

QSPR Modeling: Graph Connectivity Indices versus Line Graph Connectivity Indices<sup>†</sup>Subhash C. Basak, Sonja Nikolić,<sup>‡</sup> and Nenad Trinajstić<sup>\*,‡</sup>

Natural Resources Research Institute, University of Minnesota, Duluth, Minnesota 55811

Dragan Amić and Drago Bešlo

Faculty of Agriculture, The Josip Juraj Strossmayer University, HR-31001 Osijek, The Republic of Croatia

Received September 9, 1999

Five QSPR models of alkanes were reinvestigated. Properties considered were molecular surface-dependent properties (boiling points and gas chromatographic retention indices) and molecular volume-dependent properties (molar volumes and molar refractions). The vertex- and edge-connectivity indices were used as structural parameters. In each studied case we computed connectivity indices of alkane trees and alkane line graphs and searched for the optimum exponent. Models based on indices with an optimum exponent and on the standard value of the exponent were compared. Thus, for each property we generated six QSPR models (four for alkane trees and two for the corresponding line graphs). In all studied cases QSPR models based on connectivity indices with optimum exponents have better statistical characteristics than the models based on connectivity indices with the standard value of the exponent. The comparison between models based on vertex- and edge-connectivity indices gave in two cases (molar volumes and molar refractions) better models based on edge-connectivity indices and in three cases (boiling points for octanes and nonanes and gas chromatographic retention indices) better models based on vertex-connectivity indices. Thus, it appears that the edge-connectivity index is more appropriate to be used in the structure–molecular volume properties modeling and the vertex-connectivity index in the structure–molecular surface properties modeling. The use of line graphs did not improve the predictive power of the connectivity indices. Only in one case (boiling points of nonanes) a better model was obtained with the use of line graphs.

## INTRODUCTION

This study was motivated by two recent papers. In one Estrada and Rodriguez<sup>1</sup> have shown that the edge-connectivity index produced the best single-variable QSPR models for five out of seven physicochemical properties of octanes. In another Gutman et al.<sup>2</sup> have reported that the use of line graphs, in some cases, significantly improves the predictive power of topological indices. We decided to test both of these results by using them to reinvestigate several QSPR models from the literature. We also decided to test further the result that in many cases the optimum exponent of the vertex- and edge-connectivity indices is not  $-0.5$ .<sup>3</sup> Since we believe, along with many others,<sup>4</sup> that the QSPR modeling will become the tool of choice for many chemists-at-large in times to come, it seems to us worthwhile to search for the most reliable framework to carry out this kind of modeling. The present study is an attempt in this direction. It should also be noted that throughout this paper we will use the chemical graph theoretical concepts and language<sup>5</sup> only to simplify the analysis.

Recently, line graphs have been increasingly used in structure–property modeling,<sup>2,6–11</sup> although they may be traced back to van't Hoff, who used the line graphs of the structural formulas for representing simple organic compounds. Line graphs are described in a monograph on

chemical graph theory<sup>12</sup> and under the name bond graphs were used in deriving the molecular complexity indices.<sup>13</sup> The line graph  $L(G) = L$  of graph  $G$  is a graph derived from  $G$  in such a way that the edges in  $G$  are replaced by vertexes in  $L$ . Two vertexes in  $L$  are adjacent if the corresponding two edges in  $G$  are incident, that is, have a vertex in common. The construction of a line graph from a tree is shown in Figure 1.

The line graph  $L$  is usually a more complex structure than the corresponding graph  $G$ . Only in the case of unbranched cycloalkanes, represented by cycles,  $L$  and  $G$  coincide because in cycles the number of vertexes  $V$  and the number of edges  $E$  are identical. For  $n$ -alkanes, represented by the hydrogen-depleted chains,  $L$  is less complex than  $G$  because it has one less vertex than  $G$ , since in chains  $E = V - 1$ .

The numbers of vertexes  $V$  and edges  $E$  of the line graph  $L$  and the corresponding graph  $G$  are related by

$$V(L) = E(G) \quad (1)$$

$$E(L) = (1/2) \sum_i d_i^2(G) - E(G) \quad (2)$$

where  $d_i$  ( $i = 1, 2, \dots, V$ ) are degrees of vertexes in  $G$ . These relations can be easily confirmed by inspecting  $G$  and  $L$  depicted in Figure 1.

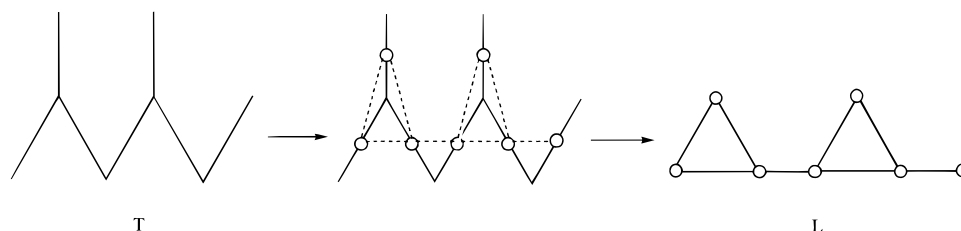
Using the equation

$$\sum_i d_i^2(G) = M_1 \quad (3)$$

where  $M_1$  is called<sup>14–16</sup> the first Zagreb-group index,<sup>17,18</sup> and

<sup>†</sup> Reported in part at the One Day Symposium on Applied Mathematical Chemistry, held on May 3, 1999, at the Natural Resources Research Institute, University of Minnesota, Duluth.

<sup>‡</sup> On leave of absence from The Rugjer Bošković Institute, HR-10001 Zagreb, The Republic of Croatia.



**Figure 1.** Construction of line graph L from tree T depicting 2,4-dimethylhexane.

introducing (3) into (2), we obtain the expression

$$E(L) = (1/2)M_1 - E(G) \quad (4)$$

Gutman and Estrada derived the same expression,<sup>19</sup> but the factor 1/2 is missing in their expression. From (4) follows an amusing result that the  $M_1$  index of the graph is simply equal to twice the number of vertexes and edges in the line graph:

$$M_1 = 2[E(L) + E(G)] = 2[E(L) + V(L)] \quad (5)$$

#### SIMPLE MODIFICATION OF THE VALENCE VERTEX- AND EDGE-CONNECTIVITY INDICES

**Vertex-Connectivity Index.** The standard definition of the vertex-connectivity index is<sup>20</sup>

$$\chi = \sum_{\text{edges}} [d(v_i) d(v_j)]^{-0.5} \quad (6)$$

where  $d(v_i)$  is the degree of the vertex  $v_i$  and  $[d(v_i) d(v_j)]^{-0.5}$  may be considered as the weight of the  $i$ - $j$  edge.<sup>21</sup> The summation in (6) goes over all edges. The vertex degree  $d(v_i)$  is equal to the number of vertexes adjacent to vertex  $i$  in a graph G. Any two vertexes in G are adjacent if there are edges connecting them.

Equation 6 is open to modification because the choice of edge weights  $[d(v_i) d(v_j)]^{-0.5}$  was based on one possible solution to the inequalities based on ordering graphs.<sup>20</sup> There are also other choices of weights possible. Hence, the quantity  $[d(v_i) d(v_j)]^{-0.5}$  can be replaced by  $[d(v_i) d(v_j)]^k$ , where  $k$  is a variable exponent that can be varied in any desired range of values, and (6) becomes<sup>3</sup>

$$\chi = \sum_{\text{edges}} [d(v_i) d(v_j)]^k \quad k \neq 0 \quad (7)$$

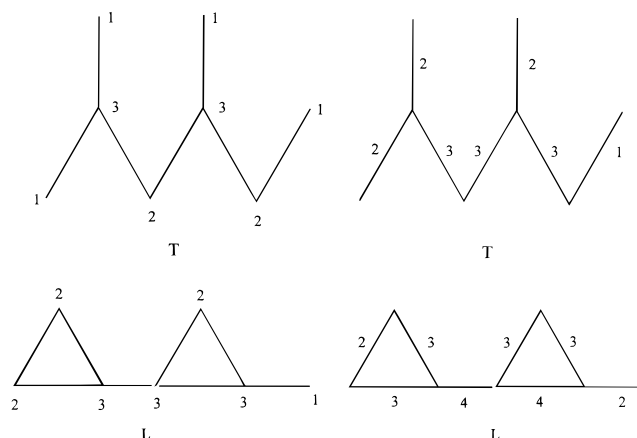
**Edge-Connectivity Index.** The standard definition of the edge-connectivity index is similar to the definition of the vertex-connectivity index, the only change being in using the edge degrees  $d(e_i)$  instead of vertex degrees  $d(v_i)$ :<sup>22</sup>

$$\epsilon = \sum_{\text{adjacent edges}} [d(e_i) d(e_j)]^{-0.5} \quad (8)$$

The edge degree  $d(e_i)$  is equal to the number of edges adjacent to edge  $i$  in a graph G. Any two edges in G are adjacent if they meet at the same vertex. Because every edge in G connects two vertexes, the edge degree  $d(e)$  can be expressed in terms of their degrees as follows:<sup>22</sup>

$$d(e) = d(v_i) + d(v_j) - 2 \quad (9)$$

This expression can be used to assign the degrees of edges in G. In Figure 2 we give the vertex and edge degrees in



**Figure 2.** Vertex degrees (digits at each vertex) and edge degrees (digits at each edge) in tree T and the corresponding line graph L from Figure 1.

tree T and the corresponding line graph L depicted in Figure 1.

A simple way to assign the degrees to edges in graph G or its line graph L is to count all adjacent bonds of a bond for which we wish to determine the edge degree. This procedure is illustrated in Figure 3.

Equation 8 can also be modified because the quantity  $[d(e_i) d(e_j)]^{-0.5}$  was the result of mimicking the original definition of Randić for the vertex-connectivity index.<sup>20</sup> Consequently,  $[d(e_i) d(e_j)]^{-0.5}$  can be replaced by  $[d(e_i) d(e_j)]^k$ , where  $k$  is a variable exponent that can be varied in any desired range of values. Thus, (8) converts into the following equation:

$$\epsilon = \sum_{\text{adjacent edges}} [d(e_i) d(e_j)]^k \quad k \neq 0 \quad (10)$$

At this point it should also be noted that the edge-adjacency matrix<sup>23</sup> of the graph G,  ${}^E\mathbf{A}(G)$ , is identical to the vertex-adjacency matrix<sup>23</sup> of the line graph L of G,  ${}^V\mathbf{A}(L)$ :

$${}^E\mathbf{A}(G) = {}^V\mathbf{A}(L) \quad (11)$$

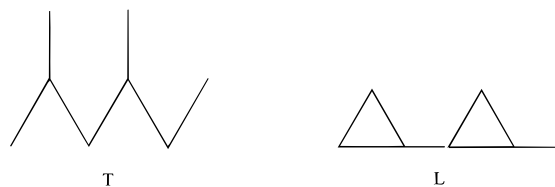
This must be so because the edge degrees in G are identical to the vertex degrees in the corresponding line graph L (see Figure 2). The consequence of (11) is that the edge-connectivity index of G is identical to the vertex-connectivity index of the corresponding line graph L:<sup>19</sup>

$$\epsilon(G) = \chi(L) \quad (12)$$

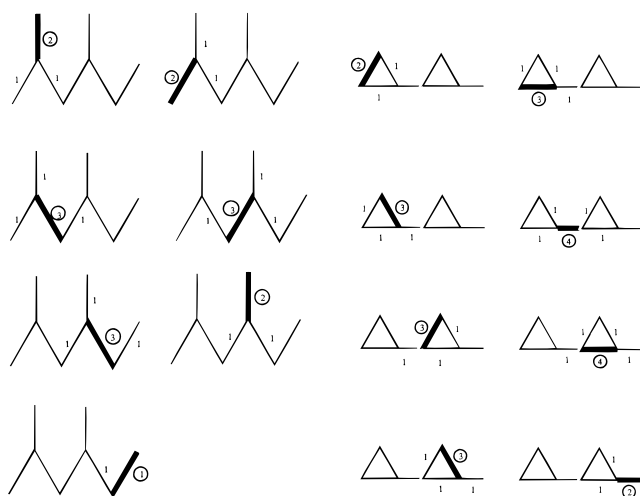
#### RESULTS AND DISCUSSION

We studied five structure–property models that were already reported in the literature. This was done on purpose because our aim was to compare the performance of the obtained models with those already published. The properties considered were boiling points of octanes and nonanes and

(1) Tree T and its line graph L



(2) Assigning degrees to edges in T and L

**Figure 3.** A simple procedure for assigning the degrees to edges in tree T and the related line graph L.

molar volumes, molar refractions, and retention indices of alkanes. Boiling points and retention indices are typical “surface”-dependent properties, while molar volumes and molar refractions are “molecular volume”-dependent properties. In all cases molecules were depicted as graphs and corresponding line graphs. The standard deviation  $S$  was used as a criterion for the comparison of the models. The optimum parameter  $k$  in (7) and (10) was determined using the procedure described in our earlier report;<sup>3</sup> that is, the parameter  $k$  was taken to be optimum when the value of  $S$  reached a minimum.

**Boiling Points of 18 Octanes.** We first considered structure–boiling point models for isomeric octanes, on the basis of their vertex-connectivity indices computed for octane trees. The best model was obtained for  $k = -1.15$ . The regression equation is given by

$$\text{bp} = 65.14(\pm 7.29) + 28.87(\pm 4.31)\chi^{[-1.15]} \quad (13)$$

$$n = 18 \quad R = 0.859 \quad S = 3.24 \quad F = 45$$

where bp is the normal boiling point,  $R$  the correlation coefficient,  $S$  the standard deviation,  $F$  the Fisher ratio, and  $\chi^{[-1.15]}$  a short-hand notation for the vertex-connectivity index computed using the value of  $-1.15$  for the exponent in (7). The notation  $\chi^{[k]}$  will be used throughout this paper. The improvement over the model based on  $k = -0.5$  is rather slight:

$$\text{bp} = 3.14(\pm 19.23) + 30.33(\pm 5.27)\chi^{[-0.50]} \quad (14)$$

$$n = 18 \quad R = 0.821 \quad S = 3.60 \quad F = 33$$

The above models are identical to structure–boiling point models for octanes published elsewhere.<sup>3,24</sup> Randić et al.<sup>25</sup> have also observed that the modified vertex-connectivity index produces better structure–boiling point models of lower ( $C_2$ – $C_7$ ) alkanes than the standard version of the vertex-connectivity index. However, they have found that the exponent value of  $-0.33$  leads to the best models of three alternatives they considered ( $k = -0.5, -0.33, -0.25$ ).

The same analysis as with the vertex-connectivity index was also carried out with the edge-connectivity index. The best model was obtained for  $k = -0.30$ . The regression equation is given by

$$\text{bp} = 179.75(\pm 11.12) - 13.66(\pm 2.30)\epsilon^{[-0.30]} \quad (15)$$

$$n = 18 \quad R = 0.830 \quad S = 3.52 \quad F = 35$$

where  $\epsilon^{[-0.30]}$  is a short-hand notation for the edge-connectivity index computed using the value of  $-0.30$  for the exponent in (10). The notation  $\epsilon^{[k]}$  will be used throughout this paper. The improvement over the model based on  $k = -0.5$  is considerable

$$\text{bp} = 162.76(\pm 29.94) - 14.52(\pm 8.85)\epsilon^{[-0.50]} \quad (16)$$

$$n = 18 \quad R = 0.379 \quad S = 5.84 \quad F = 3$$

but the model in (15) is not as good as the model in (13), though it is somewhat better than the model in (14). This result supports the work by Estrada and Rodriguez,<sup>1</sup> because one of the two physicochemical properties of octanes for which the use of the edge-connectivity index did not produce the best single-variable QSPR model was the boiling point, the other being the heat of vaporization. Estrada and Rodriguez pointed out that to describe these properties correctly it is necessary to take into account long-range contributions in the edge-connectivity index.<sup>9</sup> In both these cases better single-variable models were obtained using the Hosoya  $Z$  index.<sup>26</sup>

Finally, we considered octane line graphs. Since  $\chi^{[k]}(L) = \epsilon^{[k]}(G)$ , we derived structure–boiling point models based on the edge-connectivity index  $\epsilon^{[k]}(L)$ . The best model was obtained for  $k = -0.675$ . The regression equation is given by

$$\text{bp} = 167.56(\pm 9.03) - 20.17(\pm 3.37)\epsilon^{[-0.675]}(L) \quad (17)$$

$$n = 18 \quad R = 0.831 \quad S = 3.51 \quad F = 36$$

where  $\epsilon^{[-0.675]}(L)$  is a short-hand notation for the edge-connectivity index computed for a line graph using the value of  $-0.675$  for the exponent in (10). This notation will be used throughout this paper when the models based on line graphs and edge-connectivity indices are discussed.

The model in (17) is practically the same as the model in (15) on the basis of octane trees and the edge-connectivity index. The improvement over the model based on  $k = -0.5$  is visible:

$$\text{bp} = 138.83(\pm 5.80) - 6.11(\pm 1.39)\epsilon^{[-0.50]}(L) \quad (18)$$

$$n = 18 \quad R = 0.740 \quad S = 4.24 \quad F = 19$$

However, this model is much better than the corresponding model in (16) on the basis of octane trees.

**Boiling Points of 35 Nonanes.** The same kind of analysis as in the case of modeling boiling points of octanes is carried out for nonanes. We first considered structure–boiling point models for isomeric nonanes, on the basis of their vertex-connectivity indices computed for nonane trees. The best model was obtained for  $k = -1.25$ . The regression equation is given by

$$\text{bp} = 94.23(\pm 6.68) + 25.58(\pm 3.97)\chi^{[-1.25]} \quad (19)$$

$$n = 35 \quad R = 0.746 \quad S = 4.13 \quad F = 41$$

The improvement over the model based on  $k = -0.5$  is again rather slight:

$$\text{bp} = 31.47(\pm 19.64) + 25.67(\pm 4.77)\chi^{[-0.50]} \quad (20)$$

$$n = 35 \quad R = 0.683 \quad S = 4.53 \quad F = 29$$

The above models are comparable to the structure–boiling point models for nonanes published elsewhere.<sup>3,27</sup> The same analysis was also carried out with the edge-connectivity index. The best model was obtained for  $k = -0.375$ . The corresponding regression equation is

$$\text{bp} = 225.36(\pm 16.39) - 18.42(\pm 3.41)\epsilon^{[-0.375]} \quad (21)$$

$$n = 35 \quad R = 0.685 \quad S = 4.52 \quad F = 29$$

This model and the model in (20) are practically the same. However, it is worse than the model in (19). The improvement over the model based on  $k = -0.5$  is considerable:

$$\text{bp} = 218.35(\pm 30.12) - 21.31(\pm 7.89)\epsilon^{[-0.50]} \quad (22)$$

$$n = 35 \quad R = 0.426 \quad S = 5.61 \quad F = 7$$

Finally, we considered nonane line graphs. The best model was obtained for  $k = -0.70$ :

$$\text{bp} = 203.18(\pm 9.60) - 22.96(\pm 3.32)\epsilon^{[-0.70]}(\text{L}) \quad (23)$$

$$n = 35 \quad R = 0.769 \quad S = 3.97 \quad F = 48$$

This model is better than any regarding the relationship between structures and boiling points of nonanes. It represents an improvement over the model based on  $k = -0.5$ :

$$\text{bp} = 161.56(\pm 5.94) - 22.96(\pm 3.32)\epsilon^{[-0.50]}(\text{L}) \quad (24)$$

$$n = 35 \quad R = 0.587 \quad S = 5.03 \quad F = 17$$

Comparison between this model and the related models based on octane trees shows that the model in (24) is not as good as the model in (20), but better than the model in (22).

In this example, the edge-connectivity index did live up to the expectations based on the work by Gutman et al.:<sup>2</sup> The use of the line graph edge-connectivity index produced for nonanes the best structure–boiling point model. However, the model in (23) is still far from being satisfactory in comparison with models that use several topological indices.<sup>28</sup> For example, the best structure–boiling point model for nonanes with five descriptors has  $R = 0.981$  and  $S = 0.89$ .<sup>29</sup>

#### Gas Chromatographic Retention Indices of Alkanes.

The same methodology as above was applied to the relationship between the structures of alkanes and their gas chromatographic retention indices.<sup>30</sup> We first considered

structure–chromatographic retention data correlation for the first 157 alkanes using as the structural parameter the vertex-connectivity index. The best correlation was obtained for  $k = -0.325$ :

$$\text{RI} = 74.58(\pm 8.48) + 148.14(\pm 1.53)\chi^{[-0.325]} \quad (25)$$

$$n = 157 \quad R = 0.992 \quad S = 23.8 \quad F = 9330$$

where RI stands for the retention indices of alkanes. This model gives a very good agreement between experimental and computed retention indices of alkanes. Retention indices of alkanes cover a range from  $\text{RI}(\text{methane}) = 100$  to  $\text{RI}(\text{2,3-dimethylundecane}) = 1251.4$ . In most cases the difference between experimental and computed values is less than 3%.

The model in (25) is only slightly better than the model based on  $k = -0.5$ :

$$\text{RI} = 64.92(\pm 9.38) + 187.97(\pm 2.13)\chi^{[-0.50]} \quad (26)$$

$$n = 157 \quad R = 0.990 \quad S = 26.0 \quad F = 7801$$

The use of the edge-connectivity index produced poorer models:

$$\text{RI} = 137.98(\pm 13.79) + 200.54(\pm 3.66)\epsilon^{[-0.55]} \quad (27)$$

$$n = 157 \quad R = 0.975 \quad S = 41.3 \quad F = 3008$$

$$\text{RI} = 134.0(\pm 14.55) + 184.89(\pm 3.54)\epsilon^{[-0.50]} \quad (28)$$

$$n = 157 \quad R = 0.973 \quad S = 43.2 \quad F = 2729$$

These two models are comparable, but are much better than models based on alkane line graphs and their edge-connectivity indices:

$$\text{RI} = 206.58(\pm 21.72) + 262.94(\pm 8.30)\epsilon^{[-0.775]}(\text{L}) \quad (29)$$

$$n = 157 \quad R = 0.931 \quad S = 68.2 \quad F = 1003$$

$$\text{RI} = 365.44(\pm 36.63) + 104.00(\pm 7.24)\epsilon^{[-0.50]}(\text{L}) \quad (30)$$

$$n = 157 \quad R = 0.756 \quad S = 122.2 \quad F = 206$$

There are several structure–chromatographic retention index correlations for alkanes available in the literature.<sup>30</sup> Most of them are based on the two-dimensional and three-dimensional Wiener numbers. However, there is also a correlation available based on the vertex-connectivity index with  $k = -0.5$  which differs only slightly from (26):<sup>30</sup>

$$\text{RI} = 69.81(\pm 9.31) + 186.93(\pm 2.11)\chi^{[-0.50]} \quad (31)$$

$$n = 157 \quad R = 0.990 \quad S = 26.0 \quad F = 7827$$

The initial work on the structure–chromatographic retention data correlations is due to Randić.<sup>31</sup> The correlations based on the two-dimensional (<sup>2</sup>W) and three-dimensional (<sup>3</sup>W) Wiener numbers, which are adjusted Walker-type correlations,<sup>32</sup> are not as good as the model in (25):<sup>30</sup>

$$\text{RI} = 171.2(\pm 15.7) {}^2W^{0.335(\pm 0.013)} - 48.6(\pm 27.3) \quad (32)$$

$$n = 157 \quad R = 0.984 \quad S = 33.0 \quad F = 2403$$

$$\text{RI} = 170.6(\pm 17.0) {}^3W^{0.325(\pm 0.013)} - 31.8(\pm 30.2) \quad (33)$$

$$n = 157 \quad R = 0.982 \quad S = 35.6 \quad F = 2048$$



These models are, however, better than the ones based on edge-connectivity indices computed for either alkane trees or alkane line graphs. The best overall structure–chromatographic retention data correlation is obtained with the vertex-connectivity index with  $k = -0.325$  (model in (25)). This is to our knowledge the best structure–chromatographic retention data model of alkanes that exists in the literature.

**Molar Volumes of Alkanes.** We considered molar volumes of 69 lower alkanes taken from Estrada.<sup>22</sup> We first considered the structure–molar volume relationship using the vertex-connectivity index. The best correlation was obtained for a rather small value of  $k$  ( $-0.07$ ). The regression equation and the statistical parameters for the correlation are:

$$MV = 55.85(\pm 2.10) + 16.53(\pm 0.32)\chi^{[-0.07]} \quad (34)$$

$$n = 69 \quad R = 0.988 \quad S = 2.73 \quad F = 2649$$

where MV stands for molar volume. This regression is better as expected than the one based on the standard value of  $k$  ( $-0.50$ ):

$$MV = 53.07(\pm 4.41) + 29.60(\pm 1.18)\chi^{[-0.50]} \quad (35)$$

$$n = 69 \quad R = 0.951 \quad S = 5.38 \quad F = 632$$

These models are inferior to those based on the edge-connectivity index. The best structure–molar volume model was obtained for  $k = -0.515$ :

$$MV = 57.44(\pm 1.37) + 31.80(\pm 0.41)\epsilon^{[-0.515]} \quad (36)$$

$$n = 69 \quad R = 0.995 \quad S = 1.81 \quad F = 6094$$

This model is only very slightly better than the model based on the standard value of the exponent  $k$ :

$$MV = 58.23(\pm 1.41) + 30.80(\pm 0.41)\epsilon^{[-0.50]} \quad (37)$$

$$n = 69 \quad R = 0.994 \quad S = 1.88 \quad F = 5669$$

Equation 37 is different from the corresponding one given by Estrada<sup>22</sup> as (1) in his paper. The difference is caused by the use of erroneous values of the edge-connectivity indices for six alkanes in Table 1 in Estrada's paper. The correct values are (we use the same codes for alkanes as Estrada): (33ME5)  $-3.1160$ , (233MMM5)  $-3.2832$ , (33ME6)  $-3.6766$ , (234MMM6)  $-3.7921$ , (244MMM6)  $-3.8432$ , and (334MMM6)  $-3.7107$ . The model in (37) is in fact better than the model given in Estrada's paper (statistical parameters for Estrada's structure–molar volume model with six incorrect values of edge-connectivity indices are  $R = 0.993$ ,  $S = 2.034$ , and  $F = 4822$ ).

The statistical characteristics of models based on the edge-connectivity index also support the work by Estrada and Rodriguez,<sup>1</sup> because one of the five physicochemical properties of octanes for which the use of the edge-connectivity index produced the best single-variable QSPR model was the molar volume. This also agrees with analyses which point out that the edge-connectivity index is more appropriate to be used in the structure–molecular volume properties modeling than the vertex-connectivity index.

The structure–molar volume models based on line graphs and edge-connectivity indices possess rather inferior statistical parameters than the models shown above:

$$MV = 111.14(\pm 5.76) + 12.45(\pm 1.35)\epsilon^{[-0.50]}(L) \quad (38)$$

$$n = 69 \quad R = 0.748 \quad S = 11.54 \quad F = 85$$

$$MV = 67.12(\pm 3.59) + 44.49(\pm 1.67)\epsilon^{[-0.775]}(L) \quad (39)$$

$$n = 69 \quad R = 0.956 \quad S = 5.09 \quad F = 714$$

**Molar Refractions of Alkanes.** We considered molar refractions of 69 lower alkanes also taken from Estrada.<sup>22</sup> Among the reported experimental values one is incorrect: Molar refraction of 34MM6 is 38.8453 instead of 43.6870.<sup>33</sup> We first considered the structure–molar refraction relationship using the vertex-connectivity index. The best correlation was obtained again for a rather small value of  $k$  ( $-0.02$ ). The regression equation and the statistical parameters for the correlation are

$$MR = 6.99(\pm 0.15) + 4.70(\pm 0.02)\chi^{[-0.02]} \quad (40)$$

$$n = 69 \quad R = 0.9993 \quad S = 0.200 \quad F = 46865$$

where MR is a short-hand notation for molar refraction. This regression equation is better than the one based on the standard value of  $k$  ( $-0.50$ ):

$$MR = 5.76(\pm 1.88) + 9.11(\pm 0.32)\chi^{[-0.50]} \quad (41)$$

$$n = 69 \quad R = 0.962 \quad S = 1.45 \quad F = 824$$

The model in (40) is better than, and the model in (41) is worse than, the corresponding models based on the edge-connectivity index. The best structure–molar refraction model using edge-connectivity indices was obtained for  $k = -0.495$ :

$$MR = 7.77(\pm 0.50) + 9.26(\pm 0.14)\epsilon^{[-0.495]} \quad (42)$$

$$n = 69 \quad R = 0.992 \quad S = 0.668 \quad F = 4130$$

There is hardly any difference between this model and the model based on the standard value of exponent  $k$ :

$$MR = 7.71(\pm 0.50) + 9.36(\pm 0.15)\epsilon^{[-0.50]} \quad (43)$$

$$n = 69 \quad R = 0.992 \quad S = 0.672 \quad F = 4090$$

Equation 43 is different from the corresponding one given by Estrada<sup>22</sup> as (2) in his paper. The difference is caused by erroneous values of the edge-connectivity indices for six alkanes (see the discussion above). The model in (43) is a little better than the model in the Estrada paper when the corrected values of the edge-connectivity indices are used. We also carried out the statistical analysis of Estrada's structure–molar refraction model with six incorrect values of edge-connectivity indices and obtained different statistical parameters ( $R = 0.983$ ,  $S = 0.964$ , and  $F = 1969$ ) from those reported ( $R = 0.9913$ ,  $S = 0.698$ , and  $F = 3782$ ).

The model in (43), being better than the model in (41), supports the claim by Estrada and Rodriguez<sup>1</sup> regarding modeling the molar refraction. In their work one of the five physicochemical properties of octanes for which the use of the edge-connectivity index produced the best single-variable QSPR model was also the molar refraction. However, when the models based on vertex- and edge-connectivity indices with variable exponents are considered, the reverse is true: the model in (40) is better than the model in (42). The model in (40) is also better than the model in (43).

Structure—molar refraction models based on alkane line graphs and edge-connectivity indices are again as in the case of structure—molar volume models inferior to models based on connectivity indices computed for alkane trees:

$$\text{MR} = 11.78(\pm 1.29) + 12.30(\pm 0.56)\epsilon^{[-0.75]}(\text{L}) \quad (44)$$

$$n = 69 \quad R = 0.936 \quad S = 1.861 \quad F = 474$$

$$\text{MR} = 23.29(\pm 1.68) + 3.91(\pm 0.39)\epsilon^{[-0.50]}(\text{L}) \quad (45)$$

$$n = 69 \quad R = 0.772 \quad S = 3.358 \quad F = 99$$

Model (40), in which the value of the exponent is rather low,<sup>34</sup> supports the use of the structure-molecular refraction model based on the simplest possible topological index, the number of carbon atoms  $V$ :

$$\text{MR} = 2.60 (\pm 0.18) + 4.55 (\pm 0.02) V \quad (46)$$

$$n = 69 \quad R = 0.999 \quad S = 0.208 \quad F = 43200$$

### CONCLUSIONS

We investigated five structure—property models of alkanes. The properties considered were molecular surface-dependent properties (boiling points and gas chromatographic retention indices) and molecular volume-dependent properties (molar volumes and molar refractions). Alkanes were represented by trees and the corresponding line graphs. The vertex- and edge-connectivity indices were used as structural parameters. In each studied case we computed connectivity indices with an optimum exponent and with a standard value of  $-0.5$ . In total we generated six QSPR models for each property. The obtained results lead us to conclude the following.

(i) In all cases QSPR models based on connectivity indices with optimum exponents have better statistical parameters than the models based on connectivity indices with the standard value of the exponent ( $-0.5$ ). This is fully in agreement with our earlier study<sup>3</sup> and the ideas of Altenburg,<sup>35</sup> Randić et al.,<sup>25</sup> and Estrada.<sup>36</sup> Therefore, we suggest that the modified versions of vertex- and edge-connectivity indices should be routinely employed in the structure—property modeling rather than the standard versions of the connectivity indices.

(ii) In the five cases that we studied the structure—boiling point models for octanes and nonanes and the structure—chromatographic retention index model for alkanes based on vertex-connectivity indices are better than the corresponding models based on edge-connectivity indices. Thus, it appears that the vertex-connectivity index is more appropriate to be used in the structure—molecular surface properties modeling than the edge-connectivity index. Consequently, the vertex-connectivity index may be considered as a molecular surface descriptor.

(iii) In the five cases that we studied the structure—molar volume and the structure—molar refraction models for  $C_5$ — $C_9$  alkanes based on the edge-connectivity index produced the best single-variable model. This agrees with the findings of Estrada and Rodriguez<sup>1</sup> and is suggestive that the edge-connectivity index is the better descriptor to be used in the structure—molecular volume properties modeling than the edge-connectivity index. Thus, the edge-connectivity index may be regarded as a molecular volume descriptor. The edge-

connectivity index appears to be a promising molecular descriptor,<sup>1,10,37–39</sup> especially if the long-range contributions to this index are included in the modeling.<sup>9,11</sup>

(iv) The use of line graphs in this study did not improve the predictive power of the connectivity indices. Only in the case of structure—boiling point modeling for nonanes the model based on the nonane line graphs produced the best model among the possibilities considered. Since the construction of the line graphs is not difficult and the computation of their descriptors can be easily carried out, it is also reasonable to use them in the QSPR modeling, but to establish the usefulness of the line graph model in the structure—property studies, more work is needed.

### ACKNOWLEDGMENT

This paper was supported, in part, by the Ministry of Science and Technology of the Republic of Croatia through Grant Nos. 00980606 (S.N. and N.T.) and 079301 (D.A. and D.B.). Part of this work was completed during the stay of S.N. and N.T. at the Natural Resources Research Institute (NRRI), University of Minnesota, Duluth, MN. S.C.B. and S.N. and N.T. (during their stay at NRRI) were supported, in part, by Grants F49620-94-1-0401 and F49620-96-1-0330 from the United States Air Force. This paper is contribution number 267 from the Center for Water and the Environment, NRRI. We thank the reviewers for helpful comments. We also thank Dr. Christoph Rücker (Freiburg) for helpful comments.

### REFERENCES AND NOTES

- (1) Estrada, E.; Rodriguez, L. Edge-Connectivity Indices in QSPR/QSAR Studies. 1. Comparison to Other Topological Indices in QSPR Studies. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1037–1041.
- (2) Gutman, I.; Popović, Lj.; Estrada, E.; Bertz, S. H. The Line Graph Model. Predicting Physico-Chemical Properties of Alkanes. *ACH-Models Chem.* **1998**, *135*, 147–155 and references therein.
- (3) Amić, D.; Bešlo, D.; Lučić, B.; Nikolić, S.; Trinajstić, N. The Vertex-Connectivity Index Revisited. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 819–822 and references therein.
- (4) Katritzky, A. R.; Karelson, M.; Lobanov, V. S. QSPR as a Means of Predicting and Understanding Chemical and Physical Properties in Terms of Structure. *Pure Appl. Chem.* **1997**, *69*, 245–248.
- (5) Trinajstić, N. *Chemical Graph Theory*, 2nd revised ed.; CRC Press: Boca Raton, FL, 1992.
- (6) Diudea, M.; Horvath, D.; Bonchev, D. Molecular Topology. 14. Molord Algorithm and Real Number Subgraph Invariants. *Croat. Chem. Acta* **1995**, *68*, 131–148.
- (7) Estrada, E.; Guevara, N.; Gutman, I.; Rodriguez, L. Molecular Connectivity Indices of Iterated Line Graphs. A New Source of Descriptors for QSPR and QSAR Studies. *SAR QSAR Environ. Res.* **1998**, *9*, 229–240.
- (8) Estrada, E.; Guevara, N.; Gutman, I. Extension of Edge Connectivity Index. Relationships to Line Graph Indices and QSPR Applications. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 428–431.
- (9) Estrada, E. Edge-Connectivity Indices in QSPR/QSAR Studies. 2. Accounting for Long-Range Bond Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1042–1048.
- (10) Estrada, E. Novel Strategies in the Search of Topological Indices. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers: The Netherlands, 1999; pp 279–306.
- (11) Estrada, E. Connectivity Polynomial and Long-Range Contributions in the Molecular Connectivity Model. *Chem. Phys. Lett.*, in press.
- (12) Trinajstić, N. *Chemical Graph Theory*; CRC: Boca Raton, FL, 1983; Vol. I, p 17.
- (13) Bertz, S. H. The Bond Graph. *J. Chem. Soc., Chem. Commun.* **1981**, 818–820.
- (14) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399–404.
- (15) Basak, S. C.; Grunwald, G. D.; Niemi, G. J. Use of Graph-Theoretic and Geometrical Molecular Descriptors in Structure—Activity Rela-

- tionships. In *From Chemical Topology to Three-Dimensional Geometry*; Balaban, A. T., Ed.; Plenum Press: New York, 1997; pp 73–116.
- (16) Balaban, A. T.; Ivanciuc, O. Historical Development of Topological Indices. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers: The Netherlands, 1999; pp 21–57.
- (17) Gutman, I.; Trinajstić, N. Graph Theory and Molecular Orbitals. III. Total  $\pi$ -Electron Energy of Alternant Hydrocarbons. *Chem. Phys. Lett.* **1972**, *17*, 535–538.
- (18) Gutman, I.; Ruščić, B.; Trinajstić, N.; Wilcox, C. F., Jr. Graph Theory and Molecular Orbitals. XII. Acyclic Polyenes. *J. Chem. Phys.* **1975**, *62*, 3399–3405.
- (19) Gutman, I.; Estrada, E. Topological Indices Based on the Line Graph of the Molecular Graph. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 541–543.
- (20) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (21) Randić, M.; Jeričević, Ž.; Sabljčić, A.; Trinajstić, N. On the Molecular Connectivity and  $\pi$ -Electronic Energy of Polycyclic Hydrocarbons. *Acta Phys. Pol.* **1988**, *74*, 317–330.
- (22) Estrada, E. Edge Adjacency Relationship and a Novel Topological Index Related to Molecular Volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31–33.
- (23) Trinajstić, N. *Chemical Graph Theory*, 2nd revised ed.; CRC Press: Boca Raton, FL, 1992; Chapter 4.
- (24) Randić, M.; Trinajstić, N. Viewpoint 4—Comparative Structure—Property Studies: The Connectivity Basis. *J. Mol. Struct.: THEOCHEM* **1993**, *284*, 209–221.
- (25) 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.
- (26) Hosoya, H. Topological Index. A Newly Proposed Quantity Characterizing Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
- (27) Randić, M.; Trinajstić, N. Isomeric Variations in Alkanes: Boiling Points of Nonanes. *New J. Chem.* **1994**, *18*, 179–189.
- (28) Lučić, B.; Trinajstić, N. New Developments in QSPR/QSAR Modeling Based on Topological Indices. *SAR QSAR Environ. Res.* **1997**, *7*, 45–62.
- (29) Rücker, G.; Rücker, C. On Topological Indices, Boiling Points, and Cycloalkanes. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 788–802.
- (30) Bošnjak, N.; Mihalić, Z.; Trinajstić, N. Application of Topographic Indices to Chromatographic Data: Calculation of the Retention Indices of Alkanes. *J. Chromatogr.* **1991**, *540*, 430–440 and references therein.
- (31) Randić, M. The Structural Origin of Chromatographic Retention Data. *J. Chromatogr.* **1978**, *161*, 1–14.
- (32) Canfield, E. R.; Robinson, R. W.; Rouvray, D. H. Determination of the Wiener Molecular Branching Index for the General Tree. *J. Comput. Chem.* **1985**, *6*, 598–609.
- (33) E-mail of Dr. Christoph Rücker (Freiburg) to us on March 2, 2000. Dr. Rücker was informed about this by Dr. Ernesto Estrada on March 24, 1999.
- (34) Note that for alkanes  $\sum_{\text{edges}} [d(v_i) d(v_j)]^0 + 1 = V$ .
- (35) Altenburg, K. Eine Bemerkung zu dem Randićschen “Molekularen Bindungs-Index (Molecular Connectivity Index)”. *Z. Phys. Chem. (Leipzig)* **1980**, *261*, 389–393.
- (36) Estrada, E. Graph Theoretical Invariant of Randić Revisited. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1022–1025.
- (37) Nikolić, S.; Trinajstić, N.; Baučić, I. Comparison between the Vertex- and Edge-Connectivity Indices for Benzenoid Hydrocarbons. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 42–46.
- (38) Nikolić, S.; Trinajstić, N.; Ivaniš, S. The Connectivity Indices of Regular Graphs. *Croat. Chem. Acta* **1999**, *72*, 875–883.
- (39) Cash, G. G. Correlation of Physicochemical Properties of Alkylphenols with Their Graph-Theoretical  $\epsilon$  Parameter. *Chemosphere* **1995**, *31*, 4307–4315.

CI990119V