Relating Estrada index with spectral radius

IVAN GUTMAN[1*#], SLAVKO RADENKOVIĆ[1#], BORIS FURTULA[1],
TOUFIK MANSOUR[2] and MATTHIAS SCHORK[3]

*[1]Faculty of Science, University of Kragujevac, P. O. Box 60, 34000 Kragujevac, Serbia,
[2]Department of Mathematics, University of Haifa, 31905 Haifa, Israel and
[3]Camillo–Sitte–Weg 25, 60488 Frankfurt, Germany*

Abstract: The Estrada index *EE* is a recently proposed molecular structure-descriptor, used in the modeling of certain features of the 3D structure of organic molecules, in particular of the degree of folding of proteins and other long-chain biopolymers. The Estrada index is computed from the spectrum of the molecular graph. Therefore, finding its relation with the spectral radius *r* (= the greatest graph eigenvalue) is of interest, especially because the structure-dependency of *r* is relatively well understood. In this work, the basic characteristics of the relation between *EE* and *r*, which turned out to be much more complicated than initially anticipated, was determined.

Keywords: molecular graph, Estrada index, spectral radius, graph spectrum.

INTRODUCTION

The Cuban–Spanish scholar Ernesto Estrada designed in the year 2000 a new structure descriptor,[1] capable of representing certain features of the 3D structure of organic molecules, especially those of biochemical importance. Eventually, this structure descriptor was named the Estrada index, and is usually denoted by *EE*. It could be shown[2,3] that the *EE* is particularly suitable for characterizing the degree of folding of proteins and similar long-chain biopolymers. More recently, several other applications of the Estrada index were reported,[4−7] which are, however, of lesser chemical relevance.

As the Estrada index is, in a relatively simple manner, computed from the spectrum of the corresponding molecular graph, efforts have been made to use the powerful mathematical apparatus of graph spectral theory[8] for determining the dependence of the *EE* on molecular structure. The hitherto obtained results are available in the literature.[9−12]

Let λ_1, λ_2,..., λ_n be the eigenvalues of the molecular graph G (which in biochemical applications[1–3] may possess weighted edges).[8,13] These eigenvalues form the spectrum of G and will be labeled as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

The greatest graph eigenvalue λ_1 has the property $\lambda_1 \geq |\lambda_k|$, $k = 2,3,...,n$, and is therefore referred to as the spectral radius of the graph G. In what follows, λ_1 is denoted by r.

The spectral radius has been much investigated in graph spectral theory[8,14] and its dependence on the structure of the underlying graph is relatively well understood. This, in particular, applies to molecular graphs.[15,16]

The Estrada index is defined as:[1–7,9–12]

$$EE = \sum_{k=1}^{n} e^{\lambda_k} + \sum_{k=2}^{n} e^{\lambda_k} \tag{1}$$

Thus EE is equal to the sum of terms of the form e^x, where $x = \lambda_k$, $k = 1,2,...,n$, of which the greatest is e^r. Therefore it is plausible to expect that there is a relation (or, at least, a correlation) between the Estrada index and the spectral radius. However, this relation is not simple, as seen from the example shown in Fig. 1. The fact that the data points in Fig. 1 are grouped on several (almost) horizontal lines indicates that, in addition to the spectral radius, EE depends in a lesser manner on structural factors other than r.

In the subsequent section we elaborate an approach aimed at revealing the fine details of the structure-dependence of the EE is elaborated.
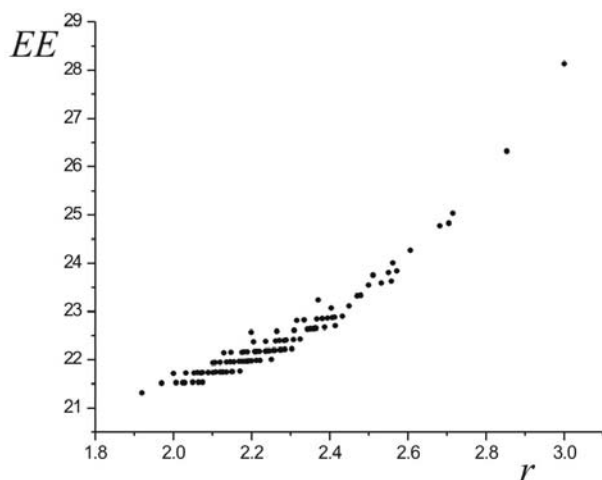

Fig. 1. The Estrada index (EE) of 10-vertex trees, plotted *versus* the spectral radius (r). There is an apparent (curvilinear) correlation between the data points. However, numerous data points lie on almost horizontal lines, indicating that factors other than r also influence the value of EE. In the case of benzenoid and acyclic molecules, the main among these "less important" factors have been identified.[10,11]

A POWER-SERIES-EXPANSION APPROACH

Formula (1) can be rewritten as:

$$EE = e^r \sum_{k=1}^{n} e^{\lambda_k - r}$$

and then, in view of the arguments outlined in the preceding section, it may be expect that the "less important" structural details influencing the value of the Estrada index are contained in the term $\sum_{k=1}^{n} e^{\lambda_k - r}$. Expanding $e^{\lambda_k - r}$ into a power series, one obtains:

$$EE = e^r \sum_{h=0}^{\infty} F_h \tag{2}$$

where

$$F_h = \sum_{k=1}^{n} \frac{(\lambda_k - r)^h}{h!} = \sum_{j=0}^{h} \frac{(-1)^{h-j}}{j!(h-j)!} r^{h-j} M_j$$

with M_j denoting the j-th spectral moment:

$$M_j = \sum_{k=1}^{n} \lambda_k^j$$

The structure dependences of the first few spectral moments of molecular graphs are known.[17–24] For instance:

$$M_0 = n; \; M_1 = 0; \; M_2 = 2m; \; M_3 = 6t$$

where n, m, and t stand, respectively, for the number of vertices, edges, and triangles. For benzenoid molecules, $M_4 = 18m - 12n$, whereas for acyclic molecular graphs, $M_4 = 2Zg - 2n + 2$, where Zg is the Zagreb index, the sum of the squares of the vertex degrees.[17,25–27] It is worth noting that for alternant hydrocarbons (*e.g.*, for acyclic and benzenoid systems), $M_j = 0$ whenever j is odd.

In view of this, the summation on the right-hand side of (2) is truncated, arriving at a series of approximate expressions for the Estrada index, *viz.*

$$EE(p) = e^r \sum_{h=0}^{2p} F_h, \; p = 1,2,\dots \tag{3}$$

If so, then $EE(p)$ will depend on the spectral radius r and on the first $2p$ spectral moments (of which many are equal to zero).

It was previously shown that in the case of alkanes, $EE(2)$ is a monotonically increasing function of the variable r and it was concluded that alkanes with maximal EE value will be those possessing maximal spectral radius.[12] The latter alkanes were earlier characterized by Simić and Tošić,[16] who established that these correspond to the so-called Volkmann trees.[28,29] In a study[12] it was (erroneously) assumed that $EE \approx EE(2)$. To obtain the correct conclusion (concerning Volkmann trees), it was sufficient that there is a positive correlation between EE and $EE(2)$. That this is indeed the case can be seen from Fig. 2.

Numerical testing revealed that for the first few values of p, in particular for $p = 2$, the approximation $EE \approx EE(p)$ is highly inaccurate and should not have

been used. In the subsequent section, this matter is clarified and it is also shown that for any value of p, $p \geq 2$, there is a reasonably good, yet non-linear, correlation between EE and EE(p).
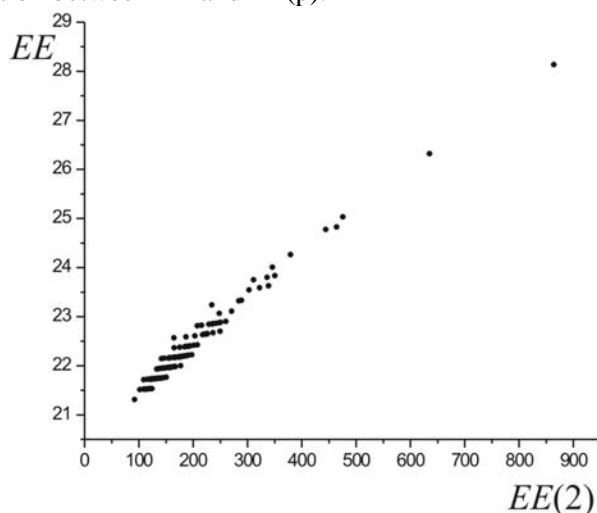

Fig. 2. The Estrada indexes (EE) of 10-vertex trees, plotted *versus* the approximate formula EE(2), *cf.* Eq. (3). The correlation coefficient is equal to 0.993.

NUMERICAL WORK

The accuracy of the approximation $EE \approx EE(p)$ was tested for the first few values of the parameter p. Some characteristic results of this kind, pertaining to trees with 8, 10 and 12 vertices (23, 106 and 551 trees, respectively) are given in Table I.

TABLE I. Average relative errors (ARE) and maximal observed errors (MRE) in % of the approximations $EE \approx EE(p)$, $p = 1,2,\ldots,7$, for 8-, 10- and 12-vertex trees. Recall that there are 23, 106, and 551 such trees, respectively.

	n					
p	8		10		12	
	ARE	MRE	ARE	MRE	ARE	MRE
1	739.36	1426.1	886.83	2327.2	935.76	5092.9
2	617.94	1545.1	792.43	2969.7	1004.33	3506.9
3	281.47	1006.5	387.25	2391.5	475.01	4824.5
4	81.84	410.6	122.53	1245.3	56.42	3044.4
5	16.56	110.2	27.55	430.6	36.90	1287.2
6	2.47	20.7	4.46	104.3	6.67	382.2
7	0.28	2.9	0.63	18.6	0.97	83.5

The data in Table I clearly show that the approximation $EE \approx EE(p)$ is highly inaccurate and that only for very large values of p, say $p > 5$, are some more-or-less satisfactory results obtained. In other words, the expressions $EE(p)$ cannot be used for approximating the Estrada index. This is the bad news.

The good news is that the quantities *EE* and *EE(p)* are reasonably well correlated. The correlation between *EE* and *EE(2)* is essentially linear, as can be seen from Fig. 2. For greater values of the parameter *p*, the correlation becomes pronouncedly curvilinear. Two typical examples are shown in Figs. 3 and 4.
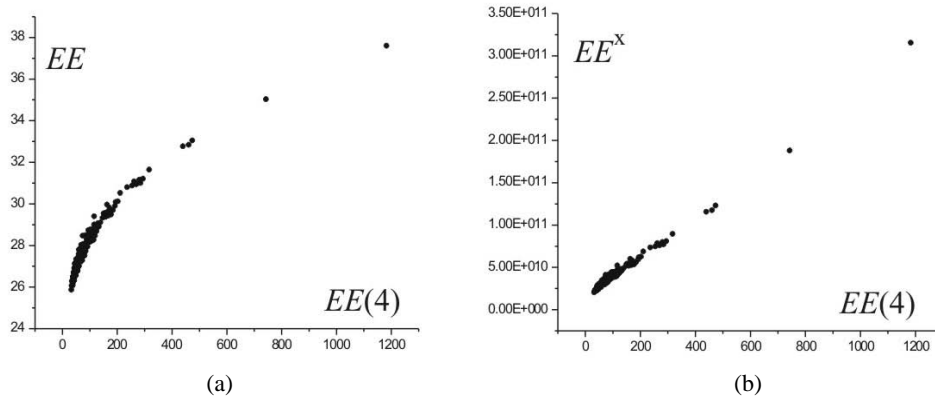

Fig. 3. (a) The Estrada index (*EE*) of 12-vertex trees, plotted *versus* the approximate formula *EE*(4). (b) Linearization of the correlation is achieved by plotting EE^x *vs.* *EE*(4), for *x* = 7.3; the correlation coefficient is 0.995.
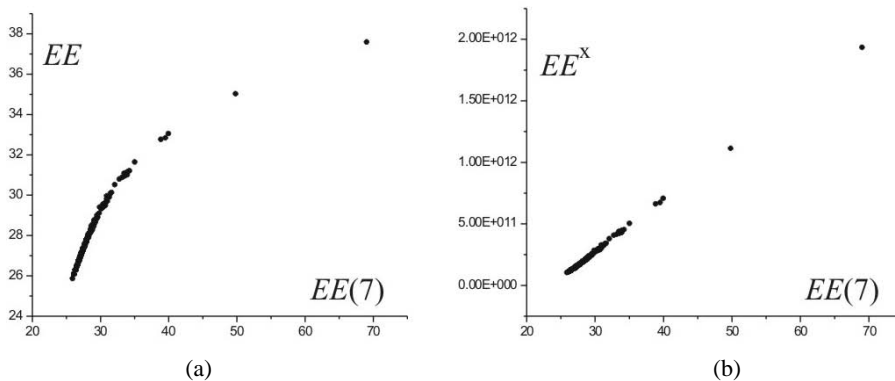

Fig. 4. Same data as in Fig. 3 for *p* = 7 and *x* = 7.8; correlation coefficient = 0.9995.

The curvature of the *EE–EE(p)* correlation is best eliminated by plotting EE^x *versus* *EE(p)*, for some appropriately chosen value of *x*. This value depends both on *p* and on the sample for which it is determined. For instance, for the sample consisting of 10-vertex trees and for *p* = 7, *x* has the value 5.6, in which case, the correlation coefficient for EE^x *vs.* *EE(p)* is as high as 0.9997. Two further examples are given in Figs. 3 and 4.

CONCLUSIONS

The expressions *EE(p)*, given by Eq. (3), provide unacceptably inaccurate approximations for the Estrada index. Only for large values of the parameter *p*

(namely, for $p > 5$) does the approximation $EE \approx EE(p)$ have an average error below 10 %. However, for such large values of p the usage of the formula $EE(p)$ would be impractical and inferior to the direct (exact) calculation of EE from its definition.

On the other hand, for any $p \geq 2$, a reasonably good correlation exists between the Estrada index and $EE(p)$. This correlation is curvilinear, except for $p = 2$ when it is linear (see Fig. 2). Thus, the Estrada index is (in a statistical sense) a monotonically increasing function of the expressions $EE(p)$, $p = 2,3,.....$ As a consequence, if $EE(p)$ is a monotonically increasing function of some variable (say, of the spectral radius r, as in the case of trees and $p = 2$), then is (in a statistical sense) also the Estrada index.

In summary: the present analysis confirms the validity of the conclusion drawn in a previous work[12] that the n-vertex Volkmann tree[28,29] has the greatest Estrada index among all n-vertex chemical trees. More generally: our analysis implies that it is justified to use $EE(2)$ and $EE(3)$ in the study of the structure dependence of the Estrada index. In other words – the structural features on which $EE(2)$ and $EE(3)$ depend (*i.e.*, on which the first few spectral moments depend, which all are known[17−24]) are to a great extent those on which the Estrada index depends. Therefore the employment of the expressions $EE(2)$ and $EE(3)$ enables the finer details of the structure dependency of the Estrada index to be resolved. This task has already been accomplished for benzenoid molecules[10] (using $EE(3)$) and alkanes[11,12] (using $EE(2)$), and now it can be achieved for any other class of molecules for which there is a chemical interest.

ИЗВОД

ВЕЗА ЕСТРАДИНОГ ИНДЕКСА И СПЕКТРАЛНОГ РАДИЈУСА

ИВАН ГУТМАН[1], СЛАВКО РАДЕНКОВИЋ[1], БОРИС ФУРТУЛА[1], TOUFIK MANSOUR[2] и MATTHIAS SCHORK[3]

[1]Природно−маѡемаѡички факулѡеѡ Универзиѡеѡа у Краѓујевцу, Србија, [2]Department of Mathematics, University of Haifa, 31905 Haifa, Israel и [3]Camillo–Sitte–Weg 25, 60488 Frankfurt, Germany

Естрадин индекс *EE* је један недавно предложени молекулски структурни дескриптор, који је примењен при моделирању извесних тродимензионалних структурних карактеристика органских молекула, нарочито степена савијања протеина и других биополимера дугачког ланца. Естрадин индекс се израчунава из спектра молекулског графа. Због тога је од интереса налажење релације између *EE* и спекралног радијуса *r* (= највеће сопствене вредности). Ово тим пре што је зависност *r* од молекулске структуре релативно добро истражена. У раду су одређене основне карактеристике релације између *EE* и *r,* која се показала много сложенијом него што се претпостављало.

REFERENCES

1. E. Estrada, *Chem. Phys. Lett.* **319** (2000) 713
2. E. Estrada, *Bioinformatics* **18** (2002) 697

3. E. Estrada, *Proteins* **54** (2004) 727
4. E. Estrada, J. A. Rodríguez–Velázquez, *Phys. Rev.* **E71** (2005) 056103
5. E. Estrada, J. A. Rodríguez–Velázquez, *Phys. Rev.* **E 72** (2005) 046105
6. E. Estrada, J. A. Rodríguez–Velázquez, M. Randić, *Int. J. Quantum Chem.* **106** (2006) 823
7. E. Estrada, *Phys. Rev.* E75 (2007) 016103
8. D. Cvetković, M. Doob, H. Sachs, *Spectra of Graphs – Theory and Application*, Academic Press, New York, 1980; 2nd revised ed., Barth, Heidelberg, 1995
9. I. Gutman, E. Estrada, J. A. Rodríguez–Velázquez, *Croat. Chem. Acta* **80** (2007) 151
10. I. Gutman, S. Radenković, *Z. Naturforsch.* **62a** (2007) 254
11. I. Gutman, B. Furtula, B. Glišić, V. Marković, A. Vesel, *Indian J Chem.* **46** (2007) 723
12. I. Gutman, B. Furtula, V. Marković, B. Glišić, *Z. Naturforsch.* **62a** (2007) 495
13. I. Gutman, O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer–Verlag, Berlin, 1986.
14. D. Cvetković, P. Rowlinson, *Lin. Multilin. Algebra* **28** (1990) 3
15. I. Gutman, D. Vidović, *Theor. Chem. Acc.* **108** (2002) 98
16. S. K. Simić, D. V. Tošić, *MATCH Commun. Math. Comput. Chem.* **54** (2005) 351
17. I. Gutman, B. Ruščić, N. Trinajstić, C. F. Wilcox, *J. Chem. Phys.* **62** (1975) 3399
18. G. G. Hall, *Theor. Chim. Acta* **70** (1986) 323
19. I. Gutman, *J. Math. Chem.* **1** (1987) 123
20. S. Marković, I. Gutman, *J. Mol. Struct. (Theochem.)* **235** (1991) 81
21. S. Marković, *Theor. Chim. Acta* **81** (1992) 237
22. S. Marković, A. Stajković, *Theor. Chem. Acc.* **96** (1997) 256
23. S. Markovic, *J. Chem. Inf. Comput. Sci.* **39** (1999) 654
24. S. Marković, Z. Marković, R. I. McCrindle, *J. Chem. Inf. Comput. Sci.* **41** (2001) 112
25. B. Zhou, I. Gutman, *MATCH Commun. Math. Comput. Chem.* **55** (2005) 233
26. B. Liu, I.Gutman, *MATCH Commun. Math. Comput. Chem.* **55** (2006) 439
27. B. Zhou, D. Stevanović, *MATCH Commun. Math. Comput. Chem.* **56** (2006) 571
28. M. Fischermann, A. Hoffmann, D. Rautenbach, L. Székely, L. Volkmann, *Discr. Appl. Math.* **122** (2002) 127
29. M. Fischermann, I. Gutman, A. Hoffmann, D. Rautenbach, D. Vidović, L. Volkmann, *Z. Naturforsch.* **57a** (2002) 49.