Coulson function and Hosoya index: extension of the relationship to polycyclic graphs and to new types of matching polynomials

Gordon G. Cash[*]

Office of Pollution Prevention and Toxics, Risk Assessment Division (7403M),
U. S. Environmental Protection Agency, 1200 Pennsylvania Av., N. W., Washington, DC 20460, USA
E-mail: cash.gordon@epa.gov

Gutman et al. [Chem. Phys. Lett. 355 (2002) 378–382] established a relationship between the Coulson function, $F(G, x) = n - [ix\phi'(G, ix)/\phi(G, ix)]$, where ϕ is the characteristic polynomial, and the Hosoya index Z, which is the sum over all k of the counts of all k-matchings. Like the original Coulson function, this relationship was postulated only for trees. The present study shows that the relationship can be extended to (poly)cyclic graphs by substituting the matching, or acyclic, polynomial for the characteristic polynomial. In addition, the relationship is extended to new types of matching polynomials that match cycles larger than edges (2-cyc1es). Finally, this presentation demonstrates a rigorous mathematical relationship between the graph adjacency matrices and the coefficients of these polynomials and describes computational algorithms for calculating them.

KEY WORDS: coulson function, Hosoya index

1. Introduction

In 1940, Coulson [1] introduced an expression for the total HMO π-electron energy, E, of a conjugated system

$$E = \frac{1}{\pi} \int_{-\infty}^{+\infty} F(x) \mathrm{d}x, \tag{1}$$

where

$$F(x) = F(G, x) = n - \frac{ix\phi'(G, ix)}{\phi(G, ix)}. \tag{2}$$

[*] Corresponding author.

In equation (1), $\phi(G, ix)$ is the ordinary characteristic polynomial of the adjacency matrix of the graph G, evaluated at ix, where $i = \sqrt{-1}$, $\phi'(G, ix)$ is the first derivative of the characteristic polynomial, also evaluated at ix, and n is the number of vertices in the graph.

Some 30 years later, Hosoya [2] introduced the counting polynomial for a graph G,

$$Q(G, x) = \sum_{k=0}^{\lfloor n/2 \rfloor} p(G, k)x^k, \tag{3}$$

where n is as above, $\lfloor n/2 \rfloor$ is the largest integer not greater than $n/2$, and the $p(G, k)$ are the number of ways of choosing k disjoint edges from G. Associated with this polynomial is the index, Z, that now bears Hosoya's name:

$$Z(G) = \sum_{k=0}^{\lfloor n/2 \rfloor} p(G, k). \tag{4}$$

The coefficients of the polynomial in equation (3) have the same absolute values as those in the much-studied acyclic, or matching polynomial, $\alpha(G, x)$:

$$\alpha(G, x) = \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k p(G, k)x^{n-2k}. \tag{5}$$

Formulas for interconverting these two polynomials were recently published [3].

The above equations have been in the literature for decades, but it was only very recently that Gutman et al. [4] established mathematical connections between $F(G, x)$ and $Z(G)$. That reference called $F(G, x)$ the *Coulson function* and defined the *Hosoya point, x_H,* as that value of x for which $F(G, x) = \ln Z(G)$. In this paper, Gutman et al. explicitly dealt only with trees, i.e., connected, acyclic graphs. As demonstrated below, however, it is possible to extend their results in several ways, revealing unexpected results. Other results from Gutman et al. [4] relevant to the present study are $F(G, 1.2) \approx \ln Z(G)$ and $\ln Z \approx aF(G, 1) + b$. In this last result, a is almost independent of the number of vertices n, while b is an (approximately) increasing function of n. Also, the values of a and b are somewhat different for all trees than they are for only chemical trees, i.e., trees with no vertex degree > 4.

2.   Extension of x_H to acyclic "reference" graphs

The much-studied acyclic, or matching, polynomial is often held to be the characterstic polynomial of a "reference" graph which is not physically realizable but has the properties of a tree. An obvious extension of the work described above, then, is to redefine the Coulson function in terms of $\alpha(G, x)$

and $\alpha'(G, x)$ in place of $\phi(G, x)$ and $\phi'(G, x)$. For reasons which will be apparent later, we call this function $F_2(G, x)$ and the associated Hosoya point $x_{H,2}$. As noted above, $Z(G)$ is just sum of the coefficients of $\alpha(G, x)$. As expected, "reference" trees for a variety of cyclic structures, e.g., benzenoids, planar non-altemants, 5,6-fullerenes, 4,6-fullerenes, polyhex toroidal fullerenes, gave values for x_H within the range found by Gutman et al. [4] for real trees. Two general observations from this exercise are that average x_H tends to be larger for planar structures than for cages and larger for bipartite graphs than for nonbipartite.

Somewhat more unexpected results arose from the study of $\ln Z \approx a F_2(G, 1) + b$ for cyclic structures. Figure 1 shows a plot of this function for 153 benzenoids of various sizes studied by Herndon et al. [5]. For trees, Gutman et al. [4] found that a and b varied with the number of vertices in the graph, but for this set of benzenoids, an excellent correlation was obtained using structures of all sizes: $\ln Z = 0.9282(0.0015) F_2(G, 1) - 0.2119(0.0225), n = 153, r^2 = 0.9996, F = 365735, s = 0.04$. Figure 1 shows distinct clustering of the data points, with each cluster containing all of the structures with a particular value of n. Benzene itself is at the lower left. Examining a few clusters in detail did not review any correlations at constant n nearly as linear as for the correlation for the entire dataset. In this respect, the reference trees for the set of benzenoids differ significantly from real trees.

Figure 2 shows the results of the same exercise with a set of 45 $C_{30}H_{12}$ indacenoids studied by Fowler and Mitchell [6]. These are planar graphs consisting of two pentagons and eight hexagons. The statistical parameters for this correlation are $\ln Z = 0.4224(0.0102) F_2(G, 1) + 8.8550(0.1764), n = 45, r^2 = 0.9756, F = 1717, s = 0.0027$. While not as good as the benzenoid results, this is a very good correlation. Even cursory inspection of Figure 2, however, reveals clear structure in the dataset, namely, the points lie approximately on three straight lines with nearly identical slopes but different intercepts. The statistical parameters for the three subsets are $\ln Z = 0.3218(0.0052) F_2(G, 1) + 10.6035(0.0906),$
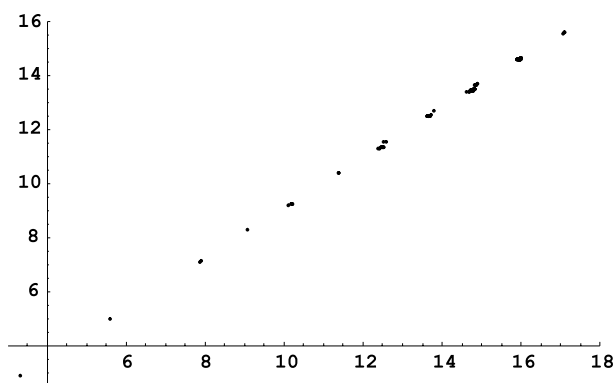

Figure 1. Plot of $\ln Z$ vs. $F_2(G, 1)$ for 153 benzenoids of various sizes.
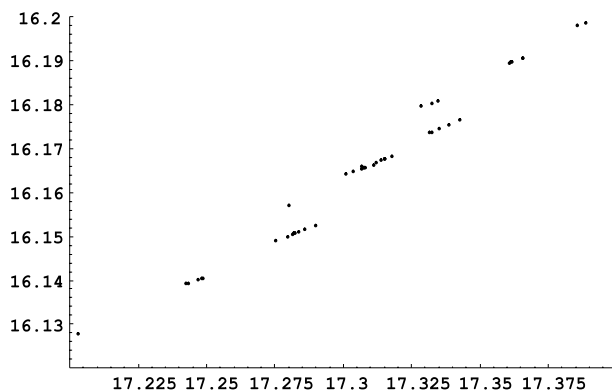
Figure 2. Plot of $\ln Z$ vs. $F_2(G, 1)$ for 45 $C_{30}H_{12}$ indacenoids.

$n = 10$, $r^2 = 0.9979$, $F = 3797$, $s = 0.0003$; $\ln Z = 0.3146(0.0053)F_2(G, 1) + 10.7208$ (0.0917), $n = 21$, $r^2 = 0.9946$, $F = 3527$, $s = 0.0003$; $\ln Z = 0.2902(0.0029)F_2(G, 1) + 11.1356(0.0502)$, $n = 14$, $r^2 = 0.9988$, $F = 9957$, $s = 0.0003$. Placing a straight-edge at various locations on the plot suggests that each of the three subsets is further divided into three smaller subsets, again lying approximately on lines with the same slope but different intercepts. There are, however, not enough data points to make a solid statistical case for this second division. Certainly, the relationship between $\ln Z$ and $F_2(G, 1)$ is much richer in information than might have been anticipated.

It is certainly tempting to speculate that the obvious data structure in Figure 2 reflects some aspects of chemical structure in the underlying graphs. However, when the 45 chemical structures are divided into the three groups represented by the regression lines, there are no obvious structural affinities within any of the groups. Relative position of the two pentagons, perhaps the most evident difference among the 45 structures, seems to be randomly distributed. Clearly, a great deal more work, and with larger datasets, is called for.

3. Extension of x_H to matchings other than edges

The previous section discussed extending the Coulson function and Hosoya point from trees to cyclic structures via the matching polynomial, which enumerates sets of disjoint edges. In some applications of graph theory, it is convenient to consider an edge as a cycle of length 2. A reasonable question, then, is whether the results for the matching polynomial can be extended further to polynomials that enumerate disjoint sets of cycles of some length > 2. (These "cycle-matching" polynomials have appeared in the literature before; their history is discussed in the next section.)

Let us define polynomials

$$\alpha_c(G, x) = \sum_{k=0}^{\lfloor n/c \rfloor} (-1)^k p_c(G, k) x^{n-ck}, \tag{6}$$

where the $p_c(G, k)$ are the number of ways of choosing k disjoint cycles of length c from the graph G. We call the Coulson function derived from these polynomials $F_c(G, x)$ and the sum of the absolute values of the coefficients $Z_c(G)$.

In the case of $c = 2$, equation (6) reduces to equation (5), the ordinary matching polynomial. Successive exponents of these polynomials will differ by c. For purposes of comparison with previous results, it would be better to define an associated set of polynomials $\beta_c(G, y)$, $y \equiv x^{c/2}$, so that successive exponents differ by 2, and another set $\gamma_c(G, z), z \equiv x^c$. Let λ, λ_β, and λ_γ be the roots of the polynomials α_c, β_c, and γ_c, respectively. No information regarding either $Z(G)$ or $F(G, x)$ is lost by using either $\beta_c(G, y)$ or $\gamma_c(G, z)$ in place of $\alpha_c(G, z)$, because

$$Z_c = \prod_{j=1}^{n}(1 + \lambda)^{1/c} = \prod_{j=1}^{2\lfloor n/c \rfloor}(1 + \lambda_\beta^2)^{1/2} = \prod_{j=1}^{\lfloor n/c \rfloor}(1 + \lambda_\gamma). \tag{7}$$

Furthermore, Gutman et al. [4] showed that

$$F(G, x) = \sum_{j=1}^{n} \frac{\lambda_j^2}{x^2 + \lambda_j^2} + i \sum_{j=1}^{n} \frac{x\lambda}{x^2 + \lambda_j^2}, \tag{8}$$

which generalizes to

$$F_c(G, x) = \sum_{\substack{j=1 \\ \lambda \neq 0}}^{\lfloor n/c \rfloor} \sum_{k=1}^{c} \frac{\lambda_\gamma^{2/c} 1^{2k/c}}{(x^2 + \lambda_\gamma^{2/c} 1^{2k/c})} + i \sum_{\substack{j=1 \\ \lambda \neq 0}}^{\lfloor n/c \rfloor} \sum_{k=1}^{c} \frac{\lambda_\gamma^{1/c} 1^{k/c}}{(x^2 + \lambda_\gamma^{2/c} 1^{2k/c})}, \tag{9}$$

where $1^{1/c} = \cos(2\pi/c) + i\sin(2\pi/c)$, the complex cth root of 1. (The corresponding expression in λ_β is rather cumbersome.) When $c = 2$, equations (8) and (9) are the same. This is a straightforward generalization of the pairing theorem, according to which, in the case of $c = 2$, the roots occur in pairs with $\lambda_j = -\lambda_{n-j+1}$. When $c > 2$, the roots occur in groups of c, the individual roots within each group related by $1^{1/c}$.

We now look at the buckminsterfullerene graph, C_{60}, as an example. For $c = 6, \alpha_6(C_{60}, x) = x^{60} - 20 x^{54} + 160 x^{48} - 660 x^{42} + 1510 x^{36} - 1912 x^{30} + 1240 x^{24} - 320 x^{18} + 5 x^{12}$. The polynomial in $y = c/2$ has the same coefficients but different powers of the variable: $\beta_6(C_{60}, y) = y^{20} - 20 y^{18} + 160 y^{16} - 660 y^{14} + 1510 y^{12} - 1912 y^{10} + 1240 y^8 - 320 y^6 + 5 y^4$. The coefficients associated with each power

of the variable, x^k, enumerate the ways of choosing $k/6$ (respectively $k/2$) disjoint hexagons from the C_{60} graph. This is not the same as the sextet polynomial, $\sigma(G, x)$, originally proposed by Hosoya and Yamaguchi [7] because that polynomial counts a selection of hexagons only if the remaining vertices can be paired into edges. For some graphs, including C_{60}, $\sigma(G, y)$ and $\beta_6(G, y)$ have the same relationship as $Q(G, x)$ and $\alpha(G, x)$ above (equations (3) and (5)), but this is not generally true. Gutman [8] discussed conditions under which these polynomials are so related for benzenoids, but no more general cases seem to have been studied. In particular, the relationship holds iff $\sigma(G, y)$ is identical to the *independence polynomial, $\omega(G, y)$*, of the reduced dual, or characteristic tree [9], of the hexagons, about which more below. It would be of value to determine which classes of structures have all sets of disjoint hexagons resonant in the Clar sense.

Another particularly interesting polynomial is $\alpha_5(C_{60}, x) = x^{60} - 12 x^{55} + 66 x^{50} - 220 x^{45} + 495 x^{40} - 792 x^{35} + 924 x^{30} - 792 x^{25} + 495 x^{20} - 220 x^{15} + 66 x^{10} - 12 x^5 + 1$. Perhaps the most obvious feature of $\alpha_5(C_{60}, x)$ is that the absolute values of its coefficients form a row of Pascal's triangle. This is so because all 12 pentagons in C_{60} are independent of each other. Furthermore, $\beta_5(C_{60}, x)$ is identical to $\alpha(G, x)$ (and $\phi(G, x)$) where G consists of 12 disjoint ethylene graphs. It follows that $Z(12C_2) = Z_6(C_{60})$ and $x_H(12C_2) = y_{H,6}(C_{60})$ ($y_{H,c}$ is the Hosoya point derived from the $\beta_c(G, y)$ polynomial). Note also that $x_H(12C_2) = x_H(C_2) = 1.3731$. This result, combined with results for larger trees from Gutman et al. [4], leads to the following:

Conjecture: The value of x_H for the ethylene graph is the maximum value of x_H for any Hosoya point derived from a polynomial whose successive exponents differ by 2.

This value for $y_{H,6}(G)$ will accrue to all isolated-pentagon fullerenes. That much prior research identifies these as the most stable isomers suggests $y_{H,6}(G)$ as a possible predictor of fullerene stability.

The summations in equations (3–6) start with $k = 0$. An empty matching, i.e., no way of selecting a subgraph, has $\alpha_c = x^n$ and $Z_c = 1$. This is relevant to the present discussion because it is mathematically possible, if not physically reasonable, to have a cycle-matching polynomial with the cycle size larger than the entire graph, i.e., $c > n$. It turns out, however, that $F_{c>n}(G, x) = 0$ for all x, and $\ln Z_{c>n} = 0$, so that $y_{H,c>n}$ is undefined. In other words, the Hosoya point is mathematically defined only when the corresponding subgraph makes physical sense. We consider this further evidence that the Hosoya point has a meaningful relationship to structure and is not merely a mathematical curiosity. A related observation is the statement of Gutman et al. [4] that the Hosoya point exists only when $F(G, 0) > \ln Z$. This is true if the Hosoya point is taken to be real-valued, but it has an imaginary value when $F(G, 0) < \ln Z$. It is an open problem whether any graph exists for which the Hosoya point has a complex value.

4. Relationship of cycle-matching polynomials to other graph polynomials

4.1. Independence polynomial

An independence polynomial is analogous to a matching polynomial, except that an independence polynomial enumerates sets of disjoint ("independent") vertices instead of sets of disjoint edges. Thus, the matching polynomial of a graph is identical to the independence polynomial of its line graph [4]. In general, Gutman and Harary [10] showed that every cycle-matching polynomial (expressed with exponents x^k like the counting polynomial in equation (3)) of the type discussed in the previous section is identical to the independence polynomial of another graph H. Gutman and Harary [10] define the independence polynomial with exponents x^k, but other sources, e.g., Cvetković et al. [11] define it with exponents $x^{2n-2k\cdot}$, compare equations (3) and (5) above. The notation of Gutman and Harary seems more convenient, but in order to place the independence and cycle-matching polynomials on the same footing, it is necessary to use the conversion formula of Miličević et al. [3], modified for $c > 2$. Specifically, to convert equation (3) to (5), Miličević et al. give $\alpha(G, x) = x^n Q(G, -x^2)$, which we modify to

$$\alpha_c(G, x) = x^n \omega(H, -x^c). \tag{10}$$

It is easy to visualize this graph for the examples above of $\alpha_c(C_{60}, x)$. It has 20 vertices at the center of each hexagon (respectively, 12 at the center of each pentagon) and edges $\{v_i, v_j\} \in E$ iff hexagons (respectively, pentagons) have an edge in common in C_{60}. If c is larger than 6, say 9, then $\alpha_c(C_{60}, x)$ would count cycles consisting of the perimeter of a hexagon and an adjacent pentagon, and $\{v_i, v_j\} \in E$ iff the perimeters have any vertex in common. This graph is easy to construct, but not intuitive to visualize. The logical conclusion of this process is $c = n$, in which $\alpha_c(G, x)$ counts Hamiltonian circuits. Here, $\alpha_c(G, x) = x^n - h$, where h is the number of Hamiltonian circuits. The graph whose independence polynomial is related to $\alpha_c(G, x)$ by equation 10 in this case is the complete graph on h vertices, K_h, since no Hamiltonian circuit is independent of any other.

4.2. Immanantal polynomials

It is mathematically possible, if computationally demanding, to derive any coefficient of any cycle-matching polynomial $\alpha_c(G, x)$ from the coefficients of suitable immanantal polynomials and a suitably selected submatrix of the irreducible character matrix of the symmetric group, S_n. For a discussion of immanantal polynomials, see Merris and Watkins [12], Balasubramanian [13], and Cash [14]. In Cash [14], the possibility of using this technique to enumerate

independent sets of both mutually resonant hexagons and all hexagons is mentioned, but no specific example is provided.

Suppose the problem were to count independent sets of three pentagons in a graph on 30 vertices. The irreducible character matrix of S_{30} is a square matrix M_{30} of order $P(30) = 5604$, where $P(n)$ is the partition P function, the number of ways of expressing n as a sum of nonincreasing integers. Choose all the columns of M_{30} indexed by conjugacy classes $\xi = \{\ldots, 1^{15}\}$, i.e., those with exactly 15 fixed points. There are $P(15) - P(14) = 41$ of these, one of which is $\{5^3, 1^{15}\}$, the one corresponding to three pentagons. Any 41 rows of M_{30} may be chosen, so long as the resulting 41×41 matrix has full rank. In practice, it is convenient to choose rows indexed by $\lambda = \{\ldots, 1^{15}\}$, where the first parts of the partitions are the *conjugates* of the corresponding parts ξ. For explanations of terms and examples of character generation, see Merris and Watkins [12] and Liu and Balasubramanian [15]. Call the 41×41 matrix $R = (r_{j,k})$. From each of the 41 immanantal polynomials associated with the selected $\lambda = \{\ldots, 1^{15}\}$, use the coefficient of x^{15}, the one to which $\xi = \{5^3, 1^{15}\}$ contributes. Call these coefficients $i_{\lambda,15}$. Now, produce 41 simultaneous equations,

$$\sum_{k=1}^{41} r_{j,k} q(G, \xi) = i_{\lambda,15}, \tag{11}$$

where the $q(G, \xi)$ are the number of permutations in the graph G belonging to each conjugacy class ξ. Because of the manner in which R was constructed, these equations are guaranteed to be linearly independent, and so they can be solved for $q(G, \{5^3, 1^{15}\})$. It is important to keep in mind that this procedure counts *permutations*. Since each cycle $(c > 2)$ can be permuted in two ways, the solution will be $2^3 = 8$ times the number of ways of choosing three independent pentagons from G. Matchings of any other type of subgraph may be calculated similarly.

5.   Conclusions

Relationships were recently discovered between the Hosoya index and the Coulson function, originally defined on the characteristic polynomials of trees. The present study extended this work by allowing definitions of the Coulson function on the matching polynomials of polycyclic structures and on polynomials that match sets of independent subgraphs other than edges. A variety of results were obtained, some of which clearly merit further study. Also, demonstrated relationships between the subgraph-matching polynomials and other graph polynomials suggested additional directions for investigation.

Disclaimer

This document has been reviewed by the Office of Pollution Prevention and Toxics, USEPA, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Agency, nor does the mention of trade names or commercial products constitute endorsement or recommendation for use.

References

[1] C.A. Coulson, Proc. Cambridge Philos. Soc. 36 (1940) 201.
[2] H. Hosoya, Bull. Chem. Soc. Jpn. 44 (1971) 2331.
[3] A. Miličević, S. Nikolić, D. Plavšić and N. Trinajstić, J. Mol. Des. 2 (2003) 160; http://www.biochempress.com
[4] I. Gutman, D. Vidović and B. Furtula, Chem. Phys. Lett. 355 (2002) 378.
[5] W.C. Herndon, P.C. Nowak, D.A. Connor and P. Lin, J. Am. Chem. Soc. 114 (1992) 41.
[6] P.W. Fowler and D. Mitchell, J. Chem. Inf. Comput. Sci. 35 (1995) 874.
[7] H. Hosoya and T. Yamaguchi, Tetrahedron Lett. (1975) 4659.
[8] I. Gutman, Z. Naturforsch. 37a (1982) 69.
[9] A.T. Balaban and F. Harary, Tetrahedron 24 (1968) 2505.
[10] I. Gutman and F. Harary, Utilitas Math. 24 (1983) 97.
[11] D.M. Cvetcović, M. Doob, I. Gutman and A. Torgašev, *Recent Results in the Theory of Graph Spectra* (North Holland, New York, 1988) p. 122.
[12] R. Merris, and W. Watkins, Linear Algebra Appl. 64 (1985) 223.
[13] K. Balasubramanian, Theor. Chim. Acta 85 (1993) 379.
[14] G.G. Cash, J. Chem. Inf. Comput. Sci. 43 (2003) 1942.
[15] X. Liu and K. Balasubramanian, J. Comput. Chem. 10 (1989) 417.