Frequency of Even and Odd Numbers in Distance Matrices of Trees

István Lukovits

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

Received March 8, 1993

The frequency of even and odd numbers in distance matrices of acyclic graphs was investigated. The concept of a triad, made up of distances between three points, was introduced. It was shown that only two classes of triads are allowed, namely even–even–even and even–odd–odd. On the basis of this result, the allowed frequencies of even and odd numbers in arbitrary distance matrices could be determined. Expressions providing the number of *eee*-triads, *eoo*-triads, $e^m o^n$ and $e^{n-1}o^{m+1}$ vertices were also given.

INTRODUCTION

The topological distance d_{ij} is the minimal number of edges separating vertices i and j in a graph. Distance matrices play an important role in chemical graph theory, where a number of graph invariants are based on topological distances.[1] The first one is the Wiener index[2] W, which is the sum of the entries in the upper triangle of the distance matrix. A closely related index to W is Seybold's atomic structural index, which is the sum of topological distances from a reference atom.[3] More complex indices are Balaban's J index[4] and Schultz' molecular topological indices.[5] Distance matrices are also used to form topological molecular transforms.[6] In trees the distances are equivalent with the paths connecting two vertices,[7,8] but in cyclic structures the distance and a path connecting the same atoms (vertices) may be non-equivalent. Paths and linear substructures may be used to relate structural and physical properties of molecules.[9] The concept of the vertex–vertex distance has been extended to structures containing heteroatoms (i.e. colored vertices).[10] More details on this topic may be found in the literature.[11]

Distance matrices may be calculated by raising the adjacency matrix to its higher powers[12] or by the more effective algorithms by Bersohn[13] and later by Mueller et al. and by Mohar and Pisanski.[14] On the eigenvalues of distance matrices, the reader is referred to the paper of Balasubramanian.[15]

The aim of this paper was to investigate the frequency of even and odd numbers in the distance matrix of a tree. The results indicated that only a few combinations, as compared to the total number of edges in a graph containing N vertices, can be realized. The results may be used to discard certain substructures in the inverse imaging procedure.[16]

THEORY

The expressions "structural formula of a molecule" and "graph" will be used interchangeably hereafter. The atoms of the molecule are the vertices, whereas chemical bonds correspond to the edges in this graph. If there are N vertices, then there are $N(N-1)/2$ distances. In this paper edges will denote distances in a complete graph consisting of N vertices of the underlying graph and $N(N-1)/2$ edges, that are colored due to the parity of the respective distance. An e-edge corresponds to an even distance and an o-edge refers to a distance that is odd. Three vertices define a triangle, and the respective distances form a triad. Triads (and the corresponding triangles) may be characterized by the parity of the constituent numbers; e.g. three even numbers make up an *eee*-triad, one even and two odd numbers make up an *eoo*-triad.

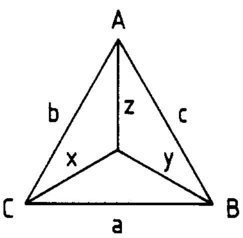

Figure 1. Scheme of a triangle. Edges a–c represent distances between vertices A–C; x–z are the lengths of the connecting paths.

Theorem 1. In distance matrices of trees, only *eee*-triads and *eoo*-triads are allowed.

Proof. Consider a triad, made up from topological distances a–c (Figure 1) and $a \geq b \geq c$, and $a \leq b + c$. a–c may be expressed in terms of paths x–z, connecting vertices A–C.

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad (1)$$

The inverse of this (matrix) equation is

$$\begin{pmatrix} +0.5 & +0.5 & -0.5 \\ +0.5 & -0.5 & +0.5 \\ -0.5 & +0.5 & +0.5 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (2)$$

x–z are integers. Because of the multiplying factor 0.5, either a–c should be even or two of them should be odd and the third number should be even in order to ensure that the resulting x–z are integers. The theorem has been proven.

This result means that neither *eeo*-triads nor *ooo*-triads are allowed. Once the parities of two edges have been fixed, the parity of the third edge is also fixed. Formally it may be written

$$e + e = e \quad (3)$$

$$e + o = o \quad (4)$$

$$o + o = e \quad (5)$$

Equations 3–5 were shown in Figure 2a–c.

In order to obtain the number of e-edges E_e and the number of o-edges E_o, we introduce the concept of the generating operator. A generating operator consists of a vertex and $N - 1$ edges with predetermined parities, incident with this vertex. If there are m e-edges incident with this vertex and n o-edges incident with this vertex, we have an $e^m o^n$ operator, or

0095-2338/93/1633-0626$04.00/0

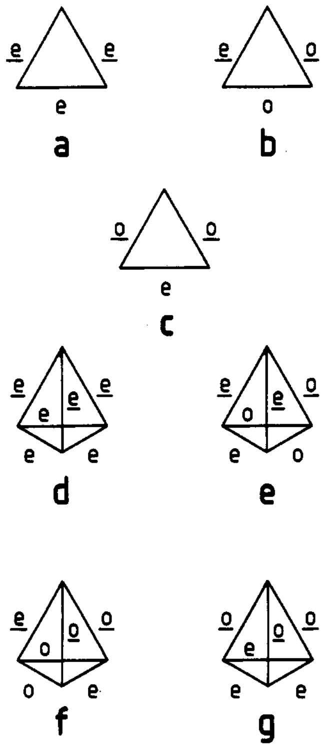

Figure 2. Generation of all possible combinations of *e*-edges and of *o*-edges in trees with three and four vertices. Underlined letters indicate that the respective edge is a part of a generating operator.

alternatively an $e^m o^n$ vertex. This operator acts onto a complete graph containing $N-1$ vertices and $(N-1)(N-2)/2$ edges by joining the "free" ends of the edges with one of the $N-1$ vertices. As a result the letter *e* or *o* may be attached to the $(N-1)(N-2)/2$ edges unambiguously, and a complete graph containing N vertices and $N(N-1)/2$ colored edges is obtained. The next theorem states that there are pairs of operators which are equivalent, with one exception, e^{N-1}, which has no equivalent pair.

Theorem 2.

$$e^m o^n = e^{n-1} o^{m+1}, \quad m+n+1=N \tag{6}$$

Proof. An *o*-edge of operator $e^m o^n$ is selected, and this edge is given as incident with a vertex x. Each of the remaining $n-1$ *o*-edges will contribute with the first *o*-edge an *e*-edge, that are incident with x, m *e*-edges will contribute together with the first *o*-edge $m+1$ *o*-edges, which are incident with vertex x. Hence, there will be $m+1$ *o*-edges and $n-1$ *e*-edges, which are incident with vertex x. Therefore operators $e^m o^n$ and $e^{n-1} o^{m+1}$ generate the same graph. If an *e*-edge is selected at first, we get again an $e^m o^n$ vertex. The theorem has been proven. $e^m o^n$ and $e^{n-1} o^{m+1}$ will be referred to as conjugate pairs of operators. If N is an even number, there is an operator which is equivalent with its own conjugate operator namely $e^{N/2-1} o^{N/2}$.

Corollary. Operator e^{N-1} has no conjugate pair.

The operator technique allows to determine the number of *e*-edges and of *o*-edges in an arbitrary acyclic graph. Examples for graphs containing four vertices and five vertices are shown in Figure 2d–g and in Figure 3, respectively. For any graph containing N vertices the number of *e*-edges, E_e, is equal the number of ways it is possible to select two *e*-edges out of m *e*-edges plus the number of ways it is possible to select two

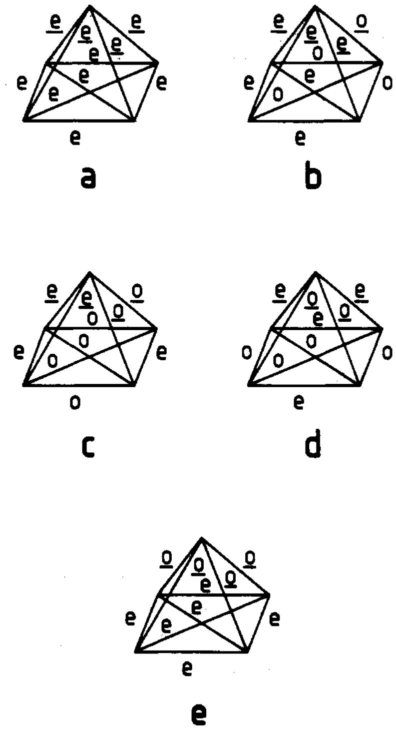

Figure 3. Generation of all possible combinations of *e*-edges and of *o*-edges in trees with five vertices. Underlined letters indicate that the respective edge is a part of the generating operator.

Table I. Number of *e*-Edges, *o*-Edges, *eee*-Triads, and *eoo*-Triads in Distance Matrices of Trees with up to Six Vertices

operator	N	*e*-edges	*o*-edges	*eee*-triads	*eoo*-triads
e	2	1	0		
o	2	0	1		
e^2	3	3	0	1	0
eo	3	1	2	0	1
o^2	3	1	2	0	1
e^3	4	6	0	4	0
e^2o	4	3	3	1	3
eo^2	4	2	4	0	4
o^3	4	3	3	1	3
e^4	5	10	0	10	0
e^3o	5	6	4	4	6
e^2o^2	5	4	6	1	9
eo^3	5	4	6	1	9
o^4	5	6	4	4	6
e^5	6	15	0	20	0
e^4o	6	10	5	10	10
e^3o^2	6	7	8	4	16
e^2o^3	6	6	9	2	18
eo^4	6	7	8	4	16
o^5	6	10	5	10	10

o-edges out of n *o*-edges plus m.

$$E_e = \binom{m}{2} + \binom{n}{2} + m \tag{7}$$

Similarly, the number of *o*-edges, E_o, is equal to m by n (the number of resulting *o*-edges) plus n, the number of edges introduced by the generating operator:

$$E_o = mn + n \tag{8}$$

The allowed partitions of *e*-edges and *o*-edges in trees with up to six vertices are listed in Table I. It can be seen that partitions due to $e^m o^n$ and $e^{n-1} o^{m+1}$ are equal. e^{N-1} produces an "all *e*-edges" graph. The partitions due to operator $e^{N/2-1} o^{N/2}$ (N is even) will occur only once in the listing (e.g. for $e^2 o^3$ for $N = 6$; Table I).

Table II. Number of Allowed Partitions of Even and Odd Numbers in Distance Matrices of Trees

N	partitions	percent of no. of edges + 1	N	partitions	percent of no. of edges + 1
3	2	50.0	15	8	7.5
4	3	42.9	20	11	5.8
5	3	27.3	30	16	3.4
6	4	25.0	40	21	2.7
7	4	18.2	50	26	2.1
8	5	17.2	100	51	1.0
9	5	13.5	500	251	0.2
10	6	13.0			

The total number of partitions P for a general tree containing N vertices is

$$P = (N-2)/2 + 2 = (N+2)/2 \quad N \text{ even} \tag{9}$$

$$P = (N-1)/2 + 1 = (N+1)/2 \quad N \text{ odd} \tag{10}$$

Thus for $N = 3$ there are two different allowed partitions of e-edges and o-edges; the total number of distances is 3. For $N = 6$, there are four different allowed partitions of e-edges and o-edges; the total number of distances is 15. The 15 adjacent edges could be marked (colored) with letters e and o in 16 different ways (15 e-edges, 14 e-edges + 1 o-edge, etc). The ratios of allowed combinations to all combinations are listed for different values of N, in Table II.

The number of eee (T_e) triads in a pattern obtained by operator $e^m o^n$ (with $m + n + 1 = N$) is equal to the number of ways three e-edges can be selected from m e-edges plus the number of ways two e-edges can be selected from m e-edges plus the number of ways three o-edges can be selected from n o-edges.

$$T_e = \binom{m}{3} + \binom{m}{2} + \binom{n}{3} \tag{11}$$

The number of eoo triads (T_o) is simply the difference between the total number of triads $\binom{N}{3}$ and T_e:

$$T_o = \binom{N}{3} - T_e \tag{12}$$

The vertices are also defined by the generating operator used to obtain the particular coloring of the edges. Thus operator $e^m o^n$ will also induce $e^{n-1}o^{m+1}$ vertices. The number of $e^m o^n$ and of $e^{n-1}o^{m+1}$ vertices is determined by theorem 3.

Theorem 3. In a distance matrix, that can be associated with an $e^m o^n$ operator, the number of $e^m o^n$ vertices is $m + 1$ and the number of $e^{n-1}o^{m+1}$ vertices is n.

Proof. The statement is clearly valid for operators e^{N-1}. In all other cases, selection of an o-edge from $e^m o^n$ gives rise to a vertex (see theorem 2) $e^{n-1}o^{m+1}$. Since there are n ways to choose an o-edge, there will be n distinct $e^{n-1}o^{m+1}$ vertices. Alternatively, selection of an e-edge gives rise to an $e^m o^n$ vertex. Since there are m ways to select an e-edge, there will be m e-edges plus one e-edge of the generating vertex, i.e. $m + 1$ distinct $e^m o^n$ vertices. This completes the proof.

Corollary. All vertices of graphs with distance matrices corresponding to the operators e^{N-1} or $e^{N/2-1}o^{N/2}$ (N is even in the latter) are e^{N-1} or $e^{N/2-1}o^{N/2}$ vertices, respectively.

NUMERICAL EXAMPLE

Figure 4 shows the hydrogen suppressed graph of 4-isopropyl-3,6,7-trimethylnonane, and Table III lists the distance matrix of this graph. In the upper triangle of this distance matrix there are 105 distances (edges), 49 of which are e-edges and 56 are o-edges. Because of this, the distance matrix (and the underlying graph) can be associated with the e^7o^7 operator (Table IV); therefore eight e^7o^7 vertices and seven e^6o^8 vertices are expected. It may be seen from Table III that vertices 1, 3, 5, 7, 8, 10, 12, and 14 are e^7o^7 vertices, whereas vertices 2, 4, 6, 9, 11, 13, and 15 are e^6o^8 vertices. Similarly, the total number of triads is 455. The number of eee-triads is 91, and the number of eoo-triads is 364 (Table IV). An algorithm could be constructed that lists all eee-triads and eoo-triads. The 91 eee-triads of 4-isopropyl-3,6,7-trimethylnonane are listed in Table V. Note that the same triad may occur several times because identical distances appear several times.

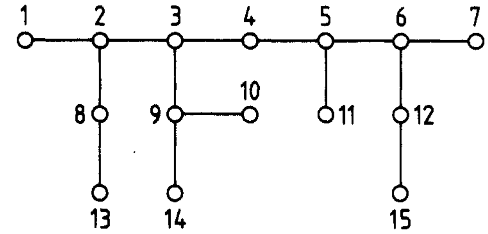

Figure 4. Hydrogen suppressed graph of 4-isopropyl-3,6,7-trimethylnonane.

Table III. Distance Matrix of 4-Isopropyl-3,6,7-trimethylnonane[a]

row	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	1	2	3	4	5	6	2	3	4	5	6	3	4	7
2	1	0	1	2	3	4	5	1	2	3	4	5	2	3	6
3	2	1	0	1	2	3	4	2	1	2	3	4	3	2	5
4	3	2	1	0	1	2	3	3	2	3	2	3	4	3	4
5	4	3	2	1	0	1	2	4	3	4	1	2	5	4	3
6	5	4	3	2	1	0	1	5	4	5	2	1	6	5	2
7	6	5	4	3	2	1	0	6	5	6	3	2	7	6	3
8	2	1	2	3	4	5	6	0	3	4	5	6	1	4	7
9	3	2	1	2	3	4	5	3	0	1	4	5	4	1	6
10	4	3	2	3	4	5	6	4	1	0	5	6	5	2	7
11	5	4	3	2	1	2	3	5	4	5	0	3	6	5	4
12	6	5	4	3	2	1	2	6	5	6	3	0	7	6	1
13	3	2	3	4	5	6	7	1	4	5	6	7	0	5	8
14	4	3	2	3	4	5	6	4	1	2	5	6	5	0	7
15	7	6	5	4	3	2	3	7	6	7	4	1	8	7	0

[a] Figure 4.

Table IV. Number of e-Edges, o-Edges, eee-Triads, and eoo-Triads in Distance Matrices of Trees with 15 Vertices

operator	e-edges	o-edges	eee-triads	eoo-triads
e^{14}	105	0	455	0
$e^{13}o$	91	14	364	91
$e^{12}o^2$	79	26	286	169
$e^{11}o^3$	69	36	221	234
$e^{10}o^4$	61	44	169	286
e^9o^5	55	50	130	325
e^8o^6	51	54	104	351
e^7o^7	49	56	91	364
e^6o^8	49	56	91	364
e^5o^9	51	54	104	351
e^4o^{10}	55	50	130	325
e^3o^{11}	61	44	169	286
e^2o^{12}	69	36	221	234
e^1o^{13}	79	26	286	169
o^{14}	91	14	364	91

CONCLUSIONS

Distances, if not arranged in a square matrix, are not sufficient to define a graph unambiguously.[17] However, it seems that distances are quite often tried in connection with other invariants to reproduce a graph. The techniques presented in this paper may contribute to the solution of the

Table V. All *eee*-Triads of 4-Isopropyl-3,6,7-trimethylnonane

(2,4,2)	(2,6,4)	(2,2,2)	(2,4,2)	(2,6,4)	(2,4,2)	(4,6,2)
(4,2,4)	(4,4,4)	(4,6,2)	(4,4,4)	(6,2,6)	(6,4,6)	(6,6,2)
(6,4,6)	(2,4,4)	(2,6,6)	(2,4,4)	(4,6,6)	(4,4,2)	(6,4,6)
(2,4,2)	(2,2,2)	(2,4,2)	(2,2,4)	(2,6,4)	(4,2,4)	(4,4,2)
(4,2,6)	(4,6,2)	(2,4,4)	(2,2,4)	(2,6,6)	(4,2,6)	(4,6,4)
(2,6,8)	(2,4,2)	(2,2,4)	(2,2,4)	(2,4,2)	(2,2,4)	(4,2,6)
(4,2,6)	(4,4,2)	(4,2,6)	(2,2,4)	(2,4,6)	(2,2,4)	(2,4,6)
(2,2,2)	(4,2,6)	(2,2,4)	(2,2,2)	(2,4,6)	(2,4,2)	(2,2,4)
(2,4,4)	(2,4,6)	(2,4,6)	(2,4,4)	(4,4,8)	(2,4,6)	(2,4,6)
(2,2,2)	(2,4,6)	(4,4,4)	(4,2,6)	(4,4,4)	(4,2,6)	(4,4,2)
(2,4,6)	(4,2,4)	(4,6,4)	(4,2,6)	(2,6,6)	(2,2,4)	(6,2,8)
(6,6,4)	(6,2,6)	(6,6,4)	(6,2,6)	(6,6,2)	(2,6,6)	(4,6,6)
(4,4,2)	(6,4,6)	(4,4,6)	(4,6,4)	(4,6,8)	(6,2,6)	(6,4,8)

graph reconstruction problem,[16] by determining the number of allowed *eee*-triads and *eoo*-triads, but alone they are insufficient for the reconstruction of a tree. In order to find an algorithm and to write a computer program that reconstructs graphs unambiguously, the sufficient conditions, that still have to be determined, should be used. The situation will be more complex in cyclic graphs, but in that case the number of *e*-edges and *o*-edges may also be controlled by some rules that have to be deduced.

REFERENCES AND NOTES

(1) (a) Rouvray, D. H. The Role of the Topological Distance Matrix in Chemistry. In *Mathematics and Computational Concepts in Chemistry*; Trinajstic, N., Ed.; Ellis Horwood: Chichester, U.K., 1986; Chapter 25, pp 295–306. (b) Bonchev, D.; Trinajstić, N. Information Theory, Distance Matrix and Molecular Branching. *J. Chem. Phys.* **1977**, *67*, 4517–4533.

(2) (a) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20. (b) Wiener, H. 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) Seybold, P. G. Topological Influences on the Carcinogenicity of Aromatic Hydrocarbons. II. Substituent Effects. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1983**, *10*, 103–108.

(4) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399–404.

(5) (a) Schultz, H. P. Topological Organic Chemistry. 1. Graph Theory and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227–228. (b) Schultz, H. P.; Schultz, E. B.; Schultz, T. P. Topological Organic Chemistry. 2. Graph Theory, Matrix Determinants and Eigenvalues and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 27–29. (c) Schultz, H. P.; Schultz, T. P. Topological Organic Chemistry. 3. Graph Theory, Binary and Decimal Adjacency Matrices and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 144–147. (d) Schultz, H. P.; Schultz, E. B.; Schultz, T. P. Topological Organic Chemistry. 4. Graph Theory, Matrix Permanents and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 69–72. (e) Schultz, H. P.; Schultz, T. P. Topological Organic Chemistry. 5. Matrix Hafnians, Pfaffians and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 364–368.

(6) Gabányi, Z.; Surján, P.; Náray-Szabó, G. Application of Topological Molecular Transforms to Rational Drug Design. *Eur. J. Med. Chem.-Chim. Ther.* **1982**, *17*, 307–311.

(7) Clark, M. J.; Kettle, S. F. A. Incidence and Distance Matrices. *Inorg. Chim. Acta* **1975**, *14*, 201–205.

(8) Randic, M.; Wilkins, C. L. Graph Theoretical Ordering of Structures as a Basis for Systematic Searches for Representation in Molecular Data. *J. Phys. Chem.* **1979**, *83*, 1525–1540.

(9) Garbalena, M.; Herndon, W. C. Optimum Graph-Theoretical Models for Enthalpic Properties of Alkanes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 37–42.

(10) Barysz, M.; Jashari, G.; Lall, R. S.; Srivastava, V.; Trinajstić, N. On the Distance Matrix of Molecules Containing Heteroatoms. *Stud. Phys. Theor. Chem.* **1983**, *28*, 222–230.

(11) (a) Mihalić, Z.; Nikolić, S.; Trinajstić, N. Comparative Study of Molecular Descriptors Derived from the Distance Matrix. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 28–37. (b) Mihalić, Z.; Veljan, D.; Amić, D.; Nikolić, S.; Trinajstić, N. The Distance Matrix in Chemistry. *J. Math. Chem.* **1992**, *11*, 223–258.

(12) Harary, F. *Graph Theory*; Addison-Wesley: 1969; p 151.

(13) Bersohn, M. A Fast Algorithm for Calculation of the Distance Matrix of a Molecule. *J. Comput. Chem.* **1983**, *4*, 110–113.

(14) (a) Muller, W. R.; Szymanski, K.; Knop, J. V.; Trinajstić, N. An Algorithm for Construction of the Molecular Distance Matrix. *J. Comput. Chem.* **1987**, *8*, 170–173. (b) Mohar, B.; Pisanski, T. How To Compute the Wiener Index of a Graph. *J. Math. Chem.* **1988**, *2*, 267–277.

(15) Balasubramanian, K. J. Computer Generation of Distance Polynomials of Graphs. *J. Comput. Chem.* **1990**, *11*, 829–836.

(16) Kier, L. B.; Hall, L. H.; Frazer, J. W. Design of Molecules from Quantitative Structure–Activity Relationship Models. 1. Information Transfer between Path and Vertex Degree Counts. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 143–147.

(17) Randić, M. Chemical Structure-What Is "She"? *J. Chem. Educ.* **1992**, *69*, 713–718.