

Computing the Variable Vertex-connectivity Index*

Nenad Kezele,^a Leo Klasinc,^{a,**} Jan von Knop,^b Sanja Ivaniš,^c
and Sonja Nikolić^a

^a *The Rugjer Bošković Institute, P. O. Box 180, HR–10002 Zagreb, Croatia*

^b *The Computer Center, The Heinrich Heine University,
D–40225 Düsseldorf 1, Germany*

^c *Belupo d.o.o., HR–43300 Koprivnica, Croatia*

Received December 12, 2000; revised April 11, 2002; accepted April 12, 2002

A computer program has been designed for efficacious computation of the variable connectivity index. It has been tested on Randić's and Basak's work²⁰ on aliphatic primary amines. We have reproduced their results for the fit model. We have also cross-validated (CV) this model. Both the fit and the CV models possess close (comparable) statistical parameters. It appears that the use of the variable connectivity index in QSPR yields very good regression equations in the case of homogeneous sets of molecules.

Key words: aliphatic amines, connectivity index, vertex-connectivity index, variable vertex-connectivity index, QSPR.

INTRODUCTION

Milan Randić introduced in 1975 a bond-additive topological index as a descriptor for characterization of molecular branching,¹ which he called the branching index and denoted it by χ . This index was soon after renamed into the connectivity index.² We call Randić's index the *vertex-connectivity index*^{3,4} to distinguish it from the *edge-connectivity index*, introduced

* Dedicated to Professor Milan Randić on the happy occasion of his 70th birthday.

** Author to whom correspondence should be addressed. (E-mail: klasinc@joker.irb.hr)

twenty years later by Estrada.⁵ The vertex-connectivity index has become the most used topological index in QSPR⁶ and QSAR⁷ studies.⁸⁻¹²

In 1991, Randić introduced the *variable* vertex-connectivity index¹³ as an alternative approach to Kier and Hall's *valence* vertex-connectivity index¹⁴ for characterization of heterosystems in QSPR and QSAR studies. The difference between the variable vertex-connectivity index and the valence vertex-connectivity index is in that the former index uses *optimized* vertex-weights and the latter index uses *fixed* vertex-weights. For example, in the case of aliphatic amines, the fixed weight for nitrogen in the primary amino group is 3, in the secondary amino group 4 and in the tertiary amino group 5, and the fixed weights for carbon are 1, 2, 3 or 4 depending on the bonding environment of carbon.

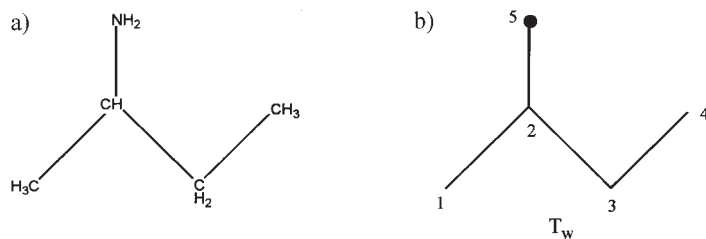
The variable vertex-connectivity index remained unnoticed for a number of years, probably because it was published in a less visible journal. To remedy this, Randić and co-workers produced in the last few years a number of papers using this index.¹⁵⁻¹⁹ To support their efforts, we produced an iterative least-squares procedure for efficacious computation of the variable vertex-connectivity index. We tested our procedure against Randić's and Basak's results for boiling points of smaller aliphatic amines,²⁰ because in that article they described in detail the procedure for generating variable connectivity indices. In addition, we have considered the boiling points of primary amines, since they have recently been rather thoroughly studied by Pogliani using molecular pseudoconnectivity indices.²¹ In the past, aliphatic amines were also studied by Kier and Hall.^{22,23}

VARIABLE VERTEX-CONNECTIVITY INDEX

The fixed (valence) vertex-connectivity index $\chi^{(v)}$ can be calculated, following suggestions by Balaban,²⁴ from the row-sums of the vertex-adjacency matrix $[(rs)_i, (rs)_j]$ using the inverse square-root algorithm making the contribution $[(rs)_i (rs)_j]^{-0.5}$ to the i - j bond. We call this matrix the *vertex*-adjacency matrix to differentiate it from the *edge*-adjacency matrix.²⁵

The variable vertex-connectivity index can also be obtained by the inverse square-root algorithm using the row-sums of the augmented (weighted) vertex-adjacency matrix. Instead of zeros, the augmented vertex-adjacency matrix has variables X, Y, Z , etc. on the main diagonal. However, a variable is added to each row-sum, thus producing bond weights $\{[(rs)_i + X] [(rs)_j + Y]\}^{-0.5}$, where X and Y are variables associated with atoms x and y , respectively. Note that, since the variable vertex-connectivity index is a function of variables, it has been denoted by the superscript f , χ^f , in order to distinguish it from the fixed valence vertex-connectivity index χ^v . For exam-

ple, in the case of aliphatic amines, variables are $X = \text{carbon } (c)$ and $Y = \text{nitrogen } (n)$. The vertex (atom)-weights are not fixed any more, they are now flexible to allow one to obtain the optimized weights leading to QSPR/QSAR models with the lowest possible errors of estimate. In Figure 1, we show the construction of the variable vertex-connectivity index for 2-aminobutane.



$$\mathbf{A}_w = \begin{bmatrix} c & 1 & 0 & 0 & 0 \\ 1 & c & 1 & 0 & 1 \\ 0 & 1 & c & 1 & 0 \\ 0 & 0 & 1 & c & 0 \\ 0 & 1 & 0 & 0 & n \end{bmatrix} \begin{matrix} 1+c \\ 3+c \\ 2+c \\ 1+c \\ 1+n \end{matrix} \text{ rs}$$

d)

$$\chi^f = [(1+c)(3+c)]^{-0.5} + [(3+c)(2+c)]^{-0.5} + [(3+c)(1+n)]^{-0.5} + [(2+c)(1+c)]^{-0.5}$$

Figure 1. Construction of the variable vertex-connectivity index χ^f for 2-aminobutane represented by the labeled weighted hydrogen-depleted tree T_w . The augmented (weighted) vertex-adjacency matrix \mathbf{A}_w and the row-sums used for calculating χ^f are shown. For convenience, variables X and Y are replaced with c (for carbon) and n (for nitrogen). (a) 2-Aminobutane; (b) Labeled weighted tree. Black dot denotes the position of nitrogen; (c) Augmented vertex-adjacency matrix and the row-sums (rs); (d) Variable vertex-connectivity index.

ITERATIVE LEAST-SQUARES PROGRAM FOR VARIABLE VERTEX-CONNECTIVITY INDICES

The variable vertex-connectivity index is to be determined for each regression analysis through an optimization procedure that, if performed us-

ing the pedestrian approach (that is, varying one variable at a time, as Randić and Basak did) could be both time-consuming and missing the optimal answer. Here, we describe an iterative least-squares program that optimizes the χ^f values in relation to a certain property.

The program optimizes the variables X, Y, Z, W of the variable vertex-connectivity indices χ^f (named C_i in the program) listed in the subroutine FUN (X, Y, Z, W) with respect to a set of experimental values E_i using a modified iterative Newton's least-squares method. Since the vertex-connectivity indices are in the subroutine, the program can handle any form of the nonlinear molecular descriptor with up to 4 variables using the same procedure.

The program starts with reading the number of variables, their starting values, their increment for calculating the partial derivative, a starting value for the sum of deviation squares (must be greater than the first one being calculated) and the accuracy of the minimal sum of deviation squares.

In the linear mode, it calculates the C_i 's using the starting values of the variables and determines the slope a and intercept b of the regression line between C_i 's and E_i 's.

The optimization procedure will be outlined for two variables, X and Y , for which the starting values were chosen to be X_0 and Y_0 . Using Taylor's expansion for X_0, Y_0 to get corrections X_1 and Y_1 for the function $C^{\text{corr}}(X, Y)$, *i.e.*, $X = X_0 + X_1$ and $Y = Y_0 + Y_1$, yields:

$$C^{\text{corr}}(X, Y) = C(X_0, Y_0) + (\delta C / \delta X_0) X_1 + (\delta C / \delta Y_0) Y_1 \quad (1)$$

with

$$\delta C^{\text{corr}} / \delta X_0 = P \text{ and } \delta C^{\text{corr}} / \delta Y_0 = Q \quad (2)$$

or for the i -th value:

$$C_i^{\text{corr}} = C_i + P_i X_1 + Q_i Y_1 \quad (3)$$

E_i should depend linearly on $C^{\text{corr}}(X, Y)$:

$$E_i^{\text{calc}} = a C_i + a P_i X_1 + a Q_i Y_1 + b \quad (4)$$

and the sum of deviation squares function:

$$\Phi(X_1, Y_1) = \sum (E_i - E_i^{\text{calc}})^2 = \sum (E_i - a C_i - a P_i X_1 - a Q_i Y_1 - b)^2 \quad (5)$$

will be minimum for:

$$\delta\Phi/\delta X_1 = \delta\Phi/\delta Y_1 = 0 \tag{6}$$

yielding:

$$\delta\Phi/\delta X_1 = a\Sigma E_i P_i - a^2\Sigma C_i P_i - a^2 X_1 \Sigma P_i^2 - a^2 Y_i \Sigma P_i Q_i - ab\Sigma P_i = 0 \tag{7}$$

$$\delta\Phi/\delta Y_1 = a\Sigma E_i Q_i - a^2\Sigma C_i Q_i - a^2 Y_1 \Sigma Q_i^2 - a^2 X_i \Sigma P_i Q_i - ab\Sigma Q_i = 0 \tag{8}$$

or:

$$\begin{pmatrix} \Sigma P_i^2 & \Sigma P_i Q_i \\ \Sigma P_i Q_i & \Sigma Q_i^2 \end{pmatrix} \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} = \begin{pmatrix} 1/a(\Sigma E_i P_i - b\Sigma P_i) - \Sigma C_i P_i \\ 1/a(\Sigma E_i Q_i - b\Sigma Q_i) - \Sigma C_i Q_i \end{pmatrix} \tag{9}$$

which allows calculation of X_1 and Y_1 and obtaining X and Y , from which the new slope a and intercept b are determined. The program then sets $X_0 = X$ and $Y_0 = Y$ and repeats the process until convergence to the desired accuracy limit of the deviation squares sum. All the partial derivatives for the variable vertex-connectivity indices are calculated using the input increment d by:

$$\delta C^{corr}/\delta X = [C(X + d \cdot X, Y) - C(X, Y)]/d \cdot X \tag{10}$$

and:

$$\delta C^{corr}/\delta Y = [C(X, Y + d \cdot Y) - C(X, Y)]/d \cdot Y \tag{11}$$

In the quadratic mode (here shown only for one variable, X) from the initial C_i 's and E_i 's, the optimal coefficients a, b, c of the function $E_i^{calc} = aC_i^2 + bC_i + c$ are determined by the least-squares regression and entered into the Taylor expanded sum of deviation squares function:

$$\Phi(X_1) = \Sigma(E_i - aC_i^2 - 2aC_i P_i X_1 - aP_i^2 X_1^2 - bC_i - bP_i X_1 - c)^2 \tag{12}$$

which will be minimum for:

$$\begin{aligned} d\Phi/dX_1 = 0 = & (a\Sigma P^4)X_1^3 + (3a\Sigma CP^3 + 1.5\Sigma P^3)X_1^2 + \\ & (a\Sigma C^2 P^2 + b\Sigma CP^2 + b/2a\Sigma P^2 + 2a\Sigma C^2 P^2 + 2b\Sigma CP^2 + c\Sigma P^2 + \Sigma P^2 E)X_1 + \\ & a\Sigma C^3 P + b\Sigma C^2 P + c\Sigma CP - \Sigma CPE - b/2a\Sigma PE + b/2\Sigma C^2 P + b^2/2a\Sigma CP = \\ & a_1 X_1^3 + a_2 X_1^2 + a_3 X_1 + a_4 \end{aligned} \tag{13}$$

The real root of this equation is the correction X_1 for X , which then replaces X_0 as in the linear mode. The polynomial root(s) are calculated in the subroutine NULT (A1, A2, A3, A4, X1) using an iterative procedure. Additional information about the program, written in FORTRAN, can be obtained from Nenad Kezele (nenad@joker.irb.hr).

TABLE I

The variable vertex-connectivity indices (χ^f) and boiling points (bp/K) of 16 aliphatic primary amines considered in this report

No.	Primary amine	bp/K	χ^f
1	Nonylamine	474.15	$1/\sqrt{(1+c)(2+c)} + 7/\sqrt{(2+c)\times(2+c)} + 1/\sqrt{(2+c)(1+n)}$
2	Octylamine	453.15	$1/\sqrt{(1+c)(2+c)} + 6/\sqrt{(2+c)\times(2+c)} + 1/\sqrt{(2+c)(1+n)}$
3	Heptylamine	428.15	$1/\sqrt{(1+c)(2+c)} + 5/\sqrt{(2+c)\times(2+c)} + 1/\sqrt{(2+c)(1+n)}$
4	Hexylamine	403.15	$1/\sqrt{(1+c)(2+c)} + 4/\sqrt{(2+c)\times(2+c)} + 1/\sqrt{(2+c)(1+n)}$
5	4-Methylpentylamine	398.15	$2/(2+c) + 2/\sqrt{(1+c)(3+c)} + 1/\sqrt{(2+c)(3+c)} + 1/\sqrt{(2+c)\times(1+n)}$
6	2-Aminohexane	387.65	$1/\sqrt{(1+c)(2+c)} + 1/\sqrt{(1+c)\times(3+c)} + 2/\sqrt{(2+c)(2+c)} + 1/\sqrt{(3+c)(2+c)}$
7	Pentylamine	377.15	$1/\sqrt{(1+c)(2+c)} + 3/\sqrt{(2+c)\times(2+c)} + 1/\sqrt{(2+c)(1+n)}$
8	2-Methylbutylamine	369.15	$2/\sqrt{(2+c)(3+c)} + 1/\sqrt{(1+c)\times(2+c)} + 1/\sqrt{(1+c)(3+c)} + 1/\sqrt{(2+c)(1+n)}$
9	3-Methylbutylamine	369.15	$2/\sqrt{(1+c)(3+c)} + 1/\sqrt{(3+c)\times(2+c)} + 1/\sqrt{(2+c)(1+c)} + 1/\sqrt{(2+c)(1+n)}$
10	2-Aminopentane	365.15	$1/\sqrt{(1+c)(3+c)} + 1/\sqrt{(3+c)\times(2+c)} + 1/\sqrt{(2+c)(2+c)} + 1/\sqrt{(1+c)(2+c)} + 1/\sqrt{(3+c)\times(1+n)}$
11	3-Aminopentane	364.15	$2/\sqrt{(1+c)(2+c)} + 2/\sqrt{(2+c)\times(3+c)} + 1/\sqrt{(3+c)(1+n)}$

TABLE I (cont.)

No.	Primary amine	bp/K	χ^f
12	2-Amino-2-methylbutane	351.15	$2/\sqrt{(1+c)(4+c)} + 1/\sqrt{(4+c)(2+c)} + 1/\sqrt{(2+c)(1+c)} + 1/\sqrt{(4+c)(1+n)}$
13	Butylamine	350.15	$1/\sqrt{(1+c)(2+c)} + 2/\sqrt{(2+c)(2+c)} + 1/\sqrt{(2+c)(1+n)}$
14	2-Methylpropylamine	342.15	$2/\sqrt{(1+c)(3+c)} + 1/\sqrt{(3+c)(2+c)} + 1/\sqrt{(2+c)(1+n)}$
15	2-Aminobutane	336.15	$1/\sqrt{(1+c)(3+c)} + 1/\sqrt{(3+c)(2+c)} + 1/\sqrt{(2+c)(1+c)} + 1/\sqrt{(3+c)(1+n)}$
16	Propylamine	322.15	$1/\sqrt{(1+c)(2+c)} + 1/\sqrt{(2+c)(2+c)} + 1/\sqrt{(2+c)(1+n)} + 1/\sqrt{(3+c)(1+n)}$

APPLICATION TO PRIMARY AMINES

Table I gives explicit expressions for the variable vertex-connectivity indices for 16 primary aliphatic amines considered by Randić and Basak.²⁰

Variables c (1.25) and n (-0.65) were reproduced as corresponding to the optimal values of χ^f for the boiling points of the primary amines considered, the standard error of estimate (S ; $N - I - 1$ in the denominator, where N = the number of molecules and I = the number of descriptors) being 1.907 K. Randić and Basak achieved their results using a stepwise procedure with 21 steps. Our procedure produced the same results within seconds. We give explicitly the fit (descriptive) model²⁶ because Randić and Basak did not give it in their article:

$$\text{bp} = 189.16 (\pm 2.29) + 83.45 (\pm 0.98) \chi^f$$

$$N = 16 \quad R_{\text{fit}} = 0.999 \quad S_{\text{fit}} = 1.907 \quad F = 7296.2 \quad (14)$$

where bp stands for the boiling point while R_{fit} and S_{fit} stand for the correlation coefficient and the standard error of estimate, respectively, of the fit model.

In addition to Randić's and Basak's investigations, we also carried out internal (cross) validation of this model using the leave-one-out procedure.

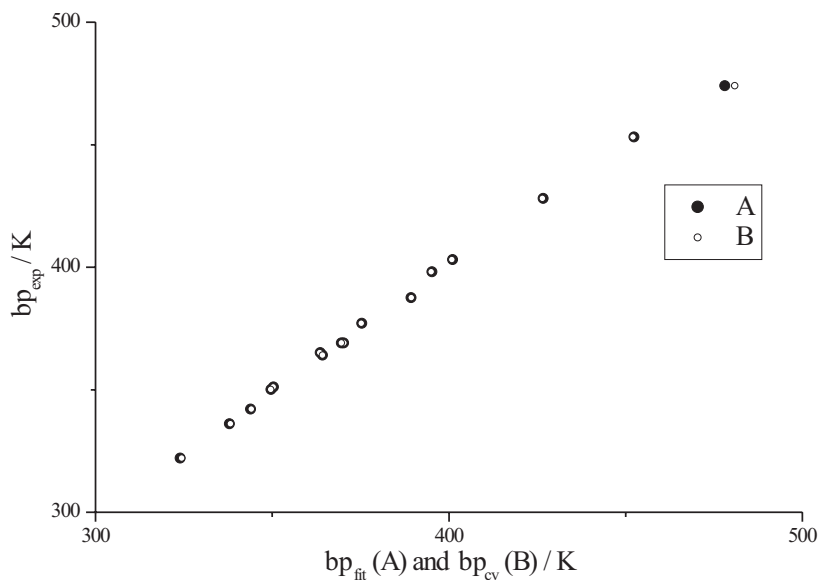


Figure 2. A plot of the experimental boiling points of aliphatic amines (bp_{exp}) versus fit (bp_{fit}), denoted by filled-in circles (A), and cross-validated boiling points (bp_{cv}), indicated by empty circles.

The statistical parameters obtained for the cross-validated model, $R_{\text{cv}} = 0.998$ and $S_{\text{cv}} = 1.907$ (cv stands for cross validation), support the above model. Figure 2 contains the plot of experimental (bp_{exp}) versus fit (bp_{fit}) and (bp_{cv}) boiling points of 16 aliphatic amines.

In Table II, we give the values of the variable connectivity indices (χ^f) computed using parameters $c = 1.25$ and $n = -0.65$, experimental (bp_{exp}) and computed (bp_{fit} , bp_{cv}) boiling points, and the differences between the experimental and computed boiling points.

In what follows, we will mention the results of other researchers, although they used somewhat larger sets of aliphatic amines and slightly different values for boiling points. We do this to show the advantage of using variable descriptors in QSPR modeling. Pogliani considered 21 primary aliphatic amines.²¹ He used molecular pseudoconnectivity indices and his best structure-boiling point models with one-, two- and three-indices have the following S values: 5, 3.4 and 3.1, respectively. The best model that Pogliani obtained has $S = 2.4$. Kier and Hall²³ obtained the best model ($S = 3.1$) with a mixed set of four indices. It is interesting to note that the simple vertex-connectivity index offers a comparable correlation ($S = 3.49$) to those obtained by Kier and Hall and Pogliani. The final point to make is that their models are descriptive or fit models; they are not predictive models.²⁶

TABLE II

The variable connectivity index χ^f for 16 aliphatic amines, their experimental $(bp)_{exp}$, cross-validated $(bp)_{cv}$ and fit $(bp)_{fit}$ boiling points, and the differences between experimental and calculated boiling points $(bp)_{exp} - (bp)_{fit}$, $(bp)_{exp} - (bp)_{cv}$

No.	$\frac{(bp)_{exp}}{K}$	χ^f	$\frac{(bp)_{cv}}{K}$	$\frac{(bp)_{fit}}{K}$	$\frac{(bp)_{exp} - (bp)_{cv}}{K}$	$\frac{(bp)_{exp} - (bp)_{fit}}{K}$
1	474.15	3.4613	480.82	478.02	-6.67	-3.87
2	453.15	3.1536	452.07	452.34	1.09	0.81
3	428.15	2.8459	426.42	426.66	1.73	1.49
4	403.15	2.5382	400.80	400.99	2.35	2.17
5	398.15	2.4688	394.97	395.19	3.18	2.96
6	387.65	2.3976	389.36	389.25	-1.71	-1.60
7	377.15	2.2305	375.18	375.31	1.97	1.84
8	369.15	2.1689	370.24	370.17	-1.09	-1.02
9	369.15	2.1611	369.54	369.52	-0.39	-0.37
10	365.15	2.0899	363.45	363.57	1.70	1.58
11	364.15	2.0977	364.23	364.22	-0.08	-0.07
12	351.15	1.9315	350.27	350.35	0.88	0.80
13	350.15	1.9228	349.57	349.63	0.58	0.52
14	342.15	1.8534	344.05	343.84	-1.90	-1.69
15	336.15	1.7822	338.16	337.89	-2.01	-1.74
16	322.15	1.6151	324.35	323.95	-2.20	-1.80

CONCLUDING REMARKS

We have produced a computer program for the computation of variable connectivity indices. The program was tested on Randić's and Basak's article,²⁰ in which, among other things, they produced the fit structure-boiling point for aliphatic primary amines. We reproduced their results and additionally computed the cross-validated model. In their article, Randić and Basak did not compute cross-validated models – they produced only fit models in all the studied cases. Therefore, there still remain the open questions about the use of the variable connectivity index in predictive models and in the case of non-homogeneous sets of molecules. These questions have been already discussed by Zefirov and Palyulin.²⁶

Acknowledgements. – This work was supported in part by the Ministry of Science and Technology of the Republic of Croatia *via* Grant No. 00980601 (NK, LK)

and Grant No. 00980606 (SN). We thank Dr Bono Lučić and Professor Milan Randić for comments.

REFERENCES

1. M. Randić, *J. Am. Chem. Soc.* **97** (1975) 6609–6615.
2. L. B. Kier, L. H. Hall, W. J. Murray, and M. Randić, *J. Pharm. Sci.* **64** (1975) 1971–1974.
3. N. Trinajstić, S. Nikolić, D. Babić, and Z. Mihalić, *Bull. Chem. Technol. Macedonia* **16** (1997) 43–51.
4. S. Nikolić, N. Trinajstić, and I. Baučić, *J. Chem. Inf. Comput. Sci.* **38** (1998) 42–46.
5. E. Estrada, *J. Chem. Inf. Comput. Sci.* **35** (1995) 31–33.
6. QSPR = quantitative structure-property relationships.
7. QSAR = quantitative structure-activity relationships.
8. L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
9. L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis*, Research Studies Press/Wiley, New York, 1986.
10. J. Devillers and A. T. Balaban (Eds.), *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon & Breach Sci. Publ., Amsterdam, 1999.
11. R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, 2000.
12. M. V. Diudea (Ed.), *QSPR/QSAR Studies by Molecular Descriptors*, Nova Sci. Publ., Huntington, N.Y., 2001.
13. M. Randić, *Chemometrics Intel. Lab. Syst.* **10** (1991) 213–227.
14. L. B. Kier and L. H. Hall, *J. Pharm. Sci.* **65** (1976) 1806–1809.
15. M. Randić and J. C. Dobrowolski, *Int. J. Quantum Chem.* **70** (1998) 1209–1215.
16. M. Randić and S. C. Basak, *J. Chem. Inf. Comput. Sci.* **40** (2000) 899–905.
17. M. Randić, D. Mills, and S. C. Basak, *Int. J. Quantum Chem.* **80** (2000) 1199–1209.
18. M. Randić and S. C. Basak, *J. Chem. Inf. Comput. Sci.* **41** (2001) 614–618.
19. M. Randić and M. Pompe, *J. Chem. Inf. Comput. Sci.* **41** (2001) 631–638.
20. M. Randić and S. C. Basak, in: D. K. Sinha, S. C. Basak, R. K. Mohanty, and I. N. Busamallick (Eds.), *Some Aspects of Mathematical Chemistry*, Visva-Bharati University Press, Santiniketan, India, in press
21. L. Pogliani, *J. Chem. Inf. Comput. Sci.* **41** (2001) 836–841.
22. Ref. 8, pp. 137–139.
23. L. B. Kier and L. H. Hall, *Molecular Structure Description. The Electrotopological State*, Academic, New York, 1999, pp. 176–177.
24. A. T. Balaban, *Chem. Phys. Lett.* **89** (1982) 399–404.
25. N. Trinajstić, *Chemical Graph Theory*, 2nd revised edition, CRC Press, Boca Raton, FL, 1992, pp. 41–44.
26. N. S. Zefirov and V. A. Palyulin, *J. Chem. Inf. Comput. Sci.* **41** (2001) 1022–1027.

SAŽETAK**O računanju varijabilnog indeksa povezanosti čvorova**

Nenad Kezele, Leo Klasinc, Jan von Knop, Sanja Ivaniš i Sonja Nikolić

Priređen je računalni program za djelotvorno računanje varijabilnoga indeksa povezanosti. Program je testiran na radu Randića i Basaka²⁰ u kojem je dan ugođeni model za predviđanje točke vrelišta alifatskih primarnih amina. Reproduciravši njihove rezultate, provjerili smo kvalitetu ugođenoga modela s pomoću unakrižno vrednovanih parametara, što nisu učinili Randić i Basak. Po svemu sudeći, varijabilni indeks povezanosti vrlo je pogodan za izvođenje dobrih regresijskih modela u slučaju homolognih skupova molekula.