

The Vertex-Connectivity Index Revisited

Dragan Amić and Drago Bešlo

Faculty of Agriculture, The Josip Juraj Strossmayer University, P.O. Box 117,
HR-31001 Osijek, The Republic of Croatia

Bono Lučić, Sonja Nikolić, and Nenad Trinajstić*

The Rugjer Bošković Institute, P. O. Box 1016, HR-10001 Zagreb, The Republic of Croatia

Received April 10, 1998

We report a search for optimum molecular descriptors based on the connectivity index. A suggestion made by several authors that the exponent -0.5 used in the standard formula for computing the connectivity index may not be the optimum for modeling some molecular properties was reexamined. We considered several molecular properties and found that in most cases the optimum value of the exponent is indeed different from -0.5 . We suggest that a modified version of the (valence) vertex-connectivity index should be routinely employed in the structure–property modeling instead of the standard version of the index.

INTRODUCTION

Years ago, Altenburg¹ proposed a modification of Randić's formula² for the first-order vertex-connectivity index. He suggested the use of numbers different than -0.5 for the exponent in the formula. Little attention was paid to this proposal³ until recently, when Estrada,⁴ in this journal, used Altenburg's modification in an attempt to remove the degeneracy of connectivity indices in regular graphs and graphs representing alkanes and cycloalkanes. He identified the exponent -0.5 as the main cause of the degeneracy of connectivity indices and found that the exponent -0.33 eliminates the degeneracy of indices in the case of the studied graphs. It is perhaps interesting to note that the exponent value of -0.33 is precisely the value reported by Randić et al.³ as the optimum exponent for modeling the boiling points of lower alkanes. We decided to reinvestigate this idea and use the modified (valence) vertex-connectivity index in several structure–property relationships. Our aim was to compare the performance of the obtained models with those already reported in the literature by the standard version of the vertex-connectivity index. In addition, we will also report the application of the modified valence vertex-connectivity index to retention times of phenolic acids and anthocyanidins. Throughout the work we will use (chemical) graph-theoretical concepts and terminology^{5,6} to simplify the analysis.

MODIFIED VERTEX-CONNECTIVITY INDEX

The standard definition of the vertex-connectivity index is:²

$$\chi = \sum_{\text{edges}} [d(i) d(j)]^{-0.5} \quad (1)$$

where $d(i)$ is the degree of the vertex i , and the quantity $[d(i) d(j)]^{-0.5}$ may be considered as the weight of the i – j edge.⁷ The summation in eq 1 goes over all edges. The vertex-degree d is equal to the number of adjacent vertexes

in a given (molecular) graph G . Two vertexes in G are adjacent if there is an edge joining them.

Because the assignment of bond weights is based on one possible solution to the inequalities based on ordering graphs, other selections of weights are possible. Hence, the quantity $[d(i) d(j)]^{-0.5}$ may be replaced by $[d(i) d(j)]^k$, and eq 1 reads as:

$$\chi = \sum_{\text{edges}} [d(i) d(j)]^k \quad k \neq 0 \quad (2)$$

where k is a variable exponent that can be varied from -0.1 to any desired value in steps of 0.1 .

In the same way, the valence vertex-connectivity index:⁸

$$\chi^v = \sum_{\text{edges}} [\delta(i) \delta(j)]^{-0.5} \quad (3)$$

can be modified:

$$\chi^v = \sum_{\text{edges}} [\delta(i) \delta(j)]^k \quad k \neq 0 \quad (4)$$

where $\delta(i)$ and $\delta(j)$ are weights (valence delta values) of vertexes i and j making up the i – j edge in a vertex-weighted graph (representing heterosystems).

RESULTS AND DISCUSSION

Boiling Points of 18 Octanes. A few years ago Randić and one of us (NT) reported a comparative structure–property study using the connectivity basis.⁹ In that work, among other results, the structure–boiling point modeling was reported for 18 octanes. The relevant statistical data for a single index model were $R = 0.821$ and $S = 3.60$ °C. We repeated this work using the modified vertex-connectivity index for the range of the k values from -0.1 to -2.0 .

In the present study we used the standard deviation S as a criterion for the comparison of the models. In Figure 1 we give the change of the standard deviation S against the exponent k of the modified vertex-connectivity index.

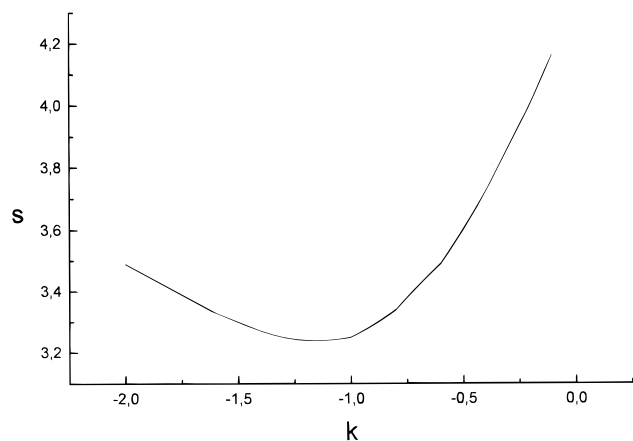


Figure 1. Boiling points of 18 octanes. A plot of the standard deviation S against the exponent k in the modified vertex-connectivity index.

The best correlation is obtained for $k = -1.15$ with statistical parameters $R = 0.859$ and $S = 3.24$ °C. The regression equation is given as follows:

$$\text{bp} = 65.14 (\pm 7.29) + 28.87 (\pm 4.31) \chi^{(-1.15)}$$

$$n = 18 \quad R = 0.859 \quad S = 3.24 \quad F = 44.8 \quad (5)$$

where bp is the boiling point and $\chi^{(-1.15)}$ is a short-hand notation for the molecular vertex-connectivity index computed using the value of -1.15 for the exponent in eq 2. Notation $\chi^{(k)}$ will be employed throughout this work.

The improvement over the earlier result is rather slight, but encouraged us to continue with this kind of analysis. It also confirms the observation by Randić et al.³ that the modified vertex-connectivity index produces better correlations with boiling points of alkanes than the standard version of the vertex-connectivity index. Here we mention that structure–boiling point modeling based on a single index, in general, does not produce particularly good correlations.^{10,11} Also, the boiling point is a difficult molecular property to model accurately.¹⁰ It should be also pointed out that when comparing the models just presented, one has to bear in mind that a simple linear regression using one independent variable (the connectivity index) has two adjusted parameters, whereas a corresponding regression using optimized $\chi^{(k)}$ is the result of adjusting three parameters.

Boiling Points of 35 Nonanes. In another study Randić and one of us (NT) reported structure-boiling point modeling for nonanes also using the connectivity basis.¹² The relevant statistical data for a single index model were $R = 0.684$ and $S = 4.54$ °C. We repeated this study in the same way as already discussed.

In Figure 2 we give the change of the standard deviation S against the exponent k of the vertex-connectivity index.

The best correlation is obtained for the $k = -1.3$ with statistical parameters $R = 0.747$ and $S = 4.14$ °C. The regression equation is given as:

$$\text{bp} = 96.10 (\pm 6.39) + 25.88 (\pm 4.02) \chi^{(-1.3)}$$

$$n = 35 \quad R = 0.747 \quad S = 4.14 \quad F = 41.5 \quad (6)$$

This time the improvement over the earlier result is slightly better. However, this new model is still far from being

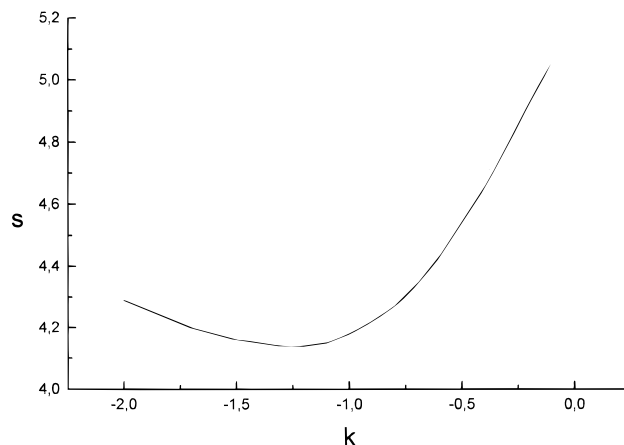


Figure 2. Boiling points of 35 nonanes. A plot of the standard deviation S against the exponent k in the modified vertex-connectivity index.

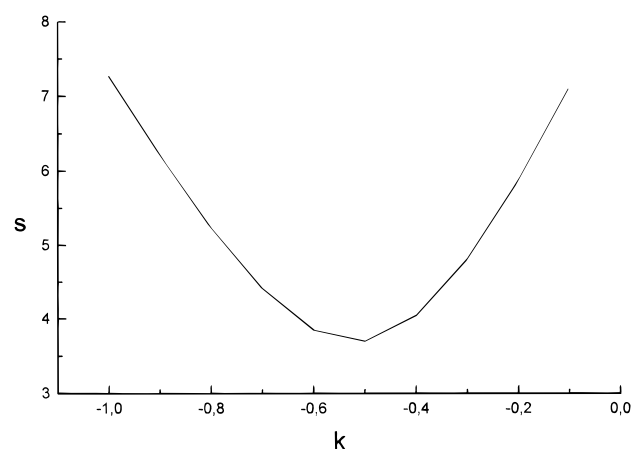


Figure 3. Boiling points of 20 cycloalkanes. A plot of the standard deviation S against the exponent k in the vertex-connectivity index.

satisfactory in comparison with models that include several indices. For example, the best model with six nonorthogonal connectivity indices has the following statistical parameters: $R = 0.987$ and $S = 1.05$ °C.¹³

Boiling Points of 20 Cycloalkanes. Plavšić et al.¹⁴ reported last year (1997) in this journal a structure–boiling point model for 20 cycloalkanes using the Z index of Hosoya¹⁵ and the vertex-connectivity index. The relevant statistical data for a model based on the connectivity index were $R = 0.992$ and $S = 3.70$ °C. We repeated this study in the same way as before, but for the range of the k values from -0.1 to -1.0 .

In Figure 3 we give the change of the standard deviation S against the exponent k of the vertex-connectivity index. The best correlation is also obtained for the $k = -0.5$ with statistical parameters $R = 0.992$ and $S = 3.70$ °C.

In this case, the previous and the present results agree. Furthermore, the obtained model is a fair one. The old-new regression equation is:

$$\text{bp} = -130.90 (\pm 5.48) + 69.28 (\pm 2.12) \chi^{(-0.5)}$$

$$n = 20 \quad R = 0.992 \quad S = 3.70 \quad F = 1068 \quad (7)$$

Solubility of 54 Aliphatic Alcohols. A few years ago Lučić et al.¹⁶ reported a structure–property modeling of the solubility of aliphatic alcohols in water. The relevant

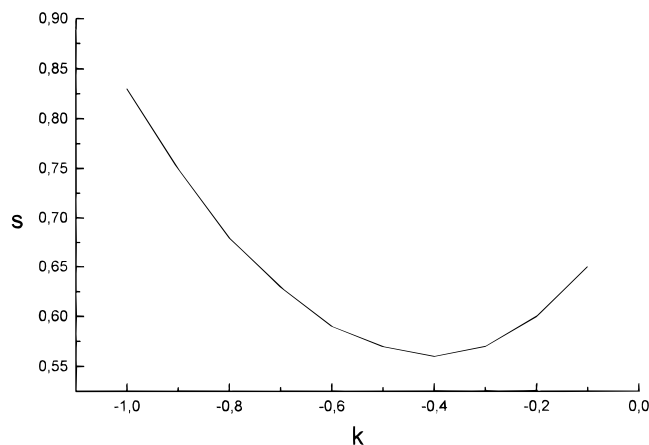


Figure 4. Solubility of 54 aliphatic alcohols. A plot of the standard deviation S against the exponent k in the modified valence vertex-connectivity index.

statistical parameters for a single index model were $R = 0.987$ and $S = 0.57$. We repeated this study using the modified vertex-connectivity index for the k values from -0.1 to -1.0 .

In Figure 4 we give the change of the standard deviation S against the exponent k of the valence vertex-connectivity index.

The best correlation is obtained for the $k = -0.4$ with statistical parameters $R = 0.988$ and $S = 0.56$. The regression equation is given as:

$$\ln \text{SOL} = 5.92 (\pm 0.22) - 2.36 (\pm 0.05) (\chi^v)^{(-0.4)}$$

$$n = 54 \quad R = 0.988 \quad S = 0.56 \quad F = 2074 \quad (8)$$

where $\ln \text{SOL}$ stands for a natural logarithm of aqueous solubility and $(\chi^v)^{(-0.4)}$ is a short-hand notation for the valence vertex-connectivity index computed using the value of -0.4 for the exponent in eq 4. The same notation will be also used in the next example.

The improvement over the earlier result is rather slight, but it also supports the idea that -0.5 is not always the optimum selection for the exponent of the valence vertex-connectivity index.

Retention Times of Phenolic Acids and Anthocyanidins.

Gao and Mazza¹⁷ recently reported retention times (in minutes) for 15 phenolic acids and anthocyanidins. We studied the structure-retention time relationship using the valence vertex-connectivity index with the exponent values in the range from -0.1 to -3.0 . In Figure 5 we give the change of the standard deviation S against the exponent k of the valence vertex-connectivity index.

The best correlation is obtained for $k = -1.6$. The corresponding regression model is given by:

$$\text{RT} = -12.96 (\pm 2.15) + 112.20 (\pm 6.95) (\chi^v)^{(-1.6)}$$

$$n = 15 \quad R = 0.976 \quad S = 2.71 \quad F = 260.5 \quad (9)$$

where RT is the retention times. At first sight, eq 9 appears to be a sound model, but there is a problem with this model: it does not correctly order retention times. We are aware that it is not strictly required that a model (even with good R and S values) correctly orders every pair of (close) points. Nevertheless, we decided to see whether we can find

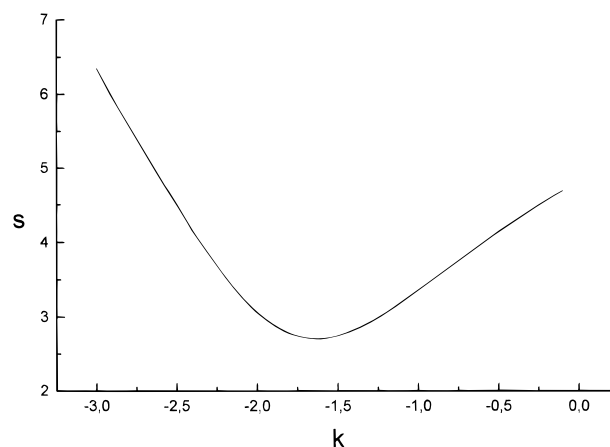


Figure 5. Retention times of phenolic acids and anthocyanidins. A plot of the standard deviation S against the exponent k in modified the valence vertex-connectivity index.

a model that will correctly order all retention times of phenolic acids and anthocyanidins without treating separately these two groups of compounds.

In an attempt to improve model we added to it the sum of π -electron charges Q_{ph} on the carbon atoms in the phenyl moiety obtained from the HMO method and have computed the retention times again using the valence vertex-connectivity index with the exponent values in the range from -0.1 to -3.0 . We achieved an interesting result. Several models had better statistical parameters than model in eq 9, but only a model given next has produced the correct order of retention times:

$$\text{RT} = 137.15 (\pm 50.76) + 103.96 (\pm 6.17) (\chi^v)^{(-1.6)}$$

$$-24.03 (\pm 8.12) Q_{\text{ph}}$$

$$n = 15 \quad R = 0.986 \quad S = 2.15 \quad F = 212.3 \quad (10)$$

However, this was not the best model according to statistical parameters. For example, a model with best parameters ($R = 0.988$, $S = 1.99$, $F = 248$) was obtained for $k = -1.9$, but the computed retention times (in minutes) for vanilic and caffeic acids (11.5, 11.0) disagree with experiment. The same result was observed for four more models with slightly better statistical parameters than eq 10 ($k = -1.7$: $R = 0.987$, $S = 2.05$, $F = 233.0$, 11.1, 11.1; $k = -1.8$: $R = 0.988$, $S = 2.00$, $F = 245.0$, 11.3, 11.1; $k = -2.0$: $R = 0.987$, $S = 2.04$, $F = 234.6$, 11.8, 10.9; $k = -2.1$: $R = 0.986$, $S = 2.14$, $F = 214.1$, 12.1, 10.8).

Calculated valence vertex-connectivity indices with $k = -1.6$, phenyl moiety π -electron charges, the experimental retention times $(\text{RT})_{\text{exp}}$ and calculated retention times $(\text{RT})_{\text{calc}}$ are given in Table 1.

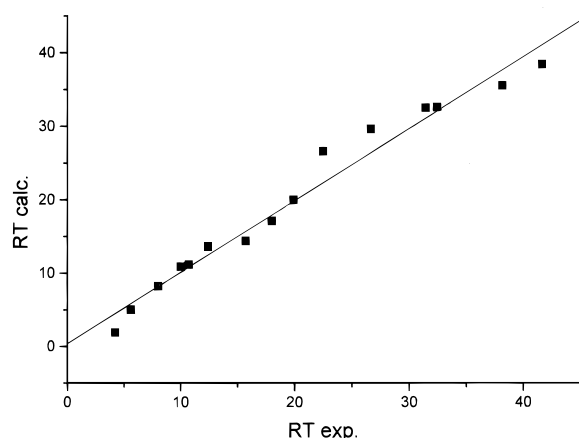
In Figure 6 we give a plot of $(\text{RT})_{\text{exp}}$ versus $(\text{RT})_{\text{calc}}$.

CONCLUDING REMARKS

This work shows that the exponent of -0.5 that is used in the standard definition of the (valence) vertex-connectivity index may not be the optimum exponent for modeling some properties and agrees with observations by Randić et al.³ and Estrada.⁴ Therefore, those who use the (valence) vertex-connectivity index in the structure-property modeling should also consider carrying out their modeling with the modified

Table 1. Calculated Modified Valence Vertex-Connectivity Indices ($\chi^{vy(-1.6)}$), Total Phenyl Moiety π -Electron Charges Q_{ph} , and the Experimental (RT)_{exp} and Calculated (RT)_{calc} Retention Times (in minutes)

phenolic acid/ anthocyanidin	$(\chi^{vy})^{(-1.6)}$	Q_{ph}	(RT) _{exp}	(RT) _{calc}
gallic acid	0.1499	6.2760	4.2	1.9
protocatechuic acid	0.1595	6.1901	5.6	5.0
<i>p</i> -hydroxybenzoic acid	0.1691	6.1002	8.0	8.2
vanillic acid	0.2143	6.1814	10.0	10.9
caffeic acid	0.2149	6.1705	10.7	11.2
syringic acid	0.2595	6.2593	12.4	13.7
<i>p</i> -coumaric acid	0.2245	6.0806	15.7	14.4
ferulic acid	0.2697	6.1619	18.0	17.1
sinapic acid	0.3149	6.2400	19.9	20.0
delphinidin	0.3586	6.1541	22.5	26.6
cyanidin	0.3682	6.0713	26.7	29.6
petunidin	0.4134	6.1460	31.5	32.5
pelargonidin	0.3778	5.9844	32.5	32.6
peonidin	0.4230	6.0628	38.2	35.5
malvidin	0.4682	6.1379	41.7	38.4

**Figure 6.** Retention times of phenolic acids and anthocyanidins. A plot of (RT)_{exp} versus (RT)_{calc}. (RT)_{calc} values are obtained using eq 10.

(valence) vertex-connectivity index. The validity of such a suggestion is rather nicely exemplified in our work on modeling the retention times of phenolic acids and anthocyanidins. When $k = -0.5$ was used, the structure-retention time model including Q_{ph} ($R = 0.954$, $S = 3.88$, $F = 60.8$) was much poorer in comparison with eq 10. We also intended with this report to draw attention to the problem of detecting optimum molecular descriptors, a topic that has not yet received enough attention.^{18,19}

ACKNOWLEDGMENT

This work was supported by the Ministry of Science and Technology of the Republic of Croatia through Grants 079301 (DA, DB) and 00980606 (BL, SN, NT). We thank the reviewers for their helpful suggestions.

REFERENCES AND NOTES

- (1) Altenburg, K. Eine Bemerkung zu dem Randićschen "Molekularen Bindungs-Index (Molecular Connectivity Index)". *Z. Phys. Chem. (Leipzig)* **1980**, *261*, 389–393.
- (2) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (3) Randić, M.; Hansen, P. J.; Jurs, P. C. Search for Useful Graph Theoretical Invariants of Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 60–68.
- (4) Estrada, E. Graph Theoretical Invariant of Randić Revisited. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1022–1025.
- (5) Harary, F. *Graph Theory*, 2nd printing; Addison-Wesley: Reading, MA, 1972.
- (6) Trinajstić N. *Chemical Graph Theory*, 2nd revised ed.; CRC: Boca Raton, FL, 1992.
- (7) Randić, M.; Jeričević, Ž.; Sabljčić, A.; Trinajstić, N. On the Molecular Connectivity and π -Electronic Energy in Polycyclic Hydrocarbons. *Acta Phys. Polon.* **1988**, *74*, 317–330.
- (8) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Property Analysis*. Wiley: Chichester, 1986.
- (9) Randić, M.; Trinajstić, N. Viewpoint 4 – Comparative Structure-Property Studies: The Connectivity Basis. *J. Mol. Struct. (Theochem)* **1993**, *284*, 209–221.
- (10) Randić, M. Comparative Regression Analysis. Regressions Based on a Single Descriptor. *Croat. Chem. Acta* **1993**, *66*, 289–312.
- (11) Randić, M.; Mihalić, Z.; Nikolić, S.; Trinajstić, N. Graph-Theoretical Correlations – Artifacts or Facts. *Croat. Chem. Acta* **1993**, *66*, 411–434.
- (12) Randić, M.; Trinajstić, N. Isomeric Variations in Alkanes: Boiling Points of Nonanes. *New J. Chem.* **1994**, *18*, 179–189.
- (13) Lučić, B.; Trinajstić, N. New Developments in QSPR/QSAR Modeling Based on Topological Indices. *SAR & QSAR Environ. Res.* **1997**, *7*, 45–62.
- (14) Plavšić, D.; Soškić, M.; Đaković, Z.; Gutman, I.; Graovac, A. Extension of the Z Matrix to Cycle-Containing and Edge-Weighted Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 529–534.
- (15) Hosoya, H. Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
- (16) Lučić, B.; Nikolić, S.; Trinajstić, N.; Jurić, A.; Mihalić, Z. A Structure-Property Study of the Solubility of Aliphatic Alcohols in Water. *Croat. Chem. Acta* **1995**, *68*, 417–434.
- (17) Gao, L.; Mazza, G. Rapid Method for Complete Chemical Characterization of Simple and Acylated Anthocyanins by High-Performance Liquid Chromatography and Capillary Gas-Liquid Chromatography. *J. Agric. Food Chem.* **1994**, *42*, 118–125.
- (18) Randić, M. On Computation of Optimal Parameters for Multivariate Analysis of Structure-Property Relationship. *J. Comput. Chem.* **1991**, *12*, 970–980.
- (19) Randić, M. Linear Combinations of Path Numbers as Molecular Descriptors. *New J. Chem.* **1997**, *21*, 945–951.

CI980039B