On Statistics of Graph Energy

Ante Graovac[a,b], Ivan Gutman[c], Peter E. John[d], Dušica Vidović[c], and Ivana Vlah[a]

[a] Faculty of Science, University of Split, Nikole Tesle 12, HR-21000 Split, Croatia
[b] The R. Bošković Institute, P.O. Box 180, HR-10002 Zagreb, Croatia
[c] Faculty of Science, University of Kragujevac, P.O. Box 60, YU-34000 Kragujevac, Yugoslavia, and
[d] Institut für Mathematik, Technische Universität Ilmenau, PF 100565, D-98684 Ilmenau, Germany

Reprint requests to Prof. I. G.; Fax: +381 34 335040; E-mail: gutman@knez.uis.kg.ac.yu

Z. Naturforsch. **56a,** 307–311 (2001); received January 26, 2001

The energy E_G of a graph G is the sum of the absolute values of the eigenvalues of G. In the case whene G is a molecular graph, E_G is closely related to the total π-electron energy of the corresponding conjugated molecule. We determine the average value of the difference between the energy of two graphs, randomly chosen from the set of all graphs with n vertices and m edges. This result provides a criterion for deciding when two (molecular) graphs are almost coenergetic.

Key words: Energy (of graph); Total π-electron Energy; Coenergetic Graphs.

Introduction

The total π-electron energy is one of the most useful quantum-chemical characteristics of a conjugated molecule that can be obtained by means of the Hückel molecular-orbital (HMO) theory. It is computed as [1–3]

$$E_\pi = \alpha n_e + \beta \sum_{i=1}^{n} g_i \lambda_i = \alpha n_e + \beta E,$$

where α and β are the standard HMO parameters, n_e is the number of π-electrons, g_i the occupation number of the i-th molecular orbital, whereas λ_i, $i = 1, 2, \ldots, n$, are the eigenvalues of the respective molecular graph. The non-trivial part of the above expression is E. For non-charged conjugated systems in their ground electronic state E assumes the form

$$E = \begin{cases} 2 \sum_{i=1}^{n/2} \lambda_i & \text{if } n \text{ is even,} \\ 2 \sum_{i=1}^{(n-1)/2} \lambda_i + \lambda_{(n+1)/2} & \text{if } n \text{ is odd,} \end{cases} \quad (1)$$

with the graph eigenvalues being labeled in a non-decreasing manner. For the vast majority of conjugated molecules (1) can be transformed into

$$E = E_G = \sum_{i=1}^{n} |\lambda_i|. \quad (2)$$

In the absence of large steric strain in the carbon-atom skeleton, by means of E_π one can calculate remarkably accurate values for the thermodynamic functions of conjugated hydrocarbons such as the enthalpy of formation, enthalpy of combustion and similar ones. This success of E_π is based on the remarkable fact, first demonstrated by Schaad and Hess [4], that not only the π-, but also the σ-electron energy is proportional to E. For more details see pp. 151–154 in the book [5], where also other chemical and physico-chemical applications of E are outlined.

The right-hand side of (2) is used as the *definition* of the so-called *energy $E = E_G$ of the graph G*. By means of this definition the considerations in the theory of total π-electron energy may be extended so as to include all graphs. In many cases such an extension proves to be advantageous and makes possible to shed more light on the structure-dependency of E_π.

For details on the graph-energy concept and a survey of its mathematical theory see [5–7].

The Average Difference between Two Graph Energies

Two graphs will be said to be *coenergetic* if their energies coincide. Cospectral graphs (i.e., graphs with equal eigenvalues) are, clearly, coenergetic. It is not too difficult to find pairs of non-cospectral coenergetic graphs. (For instance, the triangle, the quadrangle, and the graph consisting of two isolated edges all have $E = 4$; the 6-membered cycle, the graph of the trigonal prism, the 5-vertex complete graph, and the graph con-

0932-0784 / 01 / 0300-0307 $ 06.00

sisting of two triangles all have $E = 8$, etc.) Finding pairs of non-cospectral coenergetic graphs representing isomeric conjugated molecules (that must be connected, possess equal number of vertices and edges and obey a few additional structural requirements [5]) seems to be a much more difficult task. Not a single such pair is known at this moment.

As far as chemical and physico-chemical consequences are concerned, it is immaterial whether the difference of the energies of two (molecular) graphs is zero or very small: in both cases one would predict that the thermodynamic behavior of the respective compounds is nearly the same. However, in order to be able to decide when a particular energy-difference is "*very small*" we need to know what the usual (average) value of such a difference is. In what follows we provide a solution of this problem.

Let $\mathcal{G}_{n,m}$ be the set of all graphs with n vertices and m edges. As usual in chemical graph theory [1, 3, 5] the graphs considered here are *schlicht*, i.e., possess no loops and no multiple edges.

Let $G, H \in \mathcal{G}_{n,m}$. We are interested in the average value of the differenc $|E_G - E_H|$, denoted by $\mathscr{E}(|E_G - E_H|)$, when the averaging is done over all pairs of elements of the set $\mathcal{G}_{n,m}$. Instead of $\mathscr{E}(|E_G - E_H|)$ we will consider the closely similar quantity

$$\sqrt{\mathscr{E}((E_G - E_H)^2)}, \tag{3}$$

whose statistical analysis is significantly simpler.

Indeed, we have

$$\mathscr{E}(E_G) = \mathscr{E}(E_H) = \mathscr{E}(E)$$

and

$$\mathscr{E}((E_G)^2) = \mathscr{E}((E_H)^2) = \mathscr{E}(E^2),$$

and because E_G and E_H may be considered as independent random variables,

$$\mathscr{E}(E_G \cdot E_H) = \mathscr{E}(E_G) \cdot \mathscr{E}(E_H) = (\mathscr{E}(E))^2.$$

Therefore

$$\mathscr{E}((E_G - E_H)^2) = 2\,[\mathscr{E}(E^2) - (\mathscr{E}(E))^2]. \tag{4}$$

Here and later $\mathscr{E}(E)$ and $\mathscr{E}(E^2)$ stand for the average value of the graph energy and its square, respectively, averaged over all elements of the set $\mathcal{G}_{n,m}$.

From the (mathematically exact) formula (4) we see that for the calculation of the average difference (3) we must know the average values of the graph energy and of its square. In the subsequent section we show that the latter two averages are mutually related.

An Approximate Relation between $\mathscr{E}(E)$ and $\mathscr{E}(E^2)$

In order to find a general (approximate) connection between $\mathscr{E}(E)$ and $\mathscr{E}(E^2)$ we assume that in our statistical considerations the graph eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ may be replaced by numbers $x_1, x_2, \ldots, x_n$ which are random variables, *having an arbitrary probability distribution*, the same for each x_i. As we shall see later, this is a plausible assumption.

If so, then

$$\begin{aligned}\mathscr{E}(E^2) &= \mathscr{E}\left(\left(\sum_{i=1}^{n} |\lambda_i|\right)^2\right) = \mathscr{E}\left(\sum_{i=1}^{n} |\lambda_i| \sum_{j=1}^{n} |\lambda_j|\right) \\ &= \mathscr{E}\left(\sum_{i=1}^{n} |\lambda_i|^2 + 2\sum_{i<j} |\lambda_i||\lambda_j|\right) \\ &= 2m + 2\left(\sum_{i<j} \mathscr{E}(|\lambda_i|)\cdot\mathscr{E}(|\lambda_j|)\right)\end{aligned}$$

because for all graphs from $\mathcal{G}_{n,m}$,

$$\sum_{i=1}^{n} (\lambda_i)^2 = 2m. \tag{5}$$

Now, replacing λ_i and λ_j by x_i and x_j and bearing in mind that the latter two variables are statistically independent with $\mathscr{E}(|x_i|) = \mathscr{E}(|x_j|) = \mathscr{E}(|x|)$, we get

$$\begin{aligned}\mathscr{E}(E^2) &\approx 2m + 2\left(\sum_{i<j} \mathscr{E}(|x_i|)\cdot\mathscr{E}(|x_j|)\right) \\ &= 2m + 2\sum_{i<j} \mathscr{E}(|x|))^2 \\ &= 2m + 2\binom{n}{2}(\mathscr{E}(|x|))^2,\end{aligned}$$

which finally yields

$$\mathscr{E}(E^2) \approx 2m + \frac{n-1}{n}[n\,\mathscr{E}(|x|)]^2. \tag{6}$$

On the other hand,

$$\mathscr{E}(E) = \mathscr{E}\left(\sum_{i=1}^{n} |\lambda_i|\right) = \sum_{i=1}^{n} \mathscr{E}(|\lambda_i|) \approx \sum_{i=1}^{n} \mathscr{E}(|x_i|),$$

resulting in

$$\mathscr{E}(E) \approx n\,\mathscr{E}(|x|). \tag{7}$$

Combining (6) and (7) we arrive at the desired relation

$$\mathscr{E}(E^2) \approx 2m + \frac{n-1}{n}(\mathscr{E}(E))^2. \tag{8}$$

Towards a Statistics of Graph Eigenvalues

Formula (8) has been deduced without specifying the actual distribution of the random variables $x_1, x_2, \ldots, x_n$. If, however, we want to calculate the average difference (3) by applying (8) and (4), then we need to know this distribution, at least approximately. Because the true distribution of the graph eigenvalues is not known, various analytical models for its approximation have been put forward [8–12].

Let $\Gamma(x)$ denote the probability of the random variables $x_1, x_2, \ldots, x_n$. Clearly, our aim is that this function provides a statistical model for the distribution of graph eigenvalues (of the graphs from a given set $\mathcal{G}_{n,m}$). Any such function must satisfy the conditions

$$\Gamma(x) \geq 0 \quad \text{for all } x \in (-\infty, +\infty) \tag{9}$$

and

$$\int_{-\infty}^{+\infty} \Gamma(x)\,\mathrm{d}x = 1. \tag{10}$$

Because for all graphs

$$\sum_{i=1}^{n} \lambda_i = 0,$$

we additionally require

$$\int_{-\infty}^{+\infty} x\,\Gamma(x)\,\mathrm{d}x = 0. \tag{11}$$

If so, then in agreement with (2) and (7)

$$\mathcal{E}(E) = n \int_{-\infty}^{+\infty} |x|\,\Gamma(x)\,\mathrm{d}x. \tag{12}$$

Formula (12) is, of course, an approximation, the quality of which depends on how well the function Γ represents the actual distribution of the graph eigenvalues.

A Model for $\Gamma(x)$

This simplest conceivable choice for Γ is the uniform-distribution Ansatz

$$\Gamma(x_i) = \Gamma(x) = \begin{cases} 1/(2a) & \text{if } |x| \leq a, \\ 0 & \text{if } |x| > a. \end{cases} \tag{13}$$

The function (13) satisfies the conditions (9)–(11). The parameter α is determined by imposing the condition

$$\int_{-\infty}^{+\infty} x^2\,\Gamma(x)\,\mathrm{d}x = \frac{2m}{n}, \tag{14}$$

which is just a reformulation of (5).

Direct calculation yields

$$\mathcal{E}(E) = \frac{\sqrt{3}}{2}\sqrt{2mn} \tag{15}$$

which, in view of the fact that $\sqrt{3}/2 \approx 0.87$, is in excellent agreement with the previously known McClelland formula [13]

$$E = 0.91\sqrt{2mn}\,. \tag{16}$$

Recall that the multiplier 0.91 in (16) is an empirically adjusted fitting parameter. The McClelland formula (16) was shown [13–15] to be one of the most accurate (n, m)-type approximations for the total π-electron energy of conjugated hydrocarbons. Thus the Ansatz (13), in spite of its simplicity, yields a reasonably good expression for (the average value of) the graph energy.

By means of the model (13) and by using (4), (8) and (15) we obtain our first estimate of the average difference of graph energies:

$$\mathcal{E}(|E_G - E_H|) \approx \sqrt{\mathcal{E}((E_G - E_H)^2)} \approx \sqrt{m}\,. \tag{17}$$

It is worth noting that the term on the right-hand side of (17) depends only on the number of edges and is independent of the number of vertices.

An Improved Model for $\Gamma(x)$

Formula (17) is a satisfactory approximation for those values of n and m for which the McClelland formula provides a good approximation for E. Recent studies [17, 18] showed that for graphs with large number of edges (which usually are not molecular graphs) the (n, m)-dependence of the energy significantly differs from what the McClelland formula (16) or our expression (15) would predict.

In order to overcome these difficulties, and in order to get a better approximation for the average difference (3) we now put forward a somewhat more flexible model for $\Gamma(x)$:

$$\Gamma(x_i) = \Gamma(x) = \begin{cases} h_b & \text{if} \quad -b \leq x \leq 0, \\ h_a & \text{if} \quad 0 < x \leq a, \\ 0 & \text{otherwise}, \end{cases} \tag{18}$$

If one would choose $a = b$, then the model (18) would reduce to the model (13).

In order to determine a, b, h_a, and h_b we require that the conditions (10), (11), (14) and (19) be obeyed, where

$$\int_{-\infty}^{+\infty} x^3\,\Gamma(x)\,\mathrm{d}x = \frac{6t}{n}. \tag{19}$$

The condition (19) is a reformulation of the identity

$$\sum_{i=1}^{n} (\lambda_i)^3 = 6t,$$

where t is the number of triangles in the respective graph. In (19) the parameter t must be interpreted as the average number of triangles in the graphs from the considered set $\mathcal{G}_{n,m}$.

Using (10) and (11) we get

$$a\, h_a + b\, h_b = 1,$$
$$a^2\, h_a - b^2\, h_b = 0,$$

from which

$$h_a = \frac{b}{a(a+b)}; \quad h_b = \frac{a}{b(a+b)}.$$

Using (14) and (19) we get

$$ab = \frac{6m}{n},$$
$$a - b = \frac{4t}{m},$$

from which

$$a = \sqrt{\frac{4t^2}{m^2} + \frac{6m}{n}} + \frac{2t}{m}; \quad b = \sqrt{\frac{4t^2}{m^2} + \frac{6m}{n}} - \frac{2t}{m}.$$

From (12) follows

$$\mathcal{E}(E) = n\,\frac{ab}{a+b} = 3m\left(\frac{4t^2}{m^2} + \frac{6m}{n}\right)^{-1/2},$$

which combined with (8) and (4) yields within the Ansatz (18)

$$\mathcal{E}((E_G - E_H)^2) = m - \frac{10mnt^2}{3m^3 + 2nt^2}.$$

We see that because of the presence of triangles in (some) graphs, the average difference of graph energies is somewhat smaller than predicted by the uniform-distribution model (13).

Note that by setting $t = 0$ in the above formulas they reduce to the corresponding expressions for the model (13).

In order to calculate t, consider an arbitrary graph G from the set $\mathcal{G}_{n,m}$. Let u, v, w be three vertices of G. The probability that u is adjacent to v is

$$P_1 = \frac{m}{\binom{n}{2}}$$

because in G there are $\binom{n}{2}$ vertex pairs, of which m are adjacent.

Suppose that u and v are adjacent. The probability that u is adjacent to w is

$$P_2 = \frac{m-1}{\binom{n}{2} - 1}$$

because now we do not count the vertex pair u, v and the edge connecting u and v.

Similarly, if we suppose that u and v are adjacent and that u and w are adjacent, then the probability that v is adjacent to w is

$$P_3 = \frac{m-2}{\binom{n}{2} - 2}.$$

The probability that u, v, and w are simultaneously adjacent, i.e. that they form a triangle, is $P_1 \cdot P_2 \cdot P_3$. Then the average number of triangles in the graph considered is equal to

$$t = \binom{n}{3} P_1 \cdot P_2 \cdot P_3,$$

which yields

$$t = \frac{4}{3}\,\frac{m(m-1)(m-2)}{(n+1)(n^2-n-4)}. \tag{20}$$

The right-hand side of formula (20) is, in fact, the average number of triangles in a labeled graph with n vertices and m edges.

We thus arrived at our final result: The average difference of the energies of graphs with n vertices and m edges is, approximately, equal to

$$\mathcal{E}(|E_G - E_H|) \approx \sqrt{\mathcal{E}((E_G - E_H)^2)} \approx \sqrt{m - \frac{10mnt^2}{3m^3 + 2nt^2}}, \tag{21}$$

where t is given by (20).

For molecular graphs t is small and may be neglected, in which case (21) reduces to (17).

Acknowledgements

This work has been done within the DAAD project "*Akademischer Neuaufbau Südosteuropa*" that enabled scientists from Croatia and Serbia to meet (in Germany) and to resume cooperation. The financial support of these activities by DAAD is gratefully acknowledged by all the five authors.

[1] A. Graovac, I. Gutman, and N. Trinajstić, Topological Approach to the Chemistry of Conjugated Molecules, Springer-Verlag, Berlin 1977.
[2] C. A. Coulson, B. O'Leary, and R. B. Mallion, Hückel Theory for Organic Chemists, Academic Press, London 1978.
[3] J. R. Dias, Molecular Orbital Calculations Using Chemical Graph Theory, Springer-Verlag, Berlin 1993.
[4] L. J. Schaad and B. A. Hess, J. Amer. Chem. Soc. **94,** 3068 (1972).
[5] I. Gutman and O. E. Polansky, Mathematical Concepts in Organic Chemistry, Springer-Verlag, Berlin 1986.
[6] I. Gutman, Ber. Math.-Statist. Sekt. Forschungszentrum Graz **103,** 1 (1978).
[7] I. Gutman, in: A. Betten, A. Kohnert, R. Laue, and A. Wassermann (Eds.), Algebraic Combinatorics and Applications, Springer-Verlag, Berlin 2001, pp. 000–000 (in press).
[8] I. Gutman and M. Rašković, Z. Natuforsch. **40a,** 1059 (1985).
[9] A. Graovac, D. Babić, and K. Kovačević, Stud. Phys. Theor. Chem. **51,** 448 (1987).
[10] A. Graovac and J. Cioslowski, Croat. Chem. Acta **61,** 797 (1988).
[11] I. Gutman, A. Graovac, S. Vuković, and S. Marković, J. Serb. Chem. Soc. **54,** 189 (1989).
[12] I. Gutman, T. Soldatović, A. Graovac, and S. Vuković, Chem. Phys. Lett. **334,** 000 (2001) (in press).
[13] B. J. McClelland, J. Chem. Phys. **54,** 640 (1971).
[14] J. Cioslowski and I. Gutman, Z. Naturforsch. **41a,** 861 (1986).
[15] I. Gutman, Topics Curr. Chem. **162,** 29 (1992).
[16] I. Gutman, J. H. Koolen, V. Moulton, M. Parac, T. Soldatović, and D. Vidović, Z. Naturforsch. **55a,** 507 (2000).
[17] I. Gutman, T. Soldatović, and D. Vidović, Chem. Phys. Lett. **297,** 428 (1998).
[18] I. Gutman, D. Vidović, and T. Soldatović, ACH Models in Chem. **136,** 599 (1999).