Regular article

Two early branching indices and the relation between them

Ivan Gutman, Dušica Vidović

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

Abstract. In the 1970s two different structure-descriptors were put forward, aimed at quantifying the extent of branching of the carbon-atom skeleton of organic molecules: Randić's connectivity index, $\chi = \sum (\delta_r \delta_2)^{-1/2}$, where δ_r is the degree of the vertex r of the molecular graph and where the summation goes over all pairs of adjacent vertices (1975), and the greatest eigenvalue, λ_1, of the molecular graph (1973, 1977). Curiously, these two branching indices were never compared. By studying the relation between λ_1 and an auxiliary quantity $\rho = \sum (\delta_r \delta_2)^{+1/2}$, as well as the relation between ρ and χ, we establish the actual relation between χ and λ_1. For differently branched isomers, there is a (rough) decreasing correlation between λ_1 and χ; however, within groups of similarly branched isomers the correlation between λ_1 and χ increases and is nearly linear.

Key words: Connectivity index – Randić index – Greatest eigenvalue (of molecular graph) – Branching – Structure–descriptor

1 Introduction

The fact that the extent of branching of the carbon-atom skeleton of an organic molecule influences its physical and chemical properties was known already in the 19th century [1, 2]. It is therefore not surprising that the earliest attempts to design topological indices suitable for quantitative studies of structure–property relations had to use some numerical measure of branching. Extensive work on quantifying molecular branching started in the 1970s and was pursued in a number of different directions [3–14]. This problem has continued to attract the attention of researchers [15–23]; for further details, historical data, and additional bibliography see the reviews in Refs. [2, 24–33].

Correspondence to: I. Gutman
e-mail: gutman@knez.uis.kg.ac.yu

Here, we are concerned with two branching indices: Randić's connectivity index, χ, and the greatest eigenvalue, λ_1, of the molecular graph. These were among the first structure–descriptors that were explicitly proposed for measuring molecular branching.

The connectivity index was put forward by Randić in 1975 in his seminal article [3] and is defined as

$$\chi = \sum_{(rs)} \frac{1}{\sqrt{\delta_r \delta_s}} \quad , \tag{1}$$

where δ_r is the degree (number of first neighbours) of the vertex r and where the summation embraces all pairs of adjacent vertices of the molecular graph. The article [3] is entitled "On characterization of molecular branching" and in it Randić stated that χ is used for "a theoretical characterization of molecular branching". Randić himself referred to χ as the "branching index", but later the name "connectivity index" prevailed [34, 35, 36].

At first glance χ is not appropriate for measuring branching. Namely, among alkane isomers the least and most branched species have maximal and minimal χ values, respectively, and χ decreases with the increase in the extent of branching. This apparent difficulty is easily overcome by postulating that $\chi(G) < \chi(G')$ implies that G is more branched that G', provided G and G' represent isomers. Similar "inverse" behaviour is found for numerous other topological indices (e.g. Wiener, hyper-Wiener, Hosoya indices) and is not considered as a drawback.

Eventually χ and its proper generalizations became the most popular topological indices and found countless quantitive structure property relationship (QSPR) and quantitive structure activity relationship (QSAR) applications; for details see Refs. [34, 35, 36] and the references quoted therein. It should be noted that after the publication of Ref. [3], neither Randić himself nor the numerous other authors who used the connectivity index insisted on χ being a measure of branching, but rather treated it as one of the several molecular-graph-based structure–descriptors useful for designing quantitative structure–property and structure–activity relations [34, 35, 36].

If G is a graph with n vertices, then its adjacency matrix $A = \|A_{rs}\|$ is a square matrix of order n defined via

$$A_{rs} = \begin{cases} 1 & \text{if the vertices } r \text{ and } s \text{ are adjacent} \\ 0 & \text{otherwise .} \end{cases} \quad (2)$$

In 1977, on the basis of some earlier mathematical results by Lovaász and Pelikán [37] (in which the concept of branching is not mentioned at all), Cvetković and one of the present authors [7] proposed to measure branching by means of the greatest eigenvalue, λ_1, of the adjacency matrix of the molecular graph.

The article [7] is entitled "Note on branching" and in it it is claimed that "... a justification of the ... empirical finding that λ_1 is a measure of branching has been obtained" and that the "fact that the largest eigenvalue of the molecular graph is a measure of branching is ... completely understood".

Although Refs. [7, 37] are regular quoted in the relevant chemical literature, very little attention and almost no further research was devoted to this branching index. A recent article [38] by Randić shows that this displeasing situation may improve in the near future.

Anyway, until now the relation between χ and λ_1 seems to have never been investigated. The aim of this work is to contribute towards filling this gap.

The λ_1 values of the isomeric decanes are plotted versus the respective connectivity indices in Fig. 1. The first impression is that the correlation between λ_1 and χ is very weak. (Such an impression was probably the reason why nobody endeavoured to continue along these lines.)

By a more careful inspection of Fig. 1 (as well as of analogous plots for other isomeric alkanes) we observe a few regularities. These, and their mathematical analysis, are outlined in the subsequent sections.
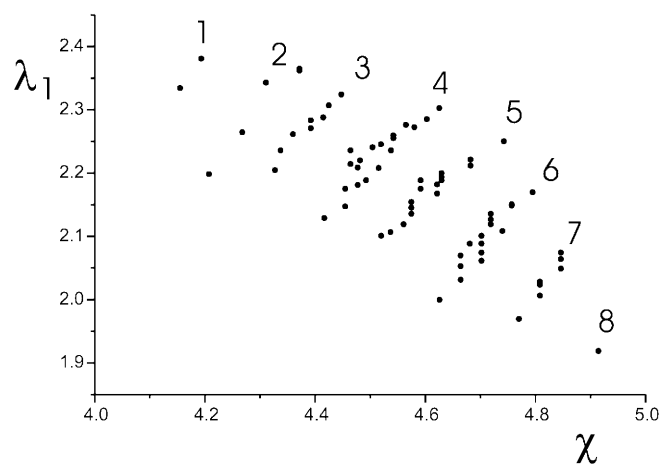
Fig. 1. The greatest eigenvalue, λ_1, of the molecular graphs of the 75 isomeric decanes $C_{10}H_{22}$ versus the connectivity index, χ. Eight groups of points are recognized. Within each group the number of tertiary (n_3) and quaternary (n_4) carbon atoms is constant – group *1*: $n_3 = 1$ and $n_4 = 2$; group *2*: $n_3 = 0$ and $n_4 = 2$; group *3*: $n_3 = 2$ and $n_4 = 1$; group *4*: $n_3 = 4$ and $n_4 = 0$ or $n_3 = 1$ and $n_4 = 1$; group *5*: $n_3 = 3$ and $n_4 = 0$ or $n_3 = 0$ and $n_4 = 1$; group *6*: $n_3 = 2$ and $n_4 = 0$; group *7*: $n_3 = 1$ and $n_4 = 0$; group *8*: $n_3 = 0$ and $n_4 = 0$. Clearly, the extent of branching decreases as one moves from group i to group $i+1$, $i = 1, 2, \ldots, 7$

2 Relations between χ and λ_1

The fundamental property of the connectivity index [3] is that (within sets of isomers, especially within sets of acyclic species) it decreases with the increasing extent of branching; on the other hand, λ_1 is expected to increase with the increasing extent of branching [7, 37]. Consequently, one would anticipate that a decreasing correlation may exist between λ_1 and χ.

This indeed is the case, although such a correlation is remarkable weak (Fig. 1).

By closer examination of the correlation between λ_1 and χ it becomes evident that the (χ, λ_1) points are clustered into certain easily recognizable groups (Fig. 1). We found that the parameters determining to which group a given (χ, λ_1) point belongs are the number of tertiary, n_3, and the number of quaternary, n_4, carbon atoms. Within each such group the correlation between λ_1 and χ increases (!) and is essentially linear (although, again, not particularly good).

An additional surprise was the finding that isomers having equal values of the sum $n_3 + 3n_4$ (but different values of n_3 and n_4) belong to the same cluster of (χ, λ_1) points. (In Fig. 1 this happens within groups 4 and 5.)

We checked these (empirically established) regularities on numerous classes of (chemical and nonchemical) trees and found not a single case of violation. Motivated by this, we tried to find an explanation for such a kind of relation between the two early branching indices.

Our analysis consists of two steps. First, we construct an approximation for λ_1, relation this graph eigenvalue with the degrees of the vertices of the molecular graph. After such an approximation has been designed, we determine the exact mathematical relation between it and the connectivity index. By this, we arrive at an approximate relation between χ and λ_1, shedding some light on the previously stated regularities.

3 Explaining the relations between χ and λ_1

As before, G denotes a molecular graph possessing n vertices and m edges. The connectivity index of G is given by Eq. (1), where δ_r is the degree of the vertex r. Recall that

$$\delta_1 + \delta_2 + \cdots + \delta_n = 2m \quad (3)$$

In what follows we need an auxiliary graph invariant, ρ, defined as [39]

$$\rho = \sum_{(rs)} \sqrt{\delta_r \delta_s} \ . \quad (4)$$

The analogy between χ and ρ should be evident by comparing Eqs. (1) and (4).

Let $C = (C_1, C_2, \ldots, C_n)$ be an n-dimensional row vector and let C^t be the respective column vector. If C^t is the first eigenvector of the adjacency matrix A, i.e., the eigenvector corresponding to the greatest eigenvalue λ_1, then $AC^t = \lambda_1 C^t$, implying

$$\lambda_1 = \frac{CAC^t}{CC^t} \quad . \tag{5}$$

By expanding the right-hand side of Eq. (5) and bearing in mind Eq. (2), we obtain

$$\lambda_1 = \frac{2\sum_{(rs)} C_r C_s}{\sum_{r=1}^{n} (C_r)^2} \quad . \tag{6}$$

Now, it has been known for some time [39, 40, 41, 42] that the rth component of the first eigenvector of a molecular graph is related to the degree of the rth vertex. In particular, as a relatively good approximation,

$$(C_r)^2 \approx \delta_r$$

By setting $C \approx (\sqrt{\delta_1}, \sqrt{\delta_2}, \ldots, \sqrt{\delta_n})$ into Eqs. (5) and (6) we arrive at

$$\lambda_1 \approx \frac{2\sum_{(rs)} \sqrt{\delta_r \delta_s}}{\sum_{r=1}^{n} \delta_r} \quad ,$$

which in view of Eqs. (3) and (4) becomes

$$\lambda_1 \approx \frac{\rho}{m} \quad . \tag{7}$$

The applicability of the approximation in Eq. (7) is illustrated in Fig. 2.

The statistical data indicating the quality of the approximation in Eq. (7) for sets of isomeric alkanes are given in Table 1.

In view of Eq. (7) it may be expected that the relation between ρ and χ will possess features similar to what previously was found for the relation between λ_1 and χ. Recall that in all the samples examined, m is constant (and is equal to $n-1$). This indeed is the case, as seen from the example depicted in Fig. 3.

Again, the (ρ, χ) points are separated into several groups and, again, the parameters n_3 and n_4 determine to which group a given point will belong. This time, however, the points within a group lie exactly on straight lines, which happen to be parallel and equidistant (Fig. 4).
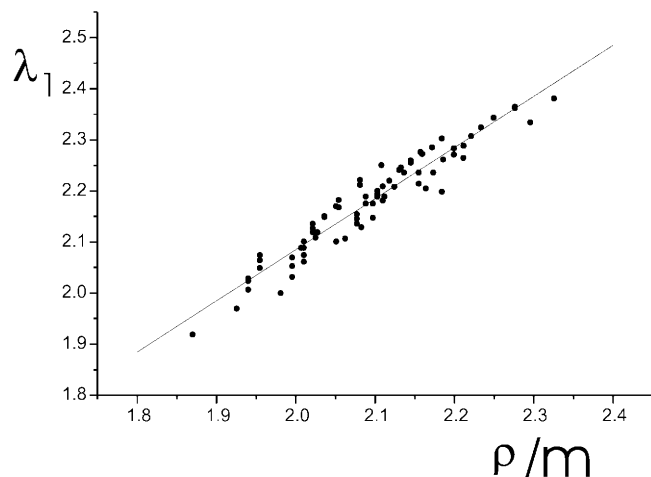
Table 1. Statistical data showing the quality of the approximation in Eq. (7) in the case of isomeric alkanes with n carbon atoms. R is the correlation coefficient, SD is the standard deviation, and A and B are the coefficients in the expression $\lambda_1 \approx A(\rho/m) + B$, calculated by least-squares fitting. Note that in all cases, A and B practically coincide with unity and zero, respectively

n	R	SD	A	B
6	0.997	0.009	0.935 ± 0.040	0.156 ± 0.077
7	0.990	0.016	0.958 ± 0.052	0.131 ± 0.104
8	0.982	0.021	0.950 ± 0.045	0.160 ± 0.092
9	0.970	0.025	0.995 ± 0.043	0.085 ± 0.089
10	0.960	0.028	1.000 ± 0.034	0.084 ± 0.072
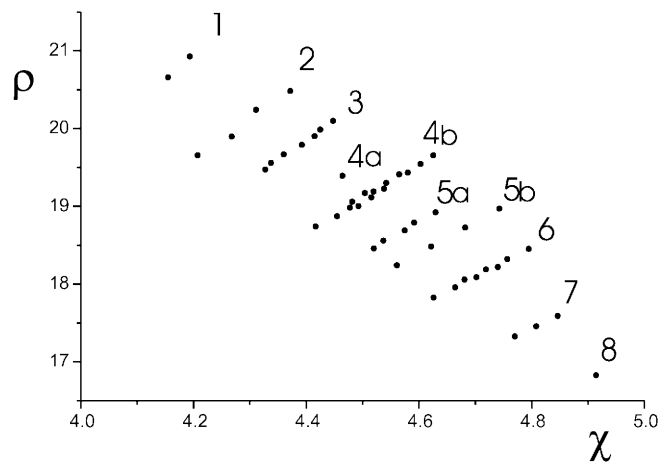
Fig. 3. The topological index, ρ, of isomeric decanes versus χ; several pairs of isomers have equal values of both ρ and χ and therefore the number of points is (seemingly) smaller than in Fig. 1. The points are grouped according to the parameters n_3 and n_4, as in Fig. 1. Exceptionally, groups 4 and 5 from Fig. 1 are now split into group *4a*: $n_3 = 4$ and $n_4 = 0$; group *4b*: $n_3 = 1$ and $n_4 = 1$; group *5a*: $n_3 = 3$ and $n_4 = 0$; and group *5b*: $n_3 = 0$ and $n_4 = 1$. Within each group the points lie on one or several mutually parallel and equidistant straight lines, cf. Fig. 4

The previously specified behaviour of the $\rho - \chi$ relation can be verified by means of the following mathematical analysis.

Denote the number vertices of degree 1, 2, 3, and 4 by n_1, n_2, n_3, and n_4, respectively. Clearly, in the case of molecular graphs of saturated hydrocarbons, n_1 is the number of methyl groups, whereas n_2, n_3, and n_4 count the secondary, tertiary, and quaternary carbon atoms.

Denote the number of edges connecting a vertex of degree i by a vertex of degree j by m_{ij}. Then, the following "bookkeeping" relations are obeyed [43, 44]:
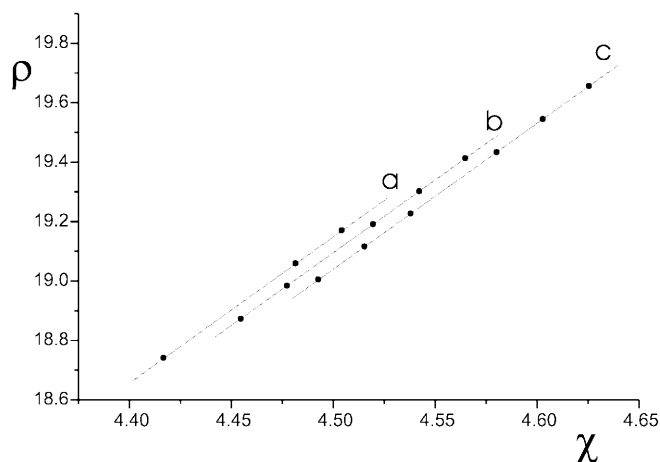
$$n_1 + n_2 + n_3 + n_4 = n \quad , \tag{8}$$

$$n_1 + 2n_2 + 3n_3 + 4n_4 = 2m \quad , \tag{9}$$

$$m_{12} + m_{13} + m_{14} = n_1 \quad , \tag{10}$$

$$m_{12} + 2m_{22} + m_{23} + m_{24} = 2n_2 \quad , \tag{11}$$

$$m_{13} + m_{23} + 2m_{33} + m_{34} = 3n_3 \quad , \tag{12}$$

Fig. 2. Exact versus approximate values of the greatest eigenvalue of the molecular graphs of isomeric decanes, Eq. (7). Note that the regression line goes through the origin and has unit slope; for details see Table 1

Fig. 4. The points from Fig. 3 belonging to group 4b. They lie on three parallel and equidistant straight lines, depending on the value of the parameter m_{22} – *line a*: $m_{22} = 2$; *line b*: $m_{22} = 1$; *line c*: $m_{22} = 0$. For all molecules in group 4b, $m_{33} = m_{44} = 0$

$$m_{14} + m_{24} + m_{34} + 2m_{44} = 4n_4 \; . \tag{13}$$

Further, the topological indices χ and ρ can be written as

$$\chi = \frac{m_{12}}{\sqrt{2}} + \frac{m_{13}}{\sqrt{3}} + \frac{m_{14}}{2} + \frac{m_{22}}{2} + \frac{m_{23}}{\sqrt{6}}$$
$$+ \frac{m_{24}}{\sqrt{8}} + \frac{m_{33}}{3} + \frac{m_{34}}{\sqrt{12}} + \frac{m_{44}}{4} \; , \tag{14}$$

$$\rho = \sqrt{2}m_{12} + \sqrt{3}m_{13} + 2m_{14} + 2m_{22} + \sqrt{6}m_{23} + \sqrt{8}m_{24}$$
$$+ 3m_{33} + \sqrt{13}m_{34} + 4m_{44} \; . \tag{15}$$

Equations (8), (9), (10), (11), (12), (13), (14), and (15) are linear in terms of the parameters n_i and m_{ij}. They can be combined in many different ways. The successful strategy happens to be the following. We discovered it by trial and error, guided by empirical findings, especially those depicted in Figs. 3 and 4.

Express m_{12}, m_{13}, m_{14}, and m_{24} from Eqs. (10), (11), (12), and (13) and substitute into Eqs. (14) and (15). Combine the expressions so obtained for χ and ρ to eliminate m_{34}. Then, fortunately, the term m_{23} will also disappear. Finally, eliminate n_1 and n_2 by means of Eqs. (8) and (9). This results in

$$\rho = 2\sqrt{6}\chi + \left(4 + 4\sqrt{3} - 5\sqrt{2} - 2\sqrt{6}\right)n$$
$$+ \left(8\sqrt{2} + 3\sqrt{6} - 6 - 7\sqrt{3}\right)m$$
$$+ \left(2 + 11\sqrt{3} - \frac{23}{2}\sqrt{2} - \sqrt{6}\right)n_3$$
$$+ \left(12 + 12\sqrt{3} - 9\sqrt{2} - 6\sqrt{6}\right)n_4$$
$$+ \left(4 + 3\sqrt{3} - 3\sqrt{2} - 2\sqrt{6}\right)m_{22}$$
$$+ \left(5 + 3\sqrt{2} - 3\sqrt{3} - \frac{5}{3}\sqrt{6}\right)m_{33}$$
$$+ \left(2 - \sqrt{2} - \sqrt{3} + \frac{1}{2}\sqrt{6}\right)m_{44} \; , \tag{16}$$

i.e.,

$$\rho = 4.898979\chi - 1.041844n + 0.537822m + 2.339613n_3$$
$$+ 5.359749n_4 + 0.054532m_{22} - 0.035995m_{33}$$
$$+ 0.078481m_{44} \; .$$

From Eq. (16) all the observed features of the relation between ρ and χ are easily understood. The noteworthy (and decisive) detail is that the coefficients associated with the parameters m_{22}, m_{33} and m_{44} are order of magnitude smaller than the coefficients of n_3 and n_4. This cause the clustering of the (ρ, χ) points to be determined solely by n_3 by n_4 and within a group (with fixed values of n_3 and n_4) the linear relation between ρ and χ to be determined solely by m_{22}, m_{33}, and m_{44}. Note that almost always (especially in chemically realistic examples) $m_{44} = 0$. All the straight lines on which the (ρ, χ) points lie are mutually parallel and have slopes of $2\sqrt{6} = 4.898979$.

Bearing in mind Eq. (7) the previous conclusions concerning the relation between ρ and χ are also applicable (as reasonable approximations) in the case of the correlation between λ_1 and χ.

In Eq. (16) the ratio between the coefficients at n_4 and n_3 is 2.3. Because this is sufficiently different from 3, the (ρ, χ) points having equal $(n_3 + 3n_4)$ values are separated into distinct, yet nearby–lying, groups. The respective (λ_1, χ) points belong to a single group (Figs. 1, 3).

4 Concluding remarks

In the chemical literature several ways to measure molecular branching have been considered [3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23]. In the absence of a generally accepted view on how branching should be quantified, it is not possible to claim that χ is a better measure of branching than λ_1, or vice versa. However, on the basis of the results reported here, one must conclude that the greatest eigenvalue of the molecular graph and the connectivity index are not compatible as measures of molecular branching, and cannot be both used for the same purpose. Curiously, a quarter of a century was needed to arrive at this simple conclusion.

From a pragmatic point of view, preference should be given to the branching index that is used by the majority of colleagues, and that found most the numerous applications in QSPR and QSAR studies. This, no doubt, is the connectivity index, χ.

References

1. Rouvray DH (1990) In: Bonchev D, Rouvray DH (eds) Chemical graph theory. Introduction and fundamentals. Abacus, New York, pp 1–39
2. Balaban AT, Ivanciuc O (1999) In: Devillers J, Balaban AT (eds) Topological indices and related descriptors in QSAR and QSPR. Gordon & Breach, Amsterdam, pp 21–57
3. Randić M (1975) J Am Chem Soc 97: 6609

4. Gutman I, Ruščić B, Trinajstić N, Wilcox CF (1975) J Chem Phys 62: 3399
5. Somayajulu GR, Zwolinski BJ (1976) J Chem Soc Faraday Trans 2 72: 2213
6. Bonchev D, Trinajstić N (1977) J Chem Phys 67: 4517
7. Cvetković DM, Gutman I (1977) Croat Chem Acta 49: 115
8. Gutman I, Randić M (1977) Chem Phys Lett 47: 15
9. Bonchev D, Trinajstić N (1978) Int J Quantum Chem Quantum Chem Symp 12: 293
10. Randić M (1978) Chem Phys Lett 55: 547
11. Balaban AT (1979) Theor Chim Acta 53: 355
12. Balaban AT, Motoc I (1979) Commun Math Chem (MATCH) 5: 197
13. Bonchev D, Knop JV, Trinajstić N (1979) Commun Math Chem (MATCH) 6: 21
14. Ruch E, Gutman I (1979) J Comb Inf Syst Sci 4: 285
15. Bertz SH (1988) Discrete Appl Math 19: 65
16. Rakshit SC, Banerjee M, Hazra B (1988) Indian J Chem Sect A 27: 183
17. Rouvray DH (1988) Discrete Appl Math 19: 317
18. Schultz HP (1989) J Chem Inf Comput Sci 29: 227
19. Bonchev D (1995) J Mol Struct (THEOCHEM) 336: 137
20. Patra SM, Mishra RK (1997) Bull Chem Soc Jpn 70: 997
21. Randić M (1997) Acta Chim Solv 44: 57
22. Randić M (1998) J Math Chem 24: 345
23. Dobrynin AA (2000) Commun Math Chem (MATCH) 41: 119
24. Balaban AT (2001) In: Diudea MV (ed) QSPR/QSAR studies by molecular descriptors. Nova, Huntington, pp 1–30
25. Motoc I, Balaban AT (1981) Rev Roum Chim 26: 593
26. Balaban AT (1985) J Chem Inf Comput Sci 25: 334
27. Randić M, Hansen PJ, Jurs PC (1988) J Chem Inf Comput Sci 28: 60
28. Randić M, Razinger M (1995) J Chem Inf Comput Sci 35: 140
29. Balaban AT (1995) J Chem Inf Comput Sci 35: 339
30. Balaban AT (1997) J Chem Inf Comput Sci 37: 645
31. Randić M (1997) J Chem Inf Comput Sci 37: 672
32. Basak SC, Balaban AT, Grunwald GD, Gute BD (2000) J Chem Inf Comput Sci 40: 891
33. Randić M, Zupan J (2001) J Chem Inf Comput Sci 41: 550
34. Kier LB, Hall LH (1976) Molecular connectivity in chemistry and drug research. Academic New York
35. Kier LB, Hall LH (1986) Molecular connectivity in structure-activity analysis. Wiley, New York
36. Pogliani L (2000) Chem Rev 100: 3827
37. Lovász L, Pelikán J (1973) Period Math Hung 3: 175
38. Randić M, Vračko M, Novič M (2001) In: Diudea MV (ed) QSPR/QSAR studies by molecular descriptors. Nova, Huntington, pp 147–211
39. Polansky OE, Gutman I (1979) Commun Math Chem (MATCH) 5: 149
40. Randić M (1980) J Comput Chem 1: 386
41. Balaban AT, Ciubotariu D, Medeleanu M (1991) J Chem Inf Comput Sci 31: 517
42. Rücker C, Rücker G (1994) J Chem Inf Comput Sci 34: 534
43. Gutman I, Miljković O, Caporossi G, Hansen P (1999) Chem Phys Lett 306: 366
44. Gutman I, Araujo O, Morales DA (2000) J Chem Inf Comput Sci 40: 572