CROATICA CHEMICA ACTA CCACAA **71** (3) 449–458 (1998)

ISSN-0011-1643
CCA-2504 *Original Scientific Paper*

Wiener Index: Formulas for Non-homeomorphic Graphs

István Lukovits

Chemical Research Center, Hungarian Academy of Sciences, H–1525, Budapest, P. O. Box 17, Hungary

Received January 20, 1998; revised April 2, 1998; accepted May 22, 1998

The Wiener index (W) is a graph invariant. Because of its importance, various methods – numerical approaches and analytical formulas – have been proposed to compute W. Earlier formulas of W were derived for homeomorphic structures, polymers, and stars. The formulas are polynomials. In this paper, a method is proposed that allows calculating the coefficients of the polynomials in sets of non-homeomorphic structures.

INTRODUCTION

The Wiener-number or Wiener-index (W) is a graph-invariant.[1] W is equal to the sum of distances (*i.e.* the number of bonds on the shortest path) $d_{i,j}$ between all pairs of vertices i and j:

$$W = \sum_{i<j} d_{i,j}. \quad (1)$$

As an example, consider the computation of the Wiener-index of isobutane (Figure 1, hydrogens will not be considered throughout this paper). W (isobutane) = $d_{1,2} + d_{1,3} + d_{1,4} + d_{2,3} + d_{2,4} + d_{3,4} = 1 + 1 + 1 + 2 + 2 + 2 = 9$.

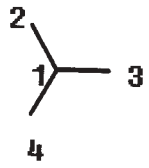

Figure 1. Hydrogen suppressed graph of isobutane. The numbers indicate vertices.

W is one of the most freqnently applied graph theoretical invariants. *W* has been used to explain the variation in boiling points, molar volumes, refractive indices, heats of isomerization and heats of vaporization of alkanes.[1] Later heats of formation, atomization, isomerization, and vaporization as well as density, critical pressure, surface tension, viscosity, melting points, partition coefficients, chromotographic retention indices, and stability of crystal lattices of various kinds of molecules were related[2,3] to *W*. Besides chemistry, the concept of the Wiener number has been used in electrical engineering[4] and mathematics.[5] Quite a few graph invariants related to *W* have been proposed, too.[6]

In addition to Eq. (1), analytical formulas have been proposed for *W*. Wiener derived a formula for chains[1] (*i.e.* for acyclic structures having two endpoints), Entringer *et al.*[7] and Bonchev *et al.*[8] proposed formulas for simple cycles, Trinajstić *et al.* obtained formulas for polymers[9] and Gutman derived formulas for various types of acyclic graphs.[10] A formula was also derived for fused bicyclic structures.[11] Canfield *et al.* proposed a recursion method to obtain appropriate formulas for any acyclic structure.[12] Each of these formulas is valid for a given set of *homeomorphic* structures, polymers or stars. Two graphs are homeomorphic if both can be obtained from the same graph by a sequence of subdivisions of lines. For example any two cycles are homeomorphic, any two chains are homeomorphic and the star depicted in Figure 1 is homeomorphic with the structure depicted in Figure 2. In this paper, a method will be proposed that allows to obtain formulas for a broader set of structures.

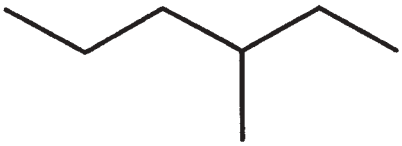

Figure 2. Example of a »starlike« graph containing three chains. $k = 2$, $m = 3$, and $n = 4$. The branching vertex belongs to all three chains.

DERIVATION OF FORMULAS

Expressions »vertex« and »atom«, »chemical structure« and »graph«, »bond« and »edge«, »valence« and »degree« are synonymous words used in chemistry and graph theory, respectively, and will be used interchangeably hereafter.

Consider a »starlike« graph (Figure 2) that has a single branching vertex, and three side-chains k, m, and n, where k, m and n denote both the number of vertices of the side chain and the side-chains themselves. A starlike graph (Figure 2) is homeomorphic with a star which has a branching

vertex of the same degree (Figure 1). The branching atom belongs to all side-chains. Example: isobutane (Figure 1) consists of three side chains, and each side chain contains two vertices. Instead of »side-chain«, we shall use the expression »string« to denote a chainlike subgraph starting with an endpoint or a branching vertex and ending with and endpoint or a branching vertex. An atom between these two vertices (if any) is bivalent.

The Wiener index of a starlike graph containing strings k, m and n (Figure 2) may be calculated by using the following formula:[2,3]

$$W = [(k^3 + m^3 + n^3) + 3(k^2m + k^2n + m^2n + km^2 + kn^2 + mn^2) - \\ - 6(k^2 + m^2 + n^2) - 6(km + kn + mn) + 5\,(k + m + n)]/6. \qquad (2)$$

For isobutane with $k = m = n = 2$, we obtain: $W = [(2^3 + 2^3 + 2^3) + 3(2^2{\times}2 + 2^2{\times}2 + 2^2{\times}2 + 2{\times}2^2 + 2{\times}2^2 + 2{\times}2^2) - 6(2^2 + 2^2 + 2^2) - 6(2{\times}2 + 2{\times}2 + 2{\times}2) + 5(2 + 2 + 2)]/6 = (24 + 144 - 72 - 72 + 30)/6 = 9$, the same value that was obtained by using Eq. (1). Eq. (2) remains valid if any of the side chains disappears, *i. e.* the size of the corresponding string – *e.g.* n – is equal to 1. Let us use Eq. (2) to obtain W for propane, then $k = 3$, $m = n = 1$, and $W = [(3^3 + 1^3 + 1^3) + 3(3^2{\times}1 + 3^2{\times}1 + 1^2{\times}1 + 3{\times}1^2 + 3{\times}1^2 + 1{\times}1^2) - 6(3^2 + 1^2 + 1^2) - 6(3{\times}1 + 3{\times}1 + 1{\times}1) + 5(3 + 1 + 1)]/6 = (29 + 78 - 66 - 42 + 25)/6 = 4$, in accordance with the value that could be obtained for the (hydrogen suppressed) graph of propane.

We want to derive a similar formula for a starlike graph containing four strings k, m, n and o (Figure 3) and the formula should have the following form:

$$W = [A_4(k^3 + m^3 + n^3 + o^3) + B_4(k^2m + k^2n + k^2o + m^2n + m^2o + n^2o + \\ + km^2 + kn^2 + ko^2 + mn^2 + mo^2 + no^2) + D_4(k^2 + m^2 + n^2 + o^2) + \\ + E_4(km + kn + ko + mn + mo + no) + F_4(k + m + n + o)]/6 \qquad (3)$$

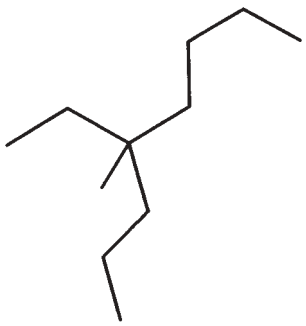

Figure 3. Example of a »starlike« graph containing four chains. The branching vertex belongs to all four chains, $k = 2$, $m = 3$, $n = 4$, and $o = 5$.

where coefficients A_4, B_4, D_4, E_4 and F_4, have to be determined. The subscripts indicate that the starlike graph contains four side-chains (strings). For similar reasons, the corresponding coefficients in Eq. (2) will be denoted by A_3, B_3, D_3, E_3 and F_3, where $A_3 = 1$, $B_3 = 3$, $D_3 = -6$, $E_3 = -6$ and $F_3 = 5$. Note that Eq. (3) cannot contain other third order terms, like kmn, a constant, fourth or higher order terms.[11]

Let as suppose that coefficients A_4, B_4, D_4, E_4 and F_4 are already known, and we want to use Eq. (3) to derive Eq. (2). Assume that $o = 1$, and substituting $o = 1$ into Eq. (3), we obtain for the constant value in Eq. (2) which equals zero:

$$A_4 + D_4 + F_4 = 0. \tag{4}$$

Similarly, the sum of the coefficients of the k^2 term in Eq. (3) is equal to –6, the coefficient of the k^2 term in Eq. (2):

$$B_4 + D_4 = D_3 = -6. \tag{5}$$

By comparing the coefficients of k in Eqs. (2) and (3), we obtain:

$$B_4 + E_4 + F_4 = F_3 = 5. \tag{6}$$

The present procedure cannot generate an equation for A_3 or B_3, since the coefficients of the third order terms k^3 and k^2m, A_4 and B_4, respectively, are not affected by setting o = 1. Therefore $A_4 = A_3 = 1$ and $B_4 = B_3 = 3$, and the set of equations (4)–(6) can be solved to obtain $D_4 = -9$, $E_4 = -6$ and $F_4 = 8$. From this, we can write down the equation for starlike structures with four sidechains:

$$\begin{aligned} W = [&(k^3 + m^3 + n^3 + o^3) + 3(k^2m + k^2n + k^2o + m^2n + m^2o + n^2o + \\ &+ km^2 + kn^2 + ko^2 + mn^2 + mo^2 + no^2) - 9(k^2 + m^2 + n^2 + o^2) - \\ &- 6(km + kn + ko + mn + mo + no) + 8\ (k + m + n + o)]/6. \end{aligned} \tag{7}$$

Example: consider structure of 3,3-dimethylpentane (Figure 4), With $k = m = 2$, n = o = 3 and $W = [(8 + 8 + 27 + 27) + 3(8 + 12 + 12 + 12 + 12 + 27 + 8 + 18 + 18 + 18 + 18 + 27) - 9(4 + 4 + 9 + 9) - 6(4 + 6 + 6 + 6 + 6 + 9) + 8(2 + 2 + 3 + 3)]/6 = (70 + 570 - 234 - 222 + 80)/6 = 44$.

The procedure can be extended for starlike structures eith n sidechains. Again $A_n = 1$ and $B_n = 3$. In full analogy with equations (4)–(6), we may write:

$$A_n + D_n + F_n = 0 \tag{8}$$

$$B_n + D_n = D_{n-1} \tag{9}$$

Figure 4. Hydrogen suppressed graph of 3,3-dimethylpentane.

$$B_n + E_n + F_n = F_{n-1}. \tag{10}$$

From Eq. (9) we obtain:

$$D_n = D_{n-1} - 3. \tag{11}$$

By expressing F_n from Eq. (8) and using Eq. (11), we obtain:

$$F_n = -D_n - 1 = 2 - D_{n-1}. \tag{12}$$

Finally, by expressing E_n from Eq. (10) and from Eqs. (11) and (12), we obtain:

$$\begin{aligned} E_n &= F_{n-1} - F_n - 3 = F_{n-1} + D_{n-1} - 2 - 3 = D_{n-1} - D_{n-2} - 3 = \\ &= D_{n-2} - D_{n-2} - 6 = -6. \end{aligned} \tag{13}$$

meaning that $E_n = -6$, irrespective of the value of n. Finally, by taking into account the numerical values of coefficients A_3, B_3, D_3, E_3 and F_3 in Eq. (2), we obtain:

$$A_n = 1 \tag{14}$$

$$B_n = 3 \tag{15}$$

$$D_n = -3(n - 1) \tag{16}$$

$$E_n = -6 \tag{17}$$

$$F_n = 3n - 4. \tag{18}$$

Coefficients A_n, B_n, D_n, E_n and F_n may be used to obtain W for any starlike graph in terms of the sizes of strings. The coefficients are still valid if $n = 2$ (see Appendix).

The present procedure can be applied to obtain the coefficients of any graph consisting of two »starlike« subgraphs. The simplest case is depicted in Figure 5. The polynomial we want to derive for this class of homeomorphic graphs should have the following form:

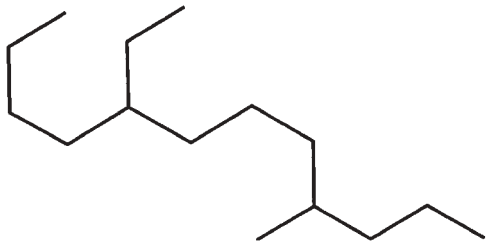

Figure 5. Example of a graph containing two starlike subgraphs. The branching vertices belong to three incident chains. $k = 2$, $j = 4$, $m = 3$, $n = 5$, and (the string linking the two branching vertices) $x = 5$.

$$W = [(A_{22}(k^3 + o^3 + m^3 + n^3) + A'_{22}x^3 + B_{22}(k^2o + ko^2 + m^2n + mn^2) + \\ + B'_{22}(k^2m + k^2n + o^2m + o^2n + km^2 + kn^2 + om^2 + on^2) + B''_{22}(k^2x + \\ + o^2x + m^2x + n^2x) + B^+_{22}(kx^2 + ox^2 + mx^2 + nx^2) + C_{22}(kxm + kxn + \\ + oxm + oxn) + D_{22}(k^2 + o^2 + m^2 + n^2) + D'_{22}x^2 + E_{22}(ko + mn) + \\ + E'_{22}(km + om + kn + on) + E'''_{22}(kx + ox + mx + nx) + \\ + F_{22}(k + o + m + n) + F'_{22}x + G_{22}]/6 \quad (19)$$

where A_{22}, A'_{22}, B_{22}, B'_{22}, B''_{22}, B^+_{22}, C_{22}, D_{22}, D'_{22}, E_{22}, E'_{22}, E'''_{22}, F_{22}, F'_{22} and G_{22} are the coefficients to be determined. The subscripts indicate that two side-chains emerge from both ends of the central chain x.

Observe that for $x = 1$ the resulting graph is a starlike graph having four strings (Figure 3). Comparison with coefficients of terms k^2, k, km and the constant in Eq. (7) yields:

$$B'''_{22} + D_{22} = D_4 = -9 \quad (20)$$

$$B^+_{22} + E'''_{22} + F_{22} = F_4 = 8 \quad (21)$$

$$C_{22} + E'_{22} = E_4 = -6 \quad (22)$$

$$A'_{22} + D'_{22} + F'_{22} + G_{22} = 0. \quad (23)$$

By choosing $m = n = 1$, a starlike graph with three side-chains (Figure 3) results. Comparison with coefficients of terms k^2, k, x^2, x, k^2x, kx, kx^2 and the constant in Eq. (7) yields:

$$2B'_{22} + D_{22} = D_3 = -6 \quad (24)$$

$$2B'_{22} + 2E'_{22} + F_{22} = F_3 = 5 \quad (25)$$

$$2B^+_{22} + D'_{22} = D_3 = -6 \quad (26)$$

$$2B''_{22} + 2E''_{22} + F'' = F_3 = 5 \tag{27}$$

$$B''_{22} = B_3 = 3$$

$$2C_{22} + E''_{22} = E_3 = -6 \tag{28}$$

$$B^{+}_{22} = B_3 = 3 \tag{29}$$

$$2A_{22} + 2B_{22} + 2D_{22} + E_{22} + 2F_{22} + G_{22} = 0. \tag{30}$$

By choosing $k = o = m = n = 1$, a simple chain results, and the chain formula does not contain second and zero order (constant) terms (see Appendix, Eq. (A1)), whereas the coefficient of x (the first order term) is equal to –1.

$$4B^{+}_{22} + D' = 0 \tag{31}$$

$$4B''_{22} + 4C_{22} + 4E''_{22} + F'' = -1 \tag{32}$$

$$4A_{22} + 4B_{22} + 8B'_{22} + 4D_{22} + 2E_{22} + 4F_{22} + G_{22} = 0 \tag{33}$$

Finally, by choosing $o = n = 1$ again a simple chain results, consisting of three parts (see Appendix, Eq. (A4)). Comparison of coefficients k^2, k, kx and the constant yields the following equations, respectively:

$$B_{22} + B'_{22} + D_{22} = -6 \tag{34}$$

$$B_{22} + B'_{22} + E_{22} + E'_{22} + F_{22} = 11 \tag{35}$$

$$C_{22} + E''_{22} = -12 \tag{36}$$

$$2A_{22} + 2B'_{22} + 2D_{22} + E'_{22} + 2F_{22} + G_{22} = -6\ . \tag{37}$$

Solution of the system of linear equations (20)–(37) yields the coefficients of Eq. (19). Numerical values of the coefficients are listed in Table I.

A similar procedure can be used to calculate the coefficients for a graph containing two branching atoms of degree four (Figure 6), and in general for any graph containing two branching atoms with a degree of $k + 1$ (Table I).

DISCUSSION

The formulas presented in this paper may be used to solve the problem of »inverse quantitative structure – activity relationship«,[13,14] at least for starlike graphs or graphs consisting of two starlike subgraphs. Quantitative structure-activity relationships (QSAR) denote statistical investigations in which the numerical values of some activity (or property) of the molecules are related to a number (or numbers) depending on the structure of the molecule. The independent variables appearing in QSARs are quite often

TABLE I

Coefficients of polynomials related to graphs containing two starlike subgraphs (Figure 5)

Coefficient	$n = 2$	$n = 3$	$n = k$
$A_{nn} = A'_{nn}$	1	1	1
$B_{nn} = B'_{nn} = B''_{nn} = B^{+}_{nn}$	3	3	3
C_{nn}	6	6	6
$D_{nn} = D'_{nn}$	–12	–18	$-6k$
E_{nn}	–6	–6	–6
E'_{nn}	–12	–12	–12
E''_{nn}	–18	–24	$-6(k + 1)$
F_{nn}	23	35	$12k - 1$
F'_{nn}	35	71	$6k^2 + 6k - 1$
G_{nn}	–24	–54	$-6k^2$

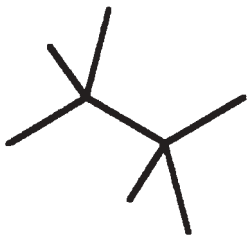

Figure 6. Scheme of a graph containing two starlike subgraphs. The branching vertices belong to four incident chains.

graph invariants, particularly W is an important variable.[1–3] As a result of QSAR investigations, the activity of a molecule is expressed as a function of graph invariants characterizing its structure. The aim of the inverse QSAR problem is to find out which molecules have the desired property. In order to solve this problem, we have to generate all graphs possessing a definite value of W (or any other graph invariant).

In such a procedure, the minimal and maximal[7,15] values of vertices for which graphs with a definite W may exist have to be determined first. Then, formulas derived in this work can be used to generate all starlike graphs or graphs consisting of two starlike subgraphs containing an allowed number of vertices and to select all those structures which correspond to the fixed value of W. The time needed to perform this task depends on the number of strings to be considered. The number of strings will necessarily be limited. As a result of this procedure, the same graph might be considered several times, and a method to avoid this difficulty has to be found.

The generation procedure described above is restricted to starlike graphs or structures for which formulas have already been dervied.[11] The solution of the inverse QSAR problem applicable to *any* three has still to be solved.

APPENDIX

Calculate coefficients A_2, B_2, D_2, E_2 and F_2. A »starlike« graph consisting of two side-chains – or stings – is in fact a chain consisting of two »sub-chains« a and b. The Wiener formula of W derived for N-chains is:[1]

$$W = (N^3 - N)/6 \tag{A1}$$

and

$$N = a + b - 1 \tag{A2}$$

$$(a + b - 1)^3 - (a + b - 1) = a^3 + b^3 + 3a^2b + 3ab^2 - 3a^2 - 3b^2 - 6ab + 2a + 2b. \tag{A3}$$

Therefore, $A_2 = 1$, $B_2 = 3$, $D_2 = 3$, $E_2 = -6$ and $F_2 = 2$, in accordance with formulas (13) – (17). Similarly,

$$(a + b + c - 2)^3 - (a + b + c - 2) = (a^3 + b^3 + c^3) + 3(a^2b + ab^2 + a^2c + ac^2 + b^2c + bc^2) + 6abc - 6(a^2 + b^2 + c^2) - 12(ab + ac + bc) + 11(a + b + c) - 6. \tag{A4}$$

REFERENCES

1. H. Wiener, *J. Am. Chem. Soc.* **69** (1947) 2636–2638.
2. I. Gutman, Y. N. Yeh, S. L. Lee, and Y. L. Luo, *Indian J. Chem.* **32A** (1993) 651–661.
3. S. Nikolić, N. Trinajstić, and Z. Mihalić, *Croat. Chem. Acta* **68** (1995) 105–129.
4. P. G. Doyle and J. L. Snell, *Random Walks and Electric Networks*, Math. Assoc., Washington, 1984.
5. J. K. Doyle and J. E. Graver, *Discrete Math.* **17** (1977) 147–154.
6. (a) D. Amić and N. Trinajstić, *Croat. Chem. Acta* **68** (1995) 53–62,
   (b) M. V. Diudea, *J. Chem. Inf. Comput. Sci.* **36** (1996) 535–540,
   (c) M. V. Diudea and I. Gutman, *Croat. Chem. Acta* (1998) (to appear),
   (d) A. A. Dobrynin and I. Gutman, *Graph Theory (New York)* **28** (1995) 21–23,
   (e) A. Graovac and T. Pisanski, *J. Math. Chem.* **8** (1991) 53–62,
   (f) O. Ivanciuc and A. Balaban, *Commun. Math. Chem. (MATCH)* **30** (1994) 141–152, P. John, *Commun. Math. Chem.* **32** (1995) 207–219, M. Randić, *Chem. Phys. Lett.* **211** (1993) 478–483, D. Rouvray and B. C. Crafford, *S. Afr. J. Sci.* **72** (1976) 47–51, S. S. Tratch, M. I. Stankevich, and N. S. Zefirov, *J. Comput. Chem.* **11** (1990) 899–908, D. J. Klein and M. Randić, *J. Math. Chem.* **12** (1993) 81–95, Z. Mihalić, S. Nikolić, and N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **32** (1992) 28–37.
7. R. C. Entringer, D. E. Jackson, and D. A. Snyder, *Czech Math. J.* **8** (1984) 1–21.

8. D. Bonchev, O. Mekenyan, J. V. Knop, and N. Trinajstić, *Croat. Chem. Acta* **52** (1979) 361–367.
9. D. Bonchev, O. Mekenyan, and N. Trinajstić, *Int. J. Quantum Chem.* **17** (1980) 845–893.
10. I. Gutman, *J. Mol. Struct. (Theochem)* **285** (1993) 137–142.
11. I. Lukovits,, *J. Chem. Inf. Comput. Sci.* **31** (1991) 503–507.
12. E. R. Canfield, R. W. Robinson, and D. H. Rouvray, *J. Comput. Chem.* **6** (1985) 598–609.
13. L. B. Kier, L. H. Hall, and J. W. Frazer, *J. Chem. Inf. Comput. Sci.* **33** (1993) 143–147.
14. M. I. Skvortsova, I. I. Baskin, O. L. Slovokhotova, V. A. Palyulin, and N. S. Zefirov, *J. Chem. Inf. Comput. Sci.* **33** (1993) 630–634.
15. D. Bonchev and N. Trinajstić, *J. Chem. Phys.* **67** (1977) 4517–4533.

SAŽETAK

Wienerov indeks: formule za nehomeomorfne grafove

István Lukovits

Wienerov indeks (W) invarijanta je grafa. Zbog njegove važnosti, predloženi su različiti postupci za njegovo izračunavanje, uključujući numeričke pristupe i analitičke formule. Već izvedene formule za W prikazane su specijalno za homeomorfne strukture, polimere i zvijezde, u polinomnom obliku. U ovom radu predložen je postupak za izračunavanje koeficijenata polinoma za skupove nehomeomorfnih struktura.