J.Serb.Chem.Soc. 66(9)605–611(2001)
JSCS – 2890

UDC 669.18é
Original scientific paper

Choosing the exponent in the definition of the connectivity index

IVAN GUTMAN[#] and MIRKO LEPOVIĆ

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

(Received 4 May 2001)

Let δ_υ denote the degree of the vertex υ of a molecular graph G. Then the connectivity index of G is defined as $C(\lambda) = G(\lambda; C) = \Sigma (\delta_u \delta_\upsilon)^\lambda$, where the summation goes over all pairs of adjacent vertices. The exponent λ is usually chosen to be equal to –1/2, but other options were considered as well, especially $\lambda = -1$. We show that whereas $C(-1/2)$ is a suitable measure of branching of the carbon-atom skeleton of organic molecules, and thus applicable as a topological index for modeling physico-chemical properties of the respective compounds, this is not the case with $C(-1)$. The value of λ is established, beyond which $C(\lambda)$ fails to correctly reflect molecular branching.

Keywords: connectivity index, branching, topological index, tree, chemical tree.

INTRODUCTION

The *connectivity index* (originally named[1] *"branching indes"*) is defined as

$$x = x(G) = \sum_{u,\upsilon} (\delta_u \delta_\upsilon)^{-1/2} \qquad (1)$$

where δ_u denotes the degree (= number of first neighbors) of the vertex u of the molecular graph G, and where the summation goes over all pairs of adjacent vertices of G. This structure-descriptor, introduced a quarter of century ago,[1] eventually became one of the most popular topological indices. Two books[2,3] and scores of papers have been written on its applications for predicting physico-chemical and pharmacologic properties of organic compounds details and further bibliography can be found in three crecent monographs.[4–6]

Formula (1) is a special case of a more general "connectivity index" C, defined as

$$C(\lambda) = C(\lambda; G) = \sum_{u,\upsilon} (\delta_u \delta_\upsilon)^\lambda \qquad (2)$$

Evidently, $x = C(-1/2)$.

Another special case of Eq. (2) is $C(\lambda)$ for $\lambda = +1$, the so-called "*2nd Zagreb group index*", put forward[7,8] as early as in 1972; for details on $C(1)$ see the book.[5] Some properties of this, otherwise not much studied, topological index were recently communicated.[9]

Serbian Chemical Society active member.

Already in the seminal paper[1] there was no convincing argument given why in formula (2) one should choose $\lambda = -1/2$. The original aim was to provide a numerical measure of the branching[10–12] of the carbon-atom skeleton of an alkane. Based on the analysis of the data obtained for butane, pentane and hexane isomers, it was concluded[1] that both choices $\lambda = -1/2$ and $\lambda = -1$ were equally plausible. The former choice was preferred because $C(-1/2)$ had greater isomer-discriminating power[13] than $C(-1)$. Eventually, the possibility that the exponent λ assumes values other than –0.5 was considered in several papers.[14–17] Indeed, if one would view λ as a variable that is adjusted so as to optimize the correlation between $C(\lambda)$ and some physico-chemical property, then λ so determined would differ from –0.5. The disadvantage of such an approach is that the chosen value of λ significantly depends both on the physico-chemical property used and on the sample of molecules employed.

The option $\lambda = -1$ was recently examined in some detail by Clark and Moon.[18] In what follows we denote $C(-1) = C(-1; G)$ by $\mu = \mu(G)$.

In this paper we are concerned with certain properties of the connectivity index $C(\lambda)$ of trees. Recall that a tree is a connected acyclic graph. A chemical tree is a tree with property $\delta_\upsilon \leq 4$ for all vertices υ. Chemical trees provide a graph representation of alkanes.

MEASURING BRANCHING BY MEANS OF CONNECTIVITY INDICES

As far as branching is concerned there are two distinguishable trees: the path P_n (with the property $\delta_\upsilon \leq 2$ for all vertices) and the star S_n (possessing a vertex u with $\delta_u = n - 1$ and $\delta_\upsilon = 1$ for all other vertices).[19]

Within the set of all n-vertex trees, the path P_n is the least branched and the star S_n the most branched species. In view of this, a necessary condition for any topological index to be an acceptable measure of branching is that its values be extremal for P_n and S_n. Indeed, if T is any n-vertex tree, but $T \neq P_n, S_n$, then

$$x(S_n) < x(T) < x(P_n) \tag{3}$$

The proofs of the general validity of the left- and right-hand side inequalities in (3) were recently given by Bollobás and Erdös[9] and by Caporossi *et al.,* [20] respectively. Paths P_n are, of course, chemical trees (representing the normal alkanes). The stars S_n are chemical trees only up to $n = 5$. For $n \geq 6$ the chemical trees with minimal χ were characterized.[21]

It was previously anticipated[1] that analogous inequalities hold when $\lambda = -1$, namely:

$$\mu(S_n) < \mu(T) < \mu(P_n) \tag{4}$$

As a kind of surprise, Clark and Moon established[18] that when n is sufficiently large the right-hand side inequality in (4) is violated, *i.e.,* P_n is not the tree with the maximal μ-value. Consequently, μ cannot be used as a measure of branching and its applicability in QSPR and QSAR studies is doubtful.

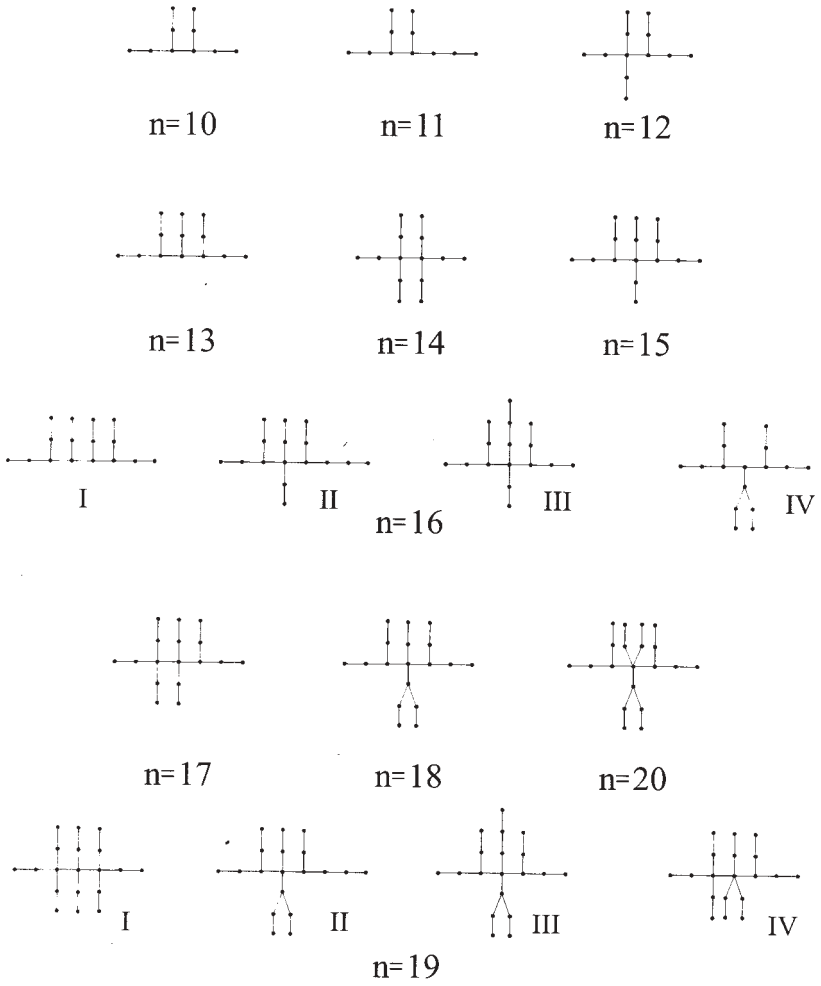

Fig. 1 The *n*-vertex trees with maximal connectivity indices $\mu = C(-1)$. For $n = 16$ and $n = 19$ there are four distinct trees with equal maximal μ-value. For $n = 10, 11, ..., 20$ the respective maximal μ-values are: 2.77778, 3.02778, 3.29167, 3.55556, 3.81250, 4.08333, 4.60417, 4.87500, 5.12500 and 5.40000.

In order to learn more on the validity of the inequalities (4), we determined the *n*-vertex tree(s) with minimal and maximal μ-values for *n* up to 20. For all the examined values of *n*, the star has the minimal μ-value, in agreement with (4). For $n \leq 9$ the path has the maximal μ-value, in agreement with (4). However, for $n \geq 10$ the trees with maximal μ differ significantly from P_n; these trees are depicted in Fig. 1.

From Fig. 1 it is evident that the trees with maximal μ-index are highly branched. Otherwise, their general structure is not easy to characterize. For $n = 16$ and $n = 19$ (and most probably for other values of $n > 20$), the tree with maximal μ is not unique.

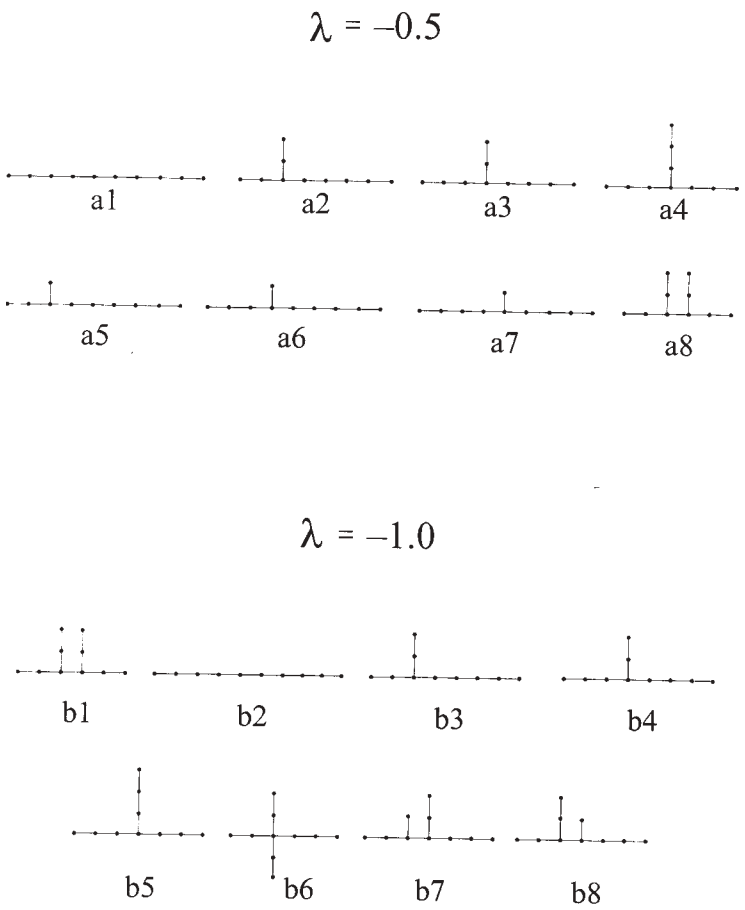

Fig. 2. The first few 10-vertex trees ordered according to decreasing connectivity indices: the species $a1$–$a8$ pertain to χ whereas $b1$–$b8$ to μ: $\chi(a1) = 4.91421$, $\chi(a2) = \chi(a3) = \chi(a4) = 4.84606$, $\chi(a5) = \chi(a6) = \chi(a7) = 4.80806$, $\chi(a8) = 4.79475$, $\mu(b1) = 2.77778$, $\mu(b2) = \mu(b3) = ... = \mu(b6) = 2.75000$, $\mu(b7) = \mu(b8) = 2.69444$. Note that $a1 = b2$, $a2 = b3$, $a3 = b4$ and $a4 = b5$; hence in the case $n = 10$ there is still some agreement between the two orderings.

Most of the species depicted in Fig. 1 are chemical trees; exceptions are only the 19-vertex tree IV and the 20-vertex tree. We expect that more non-chemical trees will be encountered for higher values of n.

Intending to shed more light on the phenomenon described above we compared the orderings of trees according to decreasing χ and μ. Two characteristic results are shown in Figs. 2 and 3, respectively. As seen from these Figures, the two orderings are significantly different, especially for larger values of n (as in Fig. 3).

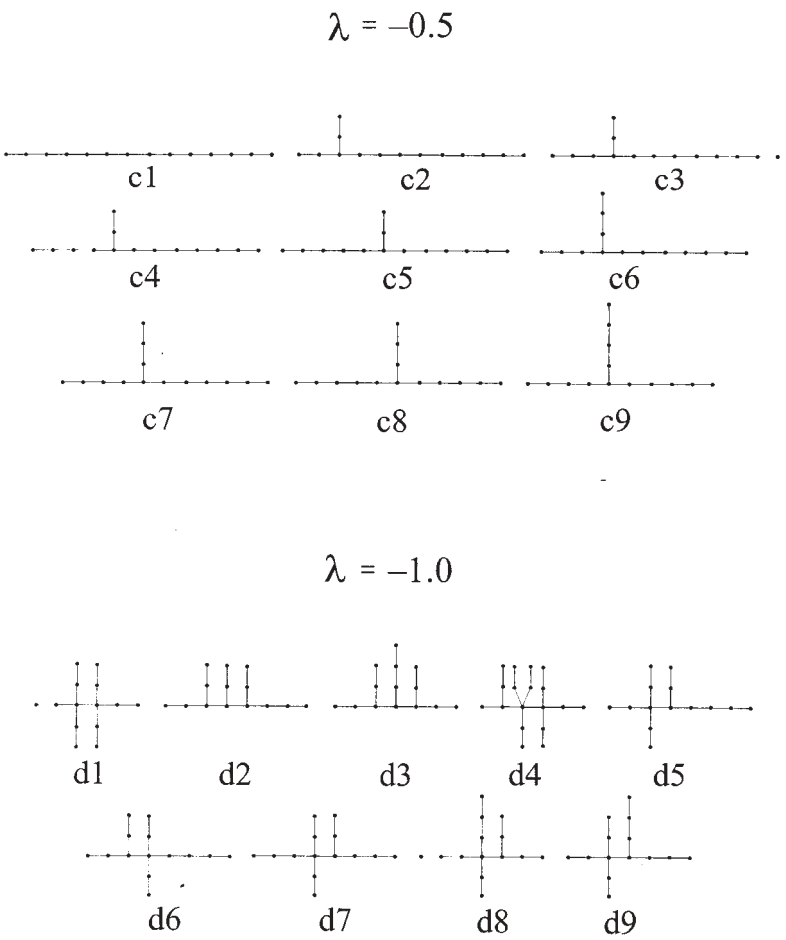

Fig. 3. Same data as in Fig. 2, for n = 14: $\chi(c1) = 6.91421$, $\chi(c2) = \chi(c3) = ... = \chi(c9) = 6.84606$, $\mu(d1) = 3.81250$, $\mu(d2) = \mu(d3) = 3.80556$, $\mu(d4) = 3.80000$, $\mu(d5) = \mu(d6) = ... = \mu(d9) = 3.79167$. No tree $c1 - c9$ coincides with any of the trees $d1 - d9$, indicating a complete disagreement between the two orderings.

THE BREAKDOWN OF THE CONNECTIVITY INDEX

The data presented in Figs. 2 and 3 clearly illustrate the fact that the connectivity index for $\lambda = -1/2$ does, and the connectivity index for $\lambda = -1$ does not provide a plausible measure of molecular branching. In other words, when the exponent λ is decreased from –0.5 to –1.0 a breakdown of the connectivity index $C(\lambda)$ occurs, making it unsuitable for QSPR and QSAR purposes (at least, as far as branching-dependent molecular properties are concerned). In the case of n-vertex chemical trees this happens for all $n \geq 10$.

The breakdown begins at the "critical" value of $\lambda = \lambda_{crit}$, for which the equality

$$C(\lambda; T) = C(\lambda; P_n) \quad (5)$$

is satisfied by the first n-vertex tree T, different from P_n. Usually (but not always) T is just one of the trees depicted in Fig. 1.

TABLE I. Values of the exponent λ below which the path P_n is no longer the tree with maximal connectivity index, Eq. (1), and the tree T which takes over the lead; for details see Eq. (5)

n	λ_{crit}	T
10	– 0.90821	Tree in Figure 1
11	– 0.90821	Tree in Figure 1
12	– 0.91833	Tree in Figure 1
13	– 0.87976	Tree in Figure 1
14	– 0.87976	3,4,5-Triethylheptane
15	– 0.87642	Tree in Figure 1
16	– 0.86594	Trees I & IV in Figure 1
17	– 0.88191	Tree in Figure 1
18	– 0.85096	Tree in Figure 1
19	– 0.85096	Trees II & III in Figure 1
20	– 0.86108	Tree in Figure I

The critical values for the exponent λ, calculated by means of Eq. (5), as well as the respective tree T, are given in Table I.

DISCUSSION

From the data given in Table I, it can be seen that the breakdown of the connectivity index occurs around λ = – 0.9, which is relatively far from the adopted value λ = – 0.5 and relatively near to λ = –1, the other option initially considered as equally plausible. This may be the reason why the problems with $C(-1)$ remained unnoticed for a long time. It turns out that choosing λ = – 0.5 for the exponent in the definition of the connectivity index (instead of λ = –1) was a rather fortunate decision. Our analysis sheds some new light on the true meaning of this choice. Remarkably, this analysis comes more than 25 years after the connectivity index was conceived. This delay seems to be caused by the fact that genuine mathematical research of the connectivity index, revealing its concealed properties, started only quite recently.[9,18,20,21]

И З В О Д

БИРАЊЕ ЕКСПОНЕНТА У ДЕФИНИЦИЈИ ИНДЕКСА ПОВЕЗАНОСТИ

ИВАН ГУТМАН и МИРКО ЛЕПОВИЋ

Природно-математички факултет у Крагујевцу

Нека δ_υ означава степен чвора υ молекулског графа G. Тада је индекс повезаности графа G дефинисан као $C(\lambda) = C(\lambda; G) = \Sigma\,(\delta_u\,\delta_\upsilon)^\lambda$, где се сумирање врши преко свих парова суседних чворова. За експонент λ се обично узима вредност –1/2, мада су биле

разматране и друге могућности, нарочито $\lambda = -1$. Показано је да док $C(-1/2)$ представља погодну меру разгранатости угљеничног скелета органских молекула, и због тога је применљив као тополошки индекс за моделовање физичко-хемијских особина одговарајућих једињења, то није случај са $C(-1)$. Одређена је вредност за λ преко које $C(\lambda)$ престаје да коректно одсликава разгранатост молекула.

(Примљено 4. маја 2001)

REFERENCES

1. M. Randić, *J. Am. Chem. Soc.* **97** (1975) 6609
2. L. B. Kier, L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976
3. L. B. Kier, L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis*, Research Studies Press, Letchworth, 1986
4. J. Devillers, A. T. Balaban (Eds.), *Topological Indices and Related Desriptors in QSAR and QSPR,* Gordon and Breach, Amsterdam, 1999
5. R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, 2000
6. M. V. Diudea (Ed.), *QSPR/QSAR Studies by Molecular Descriptors*, Nova, New York, 2001
7. I. Gutman, N. Trinajstić, *Chem. Phys. Lett.* **17** (1972) 535
8. I. Gutman, B. Ruščić, N. Trinajstić, C. F. Wilcox, *J. Chem. Phys.* **62** (1975) 3399
9. B. Bollobás, P. Erdös, *Ars Combinatoria* **50** (1998) 225
10. There is no precise definition of "*branching*". It has been argued[11,12] that branching is a non-numerical notion and that two objects from a given class (*e.g.,* two trees) are not necessarily comparable with regard to their branching. Nevertheless, χ was aimed at providing a numerical measure of branching, especially of trees and chemical trees: According to this viewpoint, if the trees T_a and T_b have equal number of vertices nad if $\chi(T_a) > \chi(T_b)$ then T_b may be considered to be more branched than T_a
11. I. Gutman, M. Randić, *Chem. Phys. Lett.* **47** (1977) 15
12. E. Ruch. I. Gutman, *J. Comb. Inf. System Sci.* **4** (1979) 285
13. Both structure-descriptors $\chi = C(-1/2)$ and $\mu = C(-1)$ have very low isomerdiscriminating powers; this is, for instance, seen from the data given in Figs. 2 and 3
14. M. Randić, P. J. Hansen, P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **28** (1988) 60
15. E. Estrada, *J. Chem. Inf. Comput. Sci.* **35** (1995) 1022
16. D. Amić, D. Beslo, B. Lučić, S. Nikolić, N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **38** (1998) 819
17. E. Estrada, *Chem. Phys. Lett.* **312** (1999) 556
18. L. H. Clark, J. W. Moon, *Ars Combinatoria* **54** (1999) 223
19. I. Gutman, O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer-Verlag, Berlin, 1986
20. G. Caporossi, I. Gutman, P. Hansen, *Comput. Chem.* **23** (1999) 469
21. I. Gutman, O Miljković, G. Caporossi, P. Hansen, *Chem. Phys. Lett.* **306** (1999) 366.