\geq 3)} j \times n_{fj} \qquad (n_G \geq 1,\ n_h \geq 0) \qquad (1)$$

This formula means that: On the one hand, D and T on the periphery belong to one and two polygonal faces, respectively, and on the other hand, D and T in the interior are shared by two and three polygonal faces, respectively; hence, the summation of the numbers of vertices of polygonal faces is to count peripheral D's one time, internal D's two times, peripheral T's two times, and internal T's three times. It should be noted that, generally speaking, even if two polygonal graphs are topologically isomorphic, the total numbers of Eq. (1) for them are not equal; e.g., Eq. (1) in Fig. 1 gives $4 + 2\times0 + 2\times5 + 3\times5 = 1\times4 + 5\times5 = 29$ on the left-hand side, and $0 + 2\times4 + 2\times5 + 3\times5 = 1\times4 + 4\times5 + 1\times9 = 33$ on the right-hand side. Note also that the term "peripheral" means both "on the periphery of holes" and "on the external perimeter"; and that the summation in the right-hand side involves no hole and no infinite face.

This paper is based on the assumption that only two kinds of vertices, T and D, constitute a polygonal graph G; hence, the cyclomatic number μ_G for G is of the form

$$2\mu_G = 3n_T + 2n_D - 2(n_T + n_D - n_G) = n_T + 2n_G$$

Whereas μ_G for polygonal graphs has the graph-theoretical meaning that

$$\mu_G = \sum_j n_{fj} + n_h$$

Therefore, we obtain

$$n_T = n_{Tp} + n_{Ti} = 2\left(\sum_j n_{fj} + n_h - n_G\right) \qquad (2)$$

It is well-known that the sum of degrees for G is equal to twice the total

number of edges:

$$2n_D + 3n_T = 2n_e \qquad (3)$$

Equations (1), (2) and/or (3) join together to yield a number of topological and/or graph-theoretical formulas in the chemical literature[1-6], as will be shown below. The present paper describes only one combination of Eqs. (1) and (2).

Subtracting Eq. (2) from Eq. (1) leads to

$$n_{Dp} + n_{Tp} = -2(n_{Di} + n_{Ti}) + 2(n_G - n_h) + \Sigma(j-2)n_{fj}$$

The total number of vertices on the periphery is, therefore, expressed as

$$n_{Dp} + n_{Tp} \equiv 2(n_G - n_h) + \Sigma(j-2)n_{fj} \pmod 4 \quad (n_G \geq 1,\ n_h \geq 0,\ \text{even } n_{Di} + n_{Ti}) \qquad (4)$$

$$n_{Dp} + n_{Tp} \equiv 2(n_G - n_h - 1) + \Sigma(j-2)n_{fj} \pmod 4 \quad (n_G \geq 1,\ n_h \geq 0,\ \text{odd } n_{Di} + n_{Ti}) \qquad (5)$$

There exist cycles in polygonal graphs; each cycle is composed of polygons, so that the length of the outer cycle can be estimated by $n_{Dp} + n_{Tp}$, and the number of vertices in the interior of that cycle can be counted by $n_{Di} + n_{Ti}$. Hence we can say that the length of any cycle of polygonal graphs is given by either (4) or (5).

Taking away treble the equation (2) from Eq. (1), we have

$$n_{Dp} + 2n_{Di} - n_{Tp} = 6(n_G - n_h) + \Sigma\,(j - 6)n_{fj} \qquad (n_G \geq 1,\ n_h \geq 0) \qquad (6)$$

A particular case for this equation, where a given polygonal graph is connected ($n_G = 1$), $n_{Di} = 0$, and $4 \leq j \leq 8$, has been suggested by Dias[1a] who

got it experimentally.

The internal graph of a polygonal graph, where all the peripheral edges of that polygonal graph are excluded, can be regarded as being a graph that is composed of n_{Ti} vertices with degree 3, n_{Di} vertices with degree 2, and n_{Tp} vertices with degree 1; then a new number μ_i for the internal graph is calculated by

$$2\mu_i = 3n_{Ti} + 2n_{Di} + n_{Tp} - 2(n_{Ti} + n_{Di} + n_{Tp} - n_G) = n_{Ti} - n_{Tp} + 2n_G$$

Then,

$$n_{Dp} + 2n_{Di} + \mu_i = 6n_G - 5n_h + \Sigma(j - 5)n_{fj} \qquad (n_G \geq 1,\ n_h \geq 0) \qquad (7)$$

Formulas for Polyhex Graphs

This chapter deals with connected hexagonal graphs (polyhexes) which correspond to polycyclic hydrocarbons. By putting $n_{Di} = 0$ and $j = 6$ into Eqs. (1) and (2) we deduce two basic formulas:

$$n_{Dp} + 2n_{Tp} + 3n_{Ti} = 6n_{f6} \qquad (n_G \geq 1,\ n_h \geq 0,\ \text{polyhex}) \qquad (8)$$

$$n_T = n_{Tp} + n_{Ti} = 2(n_{f6} + n_h - n_G) \qquad (n_G \geq 1,\ n_h \geq 0,\ \text{polyhex}) \qquad (9)$$

Table 1 and Fig. 3 show the attachment of p paths for polyhexes. A hexagonal face is composed of six edges, so that only ten cases occur; Cases I to V are obtained by the attachment of one path, Cases VI to IX by the attachment of two paths, Case X by the attachment of three paths. In every case clearly the increment $d(n_{f6})$ equals 1.

Equations (8) and (9) lead, for example, to

$$n_{Dp} - n_{Tp} = 6(n_G - n_h) \qquad (n_G \geq 1,\ n_h \geq 0,\ \text{polyhex}) \qquad (10)$$

$$n_{Dp} + n_{Tp} \equiv 2(n_G - n_h) \pmod 4 \qquad (n_G \geq 1,\ n_h \geq 0,\ \text{even } n_{Ti},\ \text{polyhex}) \qquad (11)$$

$$n_{Dp} + n_{Tp} \equiv 2(n_G - n_h - 1) \pmod 4 \qquad (n_G \geq 1,\ n_h \geq 0,\ \text{odd } n_{Ti},\ \text{polyhex}) \qquad (12)$$

$$n_{Dp} + n_{Ti} = 4(n_G - n_h) + 2n_{f6} \qquad (n_G \geq 1,\ n_h \geq 0,\ \text{polyhex}) \qquad (13)$$

$$n_{Dp} + \mu_i = 6n_G - 5n_h + n_{f6} \qquad (n_G \geq 1,\ n_h \geq 0,\ \text{polyhex}) \qquad (14)$$

Equation (8) means that the odd-even parity of n_{Ti} coincides with that of n_{Dp}. Equations (11) and (12) show that the total number of vertices on the periphery is congruent to either $2(n_G - n_h)$ for even n_{Ti} or $2(n_G - n_h - 1)$ for odd n_{Ti} with respect to the modulus 4:

$$n_{Dp} + n_{Tp} = 4k + 2(n_G - n_h) \qquad (n_G \geq 1,\ n_h \geq 0,\ \text{even } n_{Ti},\ \text{polyhex}) \qquad (15)$$

$$n_{Dp} + n_{Tp} = 4k + 2(n_G - n_h - 1) \qquad (n_G \geq 1,\ n_h \geq 0,\ \text{odd } n_{Ti},\ \text{polyhex}) \qquad (16)$$

Some of the particular forms ($n_G = 1$, $n_h = 0$ or $n_h = 1$) for Eqs. (8) to (16) appear in References 1 - 6.

Table 1. The Increments of variables in one-hexagon formation

Case No.	$d(n_{Ti})$	$d(n_{Tp})$	$d(n_{Dp})$	$d(n_{f6})$	$d(n_{TD})$	$d(n_e)$
I	0	2	2	1	4	5
II	1	1	1	1	3	4
III	2	0	0	1	2	3
IV	3	-1	-1	1	1	2
V	4	-2	-2	1	0	1
VI	2	2	-4	1	0	2
VII	1	3	-3	1	1	3
VIII	0	4	-2	1	2	4
IX	0	4	-2	1	2	4
X	0	6	-6	1	0	3

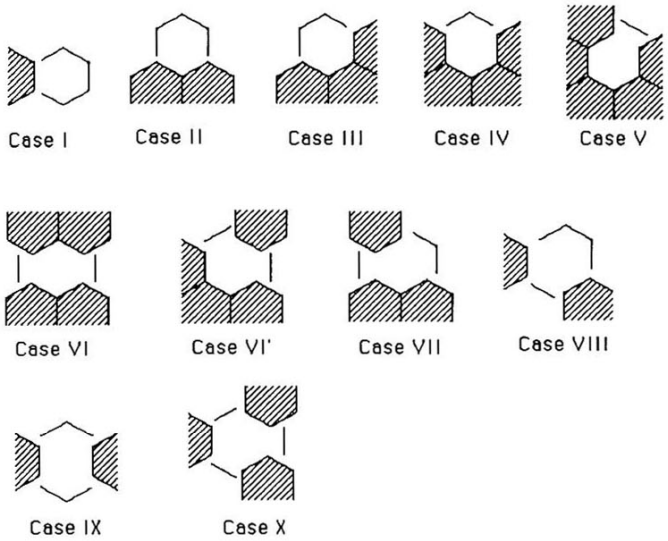

Fig. 3 One-hexagon formation by the attachment of p paths

We consider topological formulas on the periphery of holes in a given polyhex ($n_G = 1$). To extend the area of a hole is to remove paths on the periphery of that hole; such an operation is no other than the inverse process of one of Cases I to V; namely, the increment of each topological variable in the hole formation is expressed by each row of Cases I to V that is multiplied by -1. Then it can be derived from Cases I to V that $d(n_{Dp} - n_{Tp}) = 0$ on the periphery of a hole; and if only two hexagons constitute the starting hole, then $n_{Dp}' - n_{Tp}' = 2 - 8 = -6$. Therefore, the summation of all over the holes gives:

$n_{Dp}' - n_{Tp}' = -6n_h$ (on the periphery of holes, $n_G = 1$, $n_h \geq 1$, polyhex) (17)

Subtracting Eq. (17) from Eq. (10), and combining the result with the particular case ($n_h = 0$) of Eq. (10) we obtain

$n_{Dp}'' - n_{Tp}'' = 6$ (on the external perimeter, $n_G = 1$, $n_h \geq 0$, polyhex) (18)

The addition of Eqs. (17) and (18) leads to Eq. (10) because $n_{Dp} = n_{Dp}' + n_{Dp}''$ and $n_{Tp} = n_{Tp}' + n_{Tp}''$.

Choice of One Pair of Topological Variables for Polyhex Hydrocarbons

The previous chapter shows that there are a number of topological (and/or graph-theoretical) variables such as n_{Ti}, μ_i, k and n_{f6}, for polyhexes. We can choose one pair of them for the representation of hydrocarbon molecular formulas ($n_G = 1$) if the pair is linearly independent of n_C and n_H. In the enumeration of hydrocarbon molecular graphs, not all D's on the periphery are attached by hydrogen atoms; n_R stands for the number:

$$n_C = n_T + n_D = n_T + n_{Dp} \quad \text{and} \quad n_H + n_R = n_{Dp} \tag{19}$$

Dias[1] has used the pair of n_{Ti} (rows) and μ_i (columns) for the classification of benzenoid hydrocarbon molecular formulas; d_s in his notation is equivalent to $-\mu_i$. Using Eqs. (9), (10), (13) and (19) we obtain

$$n_{Ti} = n_C - 2(n_H + n_R) + 6(1 - n_h) \tag{20}$$

$$2\mu_i = n_C - 3(n_H + n_R) + 12(1 - n_h) + 2 \tag{21}$$

In other words, the pair of n_{Ti} and μ_i is linearly independent of the pair of n_C and n_H if and only if n_h and n_R are given; Dias's table assumes that in

polyhex graphs, $n_h = 0$ and $n_R = 0$.

A similar treatment for the choice of pairs of topological variables is possible by use of equations described in the previous chapter.

Upper and Lower Bounds for Topological Variables

In this chapter we deal with only polyhex graphs having no hole ($n_G = 1$, $n_h = 0$, polyhex). Not all the values for a given pair of topological variables are possible; some restrictions must be imposed on the pair. Harary and Harborth[7] proved the upper and lower limits for the total number n_{TD} of vertices:

$$2n_{f6} + 1 + \lceil (12n_{f6} - 3)^{1/2} \rceil \leq n_{TD} \leq 4n_{f6} + 2 \qquad \text{(H\&H)}$$

where the ceiling function $\lceil x \rceil$ for a real number x is defined in Reference 8 to be the smallest integer that is greater than or equal to x. Inserting $n_{TD} = n_T + n_D$ and $2n_{f6} = n_T + 2$ (Eq. (9)) into Eq. (H&H), we obtain the upper and lower bounds of n_D ($= n_{Dp}$) in terms of n_T:

$$3 + \lceil (6n_T + 9)^{1/2} \rceil \leq n_D \leq n_T + 6 \qquad (22)$$

By means of Eqs. (22) and (19) the number of hydrocarbon molecular formulas for given n_T can be calculated.

A similar treatment for topological variables is possible:

$$n_{Ti} \geq 2\mu_i + 1 + \lceil (12\mu_i - 3)^{1/2} \rceil \quad \text{for } \mu_i > 0, \text{ and } n_{Ti} \geq 0 \quad \text{for } \mu_i \leq 0 \qquad (23)$$

$$\lceil \{(12n_{f6} - 3)^{1/2} - 1\}/2 \rceil \leq k \leq n_{f6} \qquad (24)$$

$$0 \leq n_{Ti} \leq 2n_{f6} + 1 - \lceil (12n_{f6} - 3)^{1/2} \rceil \qquad (25)$$

Equation (23) makes it possible to prepare Dias's Table of Reference 1 under rigorous mathematical control. An alternative classification of hydrocarbon molecular formulas ($n_h = 0$ and $n_R = 0$) by use of Eq. (25) is presented in Table 2, where the row for $n_{Ti} = 0$ corresponds to catafusens.

Table 2. Classification of hydrocarbon molecular formulas by Eq. (25)

$n_{Ti} \backslash n_{f6}$	1	2	3	4	5	6	7	8	...
0	C_6H_6	$C_{10}H_8$	$C_{14}H_{10}$	$C_{18}H_{12}$	$C_{22}H_{14}$	$C_{26}H_{16}$	$C_{30}H_{18}$	$C_{34}H_{20}$	...
2				$C_{16}H_{10}$	$C_{20}H_{12}$	$C_{24}H_{14}$	$C_{28}H_{16}$	$C_{32}H_{18}$	...
4						$C_{22}H_{12}$	$C_{26}H_{14}$	$C_{30}H_{16}$	...
6							$C_{24}H_{12}$	$C_{28}H_{14}$	...
...									...

The authors express thanks sincerely to the referees who suggested to us several references cited hereafter and several comments.

References

1. J. R. Dias, (a) J. Chem. Inf. Comput. Sci. 22, 139 (1982); (b) J. Chem. Inf. Comput. Sci. 22, 15 (1982); (c) MATCH 14, 83 (1983); (d) J. Chem. Inf. Comput. Sci. 24, 124 (1984).
2. I. Gutman, Croat. Chim. Acta 46, 209 (1974).
3. D. Cvetkovic, I. Gutman, and N. Trinajstic, J. Chem. Phys. 61, 2700 (1974).
4. O. E. Polansky, and D. H. Rouvray, MATCH 2, 63 (1976).
5. G. G. Hall, Theor. Chim. Acta 73, 425 (1988).
6. S. J. Cyvin, and J. Brunvoll, Chem. Phys. Letters 164, 635 (1989).
7. F. Harary, and H. Harborth, J. Comb. Inf. System Sci. 1, 1 (1976).

8. R. L. Graham, D. E. Knuth, and O. Patashnik, Concrete Mathematics -- A Foundation for Computer Science, Addison-Wesley 1989, Chap. 3.

Note added: For recent results on Dias's classification, see J. R. Dias, Handbook of Polycyclic Hydrocarbons, Part A, Elsevier, Amsterdam 1987; J. Chem. Inf. Compt. Sci. 30, 61 (1990); Topics in Current Chemistry, Vol. 153, Advances in the Theory of Benzenoid Hydrocarbons, edited by I. Gutman and S. J. Cyvin, Springer, Berlin 1990. Also see J. Brunvoll, and S. J. Cyvin, Z. Naturforsch. 45A, 69 (1990); Chem.Phys. Letters 170, 364 (1990).