Formulas for the Hyper-Wiener Index of Trees

István Lukovits

Central Research Institute for Chemistry, Hungarian Academy of Sciences,
H-1525 Budapest, P.O. Box 17, Hungary

Received December 15, 1993®

Formulas for the hyper-Wiener index of chains and trees which contain one trivalent or tetravalent branching vertex are derived. The formulas are polynomials in terms of the lengths of the strings that form a tree. It is shown that the polynomial related to an arbitrary tree cannot contain fifth or higher order terms. A method to derive formulas for more complicated cases is given. It is suggested that in order to derive structure–property relationships the hyper-Wiener index has to be divided by the third power of the number of vertices of a particular graph.

INTRODUCTION

The hyper-Wiener index was proposed by Randić for quantitative structure-relationship studies.[1] This index will be denoted by R. Formally, R is equal to the sum of contributions K_{ij} of all pairs of vertices i and j of an acyclic tree:

$$R = \sum_{i<j} K_{ij} \qquad (1)$$

K_{ij} is obtained by the following algorithm: (1) replace the path between vertices i and j by a single bond; (2) multiply the number of vertices on the left-hand side by the number of vertices on the right-hand side of this bond. Diagonal elements K_{ii} are zero. n-Pentane may be used as a simple example (Figure 1). $K_{12} = 1 \times 4$, $K_{13} = 1 \times 3$, $K_{14} = 1 \times 2$, $K_{15} = 1 \times 1$, $K_{23} = 2 \times 3$, $K_{24} = 2 \times 2$, $K_{25} = 2 \times 1$, $K_{34} = 3 \times 2$, $K_{35} = 3 \times 1$, $K_{45} = 4 \times 1$, and the sum is 35 (see also eqs 5 and 6). The present definition cannot be used for graphs containing cycles.

R, like the simpler Wiener index[2] W, is a measure of the compactness of the molecule. Equation 1 is analogous to the definition of W:

$$W = \sum_{i<j} D_{ij} \qquad (2)$$

where D_{ij} denotes the distance between vertices i and j. W can also be obtained by summing contributions like in eq 1, but in this case just contributions due to neighboring vertices have to be taken into account. The procedure has been defined for trees only, but it could be extended for cyclic structures[3] too. Besides this approach, W can be obtained by numerical approaches used to construct the matrix of distances[4,5] and then using eq 2, analytical formulas derived for several classes of structures,[6-8] or the Laplacian matrix of trees.[9] More details concerning W may be found in a review.[10]

In order to calculate R, a numerical method, the Wiener matrix technique, was proposed by Randić et al.[11] The aim of the present paper was to derive formulas for simple structures. The procedure is based on a theorem that will be proved in the next section.

THEORY

The expressions "graph" and "chemical formula", "atom", and "vertex", and "chemical bond" and "edge" will be used interchangeably hereafter. Hydrogen suppressed formulas will be considered throughout the paper. First a simple formula yielding R for chains will be derived; then it will be shown that R is a fourth-order polynomial in terms of the numbers of vertices in strings, which are special subgraphs.

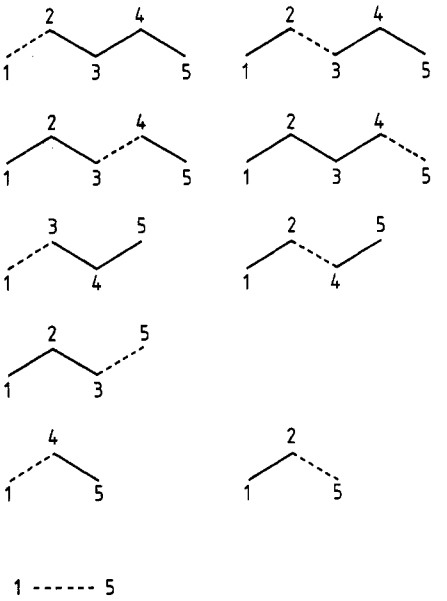

Figure 1. An example illustrating derivation of eq 5. Dashed lines denote the edge which intersects the chain.

The term "string" denotes hereafter a sequence of vertices, the first vertex of which is an end point (i.e. a point of degree 1) or a branching point (i.e. a point of degree ≥ 3) and the last vertex of which is an end point or a branching point. Strings may contain any number of vertices of degree two between end points and/or branching points (example: 3-methylhexane). This structure is composed of three strings: (1) a string involving two vertices; (2) a string involving three vertices; (3) a string involving four vertices. In this way the same branching point appears in three different strings, and the total number of vertices N is therefore equal to $2 + 3 + 4 - 2 = 7$. A chain is a single string. Strings will be denoted by small letters k, m, n, and x. Note that these letters also denote the *size* of a string, i.e. the number of vertices in k, m, n, and x.

Expression "Y-tree" denotes a tree containing a single branching point, the degree of which is 3, and the term "X-tree" denotes any tree containing a single branching point, the

® Abstract published in *Advance ACS Abstracts*, September 1, 1994.

0095-2338/94/1634-1079$04.50/0

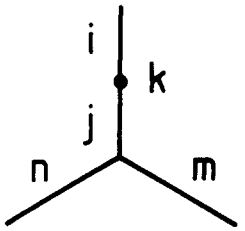

Figure 2. Scheme of a Y-tree. String k has been partitioned.

degree of which is 4. 2-Methylhexane is a Y-tree; 2,2-dimethylpropane (neopentane) is an X-tree.

The hyperconnectedness S_i of vertex i will be defined analogously to the connectedness index[12] s_i:

$$S_i = \sum_j C_{ij} \tag{3}$$

$$R = {}^1/_2 \sum_i S_i \tag{4}$$

The expression "N-chain" will denote a chain consisting of N vertices. In an N-chain there are $N-1$ distances of length 1, $N-2$ distances of length 2, etc., and one distance of length $N-1$. R is therefore equivalent to the sum of the Wiener indices of an N-chain (W_N), an $N-1$ chain (W_{N-1}), etc. and an 1-chain (W_1).

$$R_N = W_N + W_{N-1} + ... + W_2 + W_1 \tag{5}$$

where R_N denotes the R of an N-chain. The procedure is illustrated for n-pentane in Figure 1. Using the formulas for N-chains ($W_N = (N^3 - N)/6$), we obtain after some algebra the following expression:

$$R_N = N(N+1)[N(N+1) - 2]/24 = (k^4 + 2k^3 - k^2 - 2k)/24 \tag{6}$$

where $k = N$ and k was used to indicate that a chain is a string. Note that this expression is also valid for methane ($k = 1$) and $R_1 = 0$. Next it will be shown that for any tree T, R is a fourth-order polynomial in terms of the number of vertices in strings.

Theorem. Let T be a tree. The hyper Wiener index of this tree R is a fourth-order polynomial in terms of k, m, ..., x, where k, m, n, ..., x denote the sizes of strings in T.

Proof. Let r and s be strings attached to one end of k and m and n be strings attached to the other end of k. If $m = n = 1$, then k has one end point. If $m = n = r = s = 1$, then k is a chain. String k may be subdivided into components i and j, such that $k = i + j - 1$ (Figure 2). Then hyperconnectedness S_i ($i \in k$) is

$$\begin{aligned} S_i = {} & (i+r+s-2)[(j-1+m+n-2) + \\ & (j-2+m+n-2) + ... + (1+m+n-2)] + \\ & (j+m+n-2)[(i-1+r+s-2) + (i-2+r+s-2) + \\ & ... + (1+r+s-1)] = (i+r+s-2)[(k-i+1) \\ & (k-i)/2 + (k-i)(m+n-2)] + (k-i+1+m+n-2) \\ & [i(i-1)/2 + (i-1)(r+s-2)] \end{aligned} \tag{7}$$

where relation $j = k - i + 1$ has been used. S_i is a third-order polynomial in i, k, m, n, r, and s. Sum over i ($i = 1, 2, ..., k$) yields a fourth-order polynomial in terms of k, m, n, r, and s, because the sum $1^3 + 2^3 + 3^3 + ... + k^3$ yields a fourth-order polynomial in k and i^3 terms appear alone. The same procedure may be used for strings m, n, r, and s, or in fact for any number of strings attached to the left- and right-hand sides of k.

The formulas obtained for S_i do not depend on the actual structure of m, n, r, and s; therefore these may also be subgraphs containing branches. If summation of the connectedness numbers $S_i - s$ is performed for all strings and the contribution of each braching point (a third-order polynomial) is subtracted $Z_j - 1$ times from the sum, where Z_j is the number of strings incident with a braching point j, we obtain R (eq 3), which is a fourth-order polynomial. This completes the proof.

Remark. Although the derivation of formulas for R of a Y-tree (or an X-tree) is straightforward, we will use a numerical technique (next section), since the numerical method can also be used in more complicated cases.

NUMERICAL EXAMPLES

For an arbitrary Y-tree composed of k, m, and n strings (where k, m, and n denote both the strings and the number of vertices of each string), the following terms are expected to appear in R: k^4, m^4, n^4, k^3m, km^3, k^3n, kn^3, m^3n, mn^3, k^2m^2, k^2n^2, m^2n^2, k^2mn, km^2n, kmn^2, k^3, m^3, n^3, k^2m, km^2, k^2n, kn^2, m^2n, mn^2, kmn, k^2, m^2, n^2, km, kn, mn, k, m, and n, altogether 34 terms. A constant cannot appear in R, because $R_1 = 0$, and methane is a special Y-tree with $k = m = n = 1$. Hence, we may write

$$R = c_1k^4 + c_2m^4 + c_3n^4 + c_4k^3m + ... + c_{34}n \tag{8}$$

where c_i ($i = 1, ..., 34$) are the coefficients to be determined. Therefore the desired formula might be obtained[6] by calculating R of 34 different structures and by solving a system of 34 linear equations of the type of eq 8. Because of symmetry the coefficients of terms 1–3, 4–9, 10–12, 13–15, 16–18, 19–24, 25, 26–28, 29–31, and 32–34 must be equal within each group. These terms can be grouped together and by taking into account 10 different structures, with their respective values R, we can derive 10 independent linear equations, the unknowns of which are the coefficients of the terms. The coefficients could be obtained by an matrix inversion and multiplication.[6] The coefficients cannot depend on the actual structures which were used to derive them. However, since quite often the determinant is zero, we list the structures in Table 1 which were used to derive the coefficients.

Table 2 lists the formulas of R obtained in this way. Terms 13–15 and 25 do not appear in the polynomial. It should be noted that three-parameter terms also do not appear in the formulas of the Wiener index, if the respective strings do not belong to the same path.

The same technique was used to derive the respective formula for an arbitrary X-tree in terms of strings k, m, n, and x. In this case three-parameter terms were not considered, so only an eight-dimensional system of linear equations had to be solved. Table 1 lists those structures that were used to obtain the coefficients, and Table 2 shows the polynomial.

DISCUSSION

The formula of Y-trees remains valid if one or two of the strings (e.g. say m and n) are of size 1, and we have a chain. The formula obtained for X-trees remains valid if one of the strings (e.g. x) is of size one and we have a Y-tree, but it is also valid if two or three strings (e.g. m, n, and x) are of size 1 and we have a chain. All formulas in Table 2 are valid for methane, with $k = 1$ in chains, $k = m = n = 1$ in Y-trees, and $k = m = n = x = 1$ in X-trees.

Table 1. Training Sets Used To Derive Formulas for Y- and X-Trees[a]

Y-Trees

no.	k	m	n	R
1	1	1	1	0
2	1	1	2	1
3	2	1	2	5
4	2	2	2	12
5	3	2	2	28
6	2	3	3	54
7	5	2	2	108
8	3	3	4	150
9	4	4	3	231
10	5	4	3	351

X-Trees

no.	k	m	n	x	R
1	2	2	2	1	12
2	4	3	3	2	188
3	4	3	3	3	255
4	5	4	3	2	414
5	2	2	2	2	22
6	4	3	4	2	279
7	3	3	2	2	76
8	4	3	2	2	131

[a] Table 2.

Since $R = 0$ in methane, the sum of coefficients in all equations (Table 2) must be zero. In fact in the formula obtained for chains, the sum of coefficients is $1 + 2 - 1 - 2 = 0$. For Y-trees the sum of coefficients is $3 \times 1 + 6 \times 4 + 3 \times 6 - 9 \times 6 - 3 \times 1 - 3 \times 2 + 3 \times 6 = 0$, and for X-trees the sum of coefficients is $4 \times 1 + 12 \times 4 + 6 \times 6 - 4 \times 10 - 12 \times 6 - 4 \times 1 - 6 \times 2 + 4 \times 10 = 0$.

The sum of pure (i.e. one-letter) terms must also be zero; for example, if $m = 1$, then this string does not contribute to R. This sum of coefficients is in Y-trees: $1 - 6 - 1 + 6 = 0$ for k, m, and n. Similarly the sum of coefficients is in X-trees: $1 - 10 - 1 + 10 = 0$ for k, m, n, and x.

The sum of coefficients of two-letter terms, e.g. k and m, must also vanish, because if $k = m = 1$, then neither of these strings contributes to R and therefore the sum of coefficients of all $k^a m^b$ ($2 < a + b < 4$) terms must be zero. For Y-trees the sum of coefficients of terms involving k and m is $4 + 4 + 6 - 6 - 6 - 2 = 0$, and there are identical sums for k and n and m and n. For X-trees the sum of coefficients of terms involving k and m is $4 + 4 + 6 - 6 - 6 - 2 = 0$ for pairs km, kn, kx, mn, mx, and nx. The coefficients of two-letter terms are identical in Y- and X-trees.

Formulas are more effective than numerical procedures, if structures with an identical value of R are searched. In their paper Randić et al.[11] have reported that R of 2,3,4-trimethylhexane and of 3-methyl-3-ethylhexane is equal to 188. By use of the present formulas a pair of structures with a nonequal number of carbons with an identical value of R could be detected: $R = 170$ for 3-methylheptane and 3,3-diethylpentane.

Finally, the present formulas may be too complex for highly symmetrical structures, which may involve a great number of strings to be investigated. The simplest such structure is a star, which is a complete bipartite graph,[13] $K_{1,n}$, where n denotes the number of branches and $N = n + 1$. $R_{star} = n^2 + n(n-1)/2 = n[n + (n-1)/2]$. After some manipulation we get

$$R_{star} = n(1.5n - 0.5) \quad (9)$$

Therefore if $n = 0$ (methane), $R = 0$; if $n = 1$ (ethane), $R = 1$; if $n = 2$ (propane), $R = 5$; if $n = 3$ (isobutane), $R = 12$; and if $n = 4$ (2,2-dimethylpropane), $R = 22$. The formulas of more complex trees will be considered in another paper.[14]

Table 2. Formulas in Terms of Strings

graph	$R =$
chains	$(k^4 + 2k^3 - k^2 - 2k)/24$
Y-trees	$(k^4 + m^4 + n^4 + 4k^3m + 4km^3 + 4k^3n + 4kn^3 + 4m^3n + 4mn^3 + 6k^2m^2 + 6k^2n^2 + 6m^2n^2 - 6k^3 - 6m^3 - 6n^3 - 6k^2m - 6km^2 - 6k^2n - 6kn^2 - 6m^2n - 6mn^2 - k^2 - m^2 - n^2 - 2km - 2kn - 2mn + 6k + 6m + 6n)/24$
X-trees	$(k^4 + m^4 + n^4 + x^4 + 4k^3m + 4km^3 + 4k^3n + 4kn^3 + 4k^3x + 4kx^3 + 4m^3n + 4mn^3 + 4m^3x + 4mx^3 + 4n^3x + 4nx^3 + 6k^2m^2 + 6k^2n^2 + 6k^2x^2 + 6m^2n^2 + 6m^2x^2 + 6n^2x^2 - 10k^3 - 10m^3 - 10n^3 - 10x^3 - 6k^2m - 6km^2 - 6k^2n - 6kn^2 - 6k^2x - 6kx^2 - 6m^2n - 6mn^2 - 6m^2x - 6mx^2 - 6n^2x - 6nx^2 - k^2 - m^2 - n^2 - x^2 - 2km - 2kn - 2kx - 2mn - 2mx - 2nx + 10k + 10m + 10n + 10x)/24$

The polynomials may also be used to find different structures with identical R. It should be noted that the method listed in this paper could be used to derive formulas for any acyclic structure too. Since the present formulas are valid for a limited set of structures, an efficient numerical method would also be needed to calculate R for an arbitrary tree.

ACKNOWLEDGMENT

The author is indebted to Dr. D. J. Klein (Galveston, TX) for helpful discussions. Thanks are due to Dr. K. S. Pitre (India) for correcting the language of the text.

REFERENCES AND NOTES

(1) Randić, M. Novel Molecular Descriptor for Structure-Property Studies. *Chem. Phys. Lett.* **1993**, *211*, 478–483.
(2) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20. Correlation of Heats of Isomerization and Differences in Heats of Vaporization of Isomers among the Paraffin Hydrocarbons. *J. Am. Chem. Soc.* **1947**, *69*, 2636–2638.
(3) Lukovits, I. Wiener Indices and Partition Coefficients of Unsaturated Hydrocarbons. *Quant. Struct.-Act. Relat.* **1990**, *9*, 227–231.
(4) Mohar, B.; Pisanski, T. How to Compute the Wiener Index of a Graph. *J. Math. Chem.* **1988**, *2*, 267–277.
(5) Müller, W. R.; Szymanski, K.; Knop, J. V.; Trinajstić, N. An Algorithm for Calculation of the Molecular Distance Matrix. *J. Comput. Chem.* **1987**, *8*, 170–173.
(6) Lukovits, I. General Formulas for the Wiener Index. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 503–507.
(7) Bonchev, D.; Mekenyan, O.; Trinajstić, N. Topological Characterization of Cyclic Structures. *Int. J. Quantum Chem.* **1980**, *17*, 845–888.
(8) Canfield, E. R.; Robinson, R. W.; Rouvray, D. H. Determination of the Wiener Molecular Branching Index for a General Tree. *J. Comput. Chem.* **1985**, *6*, 598–609.
(9) Mohar, B.; Babić, D.; Trinajstić, N. A Novel Definition of the Wiener Index for Trees. *J. Chem. Inf. Comput. Sci.* **1993**, *23*, 153–154.
(10) Gutman, I.; Yeh, Y. N.; Lee, S. L.; Luo, Y. L. Some recent results in the theory of the Wiener number. *Indian J. Chem.* **1993**, *32A*, 651–661.
(11) Randić, M.; Guo, X. F.; Oxley, T.; Krishnapriyan, H. Wiener matrix: Source of novel graph invariants. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 709–716.
(12) Seybold, P. G. Topological influences on the carcinogenicity of aromatic hydrocarbons. I. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1983**, *10*, 95–102.
(13) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969.
(14) Lukovits, I. A formula for the hyper Wiener index of +– – –+ trees. *Comput. Chem.*, submitted for publication.