

Further Properties Derivable from the Matula Number of an Alkane

Seymour B. Elk*

Elk Technical Associates, 321 Harris Place, New Milford, New Jersey 07646

Ivan Gutman

Faculty of Science, Svetozvar Markovic University, P.O. Box 60, YU-34000 Kragujevac, Serbia

Received May 18, 1993*

In probing the use and properties of Matula numbers, a formal proof of a theorem to determine what is the maximum vertex degree of a given acyclic graph—starting from its Matula number—is presented. Using this theorem, a simple test of chemical “viability” (trees having no more than four bonds emanating from a vertex) vs “nonviability” with respect to alkanes is formulated. Additionally, a criterion for ascertaining optical isomerism is proposed on the basis of the values of one of the parameters used in the formulation of the Matula number.

A quarter century ago, David W. Matula formulated a one-to-one and onto correlation between the class of rooted trees and the prime factorization of numbers.¹ He also noted² that this relationship could be applied to the ordering (and thus the naming) of the entire class of acyclic alkanes.

Although such a scheme might appear attractive to mathematical chemists, because of the peculiar needs of chemical nomenclature (see, for example, Read and Milner³), the organic chemistry community showed no interest in such a “nomenclature”. Consequently, the virtues of Matula’s system remained buried in applied mathematics literature until, 3 years ago, one of us started probing the underlying logic of Matula’s method. In doing so, it was realized that this type of procedure is applicable to a much wider class of chemical compounds. In particular, it was noted⁴ that any class of compounds that could be uniquely represented by a single parameter—in the same manner as Matula had noted the relation between the graph theoretical tree and the alkanes—would be amenable to a similar canonical ordering (and thus a canonical name). Two such classes of chemical interest are polycyclic aromatic compounds comprised of fused benzene modules (referred to by various authors as polybenzenes, polyhexes, PAH6, etc.) and face-fused modules of adamantane (referred to as polymantanes). [Also, we may expand this idea by preselecting a sequence of properties and using various ordered concatenations of individual modules. For example, use modulo 3 integers so that 1 represents a single bond, 2 a double bond, and 0 a triple bond⁴, etc.]

Since that 1990 resurrection of this technique, we have described quite a few properties about this class of numbers and formally proved the following two important theorems.⁵

Theorem 1. An integer M is the Matula number of an n -alkane (often incorrectly referred to as a “straight-chain” alkane⁶) C_nH_{2N+2} if and only if there exists an integer k between 1 and N such that $M = q_{k-1}q_{N-k}$, where the values of q_p are defined recursively in terms of subscripts of subscripts as follows: For the first term, let $p = 0$ and the subscript of q be the 0th prime (p_0); thus the value of $q(q_{p_0}) = 1$. For the next term, use the previous value of p and take the p th prime number; i.e., q_1 is the first prime, namely, 2. This process is continued by using the q value obtained as the next sequential p : q_2 is the second prime equal to 3; q_3 is the third prime equal to 5; q_4 is the fifth prime equal to 11; q_5 is the eleventh prime equal to 31; etc.

Theorem 2. For large values of k , $\ln q_k / \ln(\ln q_k) \approx k$.

As well as the algorithm used in computing the Matula number of a given rooted tree, three additional items are worth

reiterating at this time. (Note that some of these additional items were implied, rather than formally spelled out in ref 5.)

Algorithm. Setting the rooted vertex at the top of the tree, work upward from each of the leaves of the tree as follows: (a) Label each leaf of the given rooted tree by 1; (b) label each stem by the first prime number, namely, 2; (c) label the nodes from which these stems emanated by the product of the names of the stems below it if there is more than one such stem, or by the single number if there is only a single stem below that vertex; (d) label the stem above the node described in c by the corresponding prime number; i.e., if the node is numbered 4, the stem above it is the fourth prime number, namely, 7, etc.; (e) repeat steps c and d until the rooted vertex is reached. The name of that vertex is the name of the corresponding alkane.

(1) In the hydrogen-surpressed graph of an n -alkane, the degree of the respective vertices (herein referred to as the “valency”) is either 1 (for the end carbons) or 2 (for all interior carbon atoms). This property manifests itself in Theorem 1 in the existence of the integer k . [The term “valence” in graph theory corresponds to the hydrogen-surpressed graph of an alkane, rather than, as in usual chemical parlance, each carbon atom of an n -alkane has a valence equal to 4, counting the hydrogen atoms. It was for this reason in ref 7 that we chose to use the term “valency” for the graph theoretical idea and to continue to use the chemically familiar word “valence” counting the hydrogen atoms.] (2) There exist exactly $[(N + 1)/2]$ values that M can assume for a given n -alkane, where the square brackets, $[]$, represent the greatest integer function. (3) For an n -alkane, the values that q_i can assume are ONLY the numbers described in theorem 1, namely, 1, $q_1 = 2$, $q_2 = 3$, $q_3 = 5$, $q_4 = 11$, $q_5 = 31$, ... All other values of M , including denumerably many prime integers, such as 7, 13, etc., will correspond to either a branched alkane or a chemically nonviable graph.

Additionally we note that every rooted tree can be represented by a unique integer, the Matula number, and, so long as we make no distinction between the different possible sequencings of the various branches of a tree, every positive integer corresponds to a single rooted tree. Furthermore, we note that the establishment of a canonical name for alkanes in chemistry was not a consideration in Matula’s original idea;¹ consequently, no distinction was originally made as to which of the different carbon atoms (vertices) should be designated as the root. It was only when the method was resurrected in ref 4 that the obvious choice of using the lowest value of M as the canonical name was suggested. This had the added advantage that (except for very small trees) one of the “center” vertices would almost always be the root. (Table 1 is a

* Abstract published in *Advance ACS Abstracts*, February 15, 1994.

Table 1. Correlation of Matula Number with Root Designated Alkane

<i>M</i>	carbon skeleton	IUPAC name	<i>M</i>	carbon skeleton	IUPAC name
1	C*	methane	19		(1-hydro)2,2-dimethylpropane
2	C*-C	ethane	20		(2-hydro)2-methylpentane
3	C*-C-C	(1-hydro)propane	21		(3-hydro)2-methylpentane
4	C-C*-C	(2-hydro)propane	22	C-C*-C-C-C-C	(2-hydro)hexane
5	C*-C-C-C	(1-hydro)butane	23		(1'-hydro)3-methylhexane
6	C-C*-C-C	(2-hydro)butane	24		(0-hydro)2,2-dimethylbutane
7		(1-hydro)2-methylpropane	25	C-C-C-C*-C-C-C	(4-hydro)heptane
8		(2-hydro)2-methylpropane	26	C-C*-C-C-C-C	(2-hydro)3-methylpentane
9	C-C-C*-C-C	(3-hydro)pentane	27		(3-hydro)3-ethylpentane
10	C-C*-C-C-C	(2-hydro)pentane	28		(2-hydro)2,3-dimethylbutane
11	C*-C-C-C-C	(1-hydro)pentane	29		(1-hydro)2-methylpentane
12		(2-hydro)2-methylbutane	30	C-C-C*-C-C-C-C	(3-hydro)3-methylhexane
13		(1-hydro)2-methylbutane	31	C*-C-C-C-C-C-C	(1-hydro)hexane
14		(3-hydro)2-methylbutane	32		impossible—five ligands on center carbon atom
15	C-C-C*-C-C-C	(4-hydro)2-methylpentane	33	C-C-C*-C-C-C-C-C	(3-hydro)heptane
16		(3-hydro)hexane	34		(3-hydro)2-methylhexane
17		(0-hydro)2,2-dimethylpropane	35		(0-hydro)3,3-dimethylpentane
18		(4-hydro)2-methylbutane	36		(3-hydro)3-methylpentane

sequential listing of the first 36 values of *M* and the alkane that such a number would correspond to. Also, included is a minor modification of the IUPAC name—in which the root has been singled out.) In retrospect, we see that each of the different Matula names that a specific alkane has (depending on which atom is selected as the root) corresponds to the name of that particular alkyl group as a substituent. Note in Table 1, we have used as the "substituent" the hydrogen atom that is already there. In other words, in the IUPAC name, the part in parentheses would be dropped for the unsubstituted alkane.

Later in this report we shall note that, by examining only the Matula number—without actually constructing the entire tree—we can ascertain whether the represented alkane is (A) chemically NOT viable, (B) chemically viable and existing as a pair of enantiomers, or (C) chemically viable and having

no optical isomerism. Meanwhile, we shall present a straightforward method that can be easily implemented on a computer, for determining the maximum vertex degree $D(T)$ of the rooted tree *T* associated with Matula number *M*, without actually constructing tree *T*.

Consider the integer, *M*, decomposed into its prime factorization:

$$M = \prod_{i=1}^d p_i \quad (1)$$

where *d* is the number of prime factors of *M*. Define a function

$$F(M) = \max \{d, F(2t_i) | i = 1, 2, \dots, d\} \quad (2)$$

with initial condition $F(2) = 1$. We now wish to show that $F(M)$ is the maximum vertex degree of the tree corresponding

Table 2. Determination of $F(M)$

M	d	primes	t_i	$F(2t_i)$	$F(M)$	N
1	0					1
2	1	2	1	$F(2) = 1$	1	2
3	1	3	2	$F(4) = 1$	2	3
4	2	2, 2	1	$F(2) = 1$	2	3
5	1	5	3	$F(6) = 1$	2	4
6	2	2, 3	1, 2	$F(2) = 1, F(3) = 2$	2	4
7	1	7	4	$F(8) = 1$	3	4
8	3	2, 2, 2	1, 1, 1	$F(2) = 1$	3	4
9	2	3, 3	2, 2	$F(4) = 2$	2	5
10	2	2, 5	1, 3	$F(2) = 1, F(6) = 2$	2	5
11	1	11	5	$F(10) = 2$	2	5
12	3	2, 2, 3	1, 1, 2	$F(2) = 1, F(4) = 2$	3	5
13	1	13	6	$F(12) = 3$	3	5
14	2	2, 7	1, 4	$F(2) = 1, F(8) = 3$	3	5
15	2	3, 5	2, 3	$F(4) = 2, F(6) = 2$	2	6
16	4	2, 2, 2, 2	1, 1, 1, 1	$F(2) = 1$	4	5
17	1	17	7	$F(14) = 3$	3	5
18	3	2, 3, 3	1, 2, 2	$F(2) = 1, F(4) = 2$	3	6
19	1	19	8	$F(16) = 4$	4	5
20	3	2, 2, 5	1, 1, 3	$F(2) = 1, F(6) = 2$	3	6
21	2	3, 7	2, 4	$F(4) = 2, F(8) = 3$	3	6
22	2	2, 11	1, 5	$F(2) = 1, F(10) = 2$	2	6
23	1	23	9	$F(18) = 3$	3	6
24	4	2, 2, 2, 3	1, 1, 1, 2	$F(2) = 1, F(4) = 2$	4	6
25	2	5, 5	3, 3	$F(6) = 2$	2	7
26	2	2, 13	1, 6	$F(2) = 1, F(12) = 3$	3	6
27	3	3, 3, 3	2, 2, 2	$F(4) = 2$	3	7
28	3	2, 2, 7	1, 1, 4	$F(2) = 1, F(8) = 3$	3	6
29	1	29	10	$F(20) = 3$	3	6
30	3	2, 3, 5	1, 2, 3	$F(2) = 1, F(4) = F(6) = 2$	3	7
31	1	31	11	$F(22) = 2$	2	6
32	5	2, 2, 2, 2	1, 1, 1, 1, 1	$F(2) = 1$	5	6
33	2	3, 11	2, 5	$F(4) = 2, F(10) = 2$	2	7
34	2	2, 17	1, 7	$F(2) = 1, F(14) = 3$	3	6
35	2	5, 7	3, 4	$F(6) = 2, F(8) = 3$	3	7
36	4	2, 2, 3, 3	1, 1, 2, 2	$F(2) = 1, F(4) = 2$	4	7

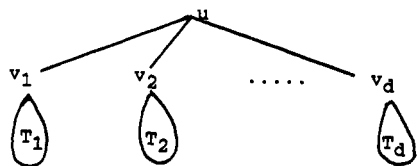


Figure 1.

to Matula number M . If this is true, then when $F(M) \leq 4$, the alkane represented is viable, while when $F(M) > 4$, it is not viable.

As motivation for this proposition, in Table 2, we present a tabulation of the values of d and $F(2t)$ for all values of M from 1 to 36. We are now ready to prove theorem 3.

Theorem 3. $F(M)$ is the maximum vertex degree of the tree corresponding to the Matula number M .

Consider a rooted tree T whose root is u . Designate the degree of vertex u by d , where $d \geq 1$. We may now express T in terms of subtrees emanating from each of the "level 2" vertices (Figure 1). Observe that every subtree T_i has at least one vertex, namely, v_i . Now, the Matula number of the tree rooted at vertex u is defined as

$$M(T, u) = \prod_{i=1}^d p_{M(T_i, v_i)} \quad (3)$$

where p_i stands for the i th prime number and where $M(T_0) = 1$ if T_0 consists of a single vertex.

Formula 3 shows that the Matula number of the rooted tree T depends on the Matula numbers of its branches T_1, T_2, \dots, T_d (provided these are considered as trees rooted at the respective vertices v_1, v_2, \dots, v_d). On the other hand, the maximum vertex degree of T is either equal to the degree of

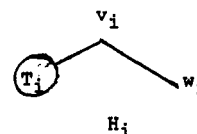


Figure 2.

u or to the degree of some vertex in one (or more, if equal) of the branches. Consequently, the maximum vertex degree of T can be obtained by separately examining the root u and the branches T_1, T_2, \dots, T_d .

Now, however, in lieu of examining the branches T_i , consider the auxiliary graph H_i obtained by adding a new vertex, w_i , to the vertex v_i of T_i (Figure 2). Note that T_i (considered as a branch of T) and H_i (considered as a separate graph) have equal maximum vertex degrees. For H_i rooted at w_i , we can now write formula 3 as

$$M(T, u) = \prod_{i=1}^d M(H_i, w_i) \quad (4)$$

Next, designating the maximum vertex degree of T as D and letting M be the Matula number corresponding to T , because of the 1:1 correspondence between M and T , we may consider D as a function of M . We may now determine the maximum vertex degree of T by separately examining the two possible cases: (1) The vertex with maximum degree is the root of T —in which case $D = d$; OR (2) the vertex with maximum degree is in one of the branches. For this second case, we note that M had been expressed by formula 1. Now from formula 4, the maximum vertex degree in $H_i = D(p_i)$. Consequently,

$$D(M) = \max \{d, D(p_i) | i = 1, 2, \dots, d\} \quad (5)$$

Now because the value of the maximum vertex degree of H_i does not change if we choose vertex v_i as the root, the Matula number of H_i becomes equal to $2t_i$. Consequently, $D(p_i) = D(2t_i)$ and eq 5 becomes

$$D(M) = \max \{d, D(2t_i) | i = 1, 2, \dots, d\} \quad (6)$$

The final step of the proof is to note that $M = 2$ is the unique two-vertex tree (chemically this corresponds to ethane), for which $D = 1$. Thus the function $D(M)$ is precisely the same as the previously defined function $F(M)$.

Having just proved that $F(M)$ corresponds to the maximum degree vertex, we can now give a geometric interpretation to what the various values of $F(M)$ imply. First of all we note that, in accordance with theorem 1, when $F(M) = 2$, the alkane is an n -alkane. This is readily seen by comparing Tables 1 and 2. In both tables, we see that the set of n -alkanes through $M = 36$ is

$$M = \{3, 4, 5, 6, 9, 10, 11, 15, 22, 25, 31, 33\}$$

Next, using Table 2, let us examine the set of alkanes having $F(M) = 3$:

$$M = \{7, 8, 12, 13, 14, 17, 18, 20, 21, 23, 26, 27, 28, 29, 30, 34, 35\}$$

Here, we see that this set can be segregated into two disjoint sets. M' was formed using two or more common primes, while M'' has three distinct primes:

$$M' = \{7, 8, 12, 13, 14, 17, 18, 20, 21, 23, 26, 27, 28, 29, 34, 35\}$$

$$M'' = \{30\}$$

Examining the members of M' in Table 1, we note that every member of this set lies in the plane and would have a line of symmetry running through the middle of the molecule if we had foregone the printing convenience of using a strictly rectangular format. On the other hand, the only member of M'' below 36 ($M = 30$) has a chiral center at the root and thus gives rise to optical isomerism.

Turning now to the set of alkanes having $F(M) = 4$

$$M = \{16, 19, 24, 36\}$$

since every one of these numbers either has a prime factorization which contains a repeated prime or is the r th prime number, where r has a repeated prime in its factorization, we would expect and shall find that each has a line of symmetry and thus does not give rise to optical isomerism; i.e., they would be categorized as members of set M' . The smallest value of M having $F(M) = 4$ that would fall in M'' and thus have four different non-hydrogen ligands is $M = 2 \times 3 \times 5 \times 7 = 210$ (which corresponds to 3-ethyl-2,3-dimethylhexane).

Finally, let us turn to integers that produce values of $F(M) > 4$. These would correspond to a central atom having more than four ligands, which is chemically unattainable. This we see for the smallest such number, $M = 32$, which was marked accordingly in Table 1 as "impossible".

Recapitulating, we see that if M can be decomposed into four or less primes, it MAY exist chemically; however, if it requires five or more prime factors, then this is a sufficient condition that such an "alkane" is not viable. Similarly, if M has three or four distinct prime factors, then this is a sufficient condition for the existence of enantiomers.

Furthermore, in order to ascertain an important class of nonviable values of M , we note that $M = 53$ is the 16th prime. Now since we used $F(2t_i)$ in determining $F(M)$ and since $F(32) = 5$, this is a nonviable value of M which is the same alkane as $M = 32$, but with the root on one of the leaf carbons, rather than the central one. We can generalize this to include not only the product of five primes, but also the a th prime, where a is the product of only four primes; i.e., we have nonviability for the 16th prime (53), the 24th prime (89), the 36th prime (151), etc., as well as for any multiple of one of these numbers, or if any of these numbers is the p value of the p th prime, e.g., the 53rd prime (241), the 89th prime (461), etc.

As a final note, in order to establish necessary conditions for enantiomers, let us examine the different roots of a given tree. In particular, let us focus on the smallest Matula number

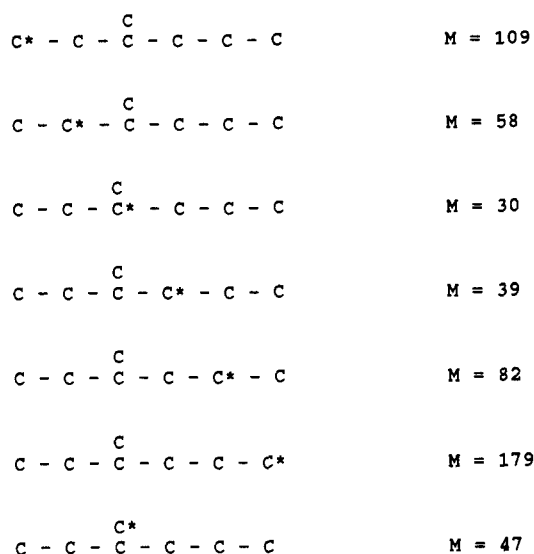


Figure 3.

that gave rise to enantiomers ($M = 30$) and its associated tree. Note that $M = 30$ is just one of the seven rooted trees that is associated with this alkane (3-methylhexane). The other six are shown in Figure 3, along with their respective Matula numbers. Note that, in conformity, with the observation that the central most atom being the root gives rise to the minimum Matula number, $M = 30$ is the only value that was evidently optically active; all the rest required more than just a single prime factorization in order to identify the four distinct ligands on the central atom.

Consequently, realistically, we believe that necessary conditions are not going to be forthcoming unless some means of rapidly correlating Matula numbers of the same tree having different roots can be discovered. At present, such a protocol is not in evidence.

REFERENCES AND NOTES

- (1) Matula, D. W. *SIAM Rev.* **1968**, *10*, 273.
- (2) Personal conversations with one of us (S.B.E.).
- (3) Read, R. C.; Milner, R. S. Report to University of West Indies, 1968. Revised as: Research Report CORR-78-42; Department of Combinatorics and Optimization, University of Waterloo: Ontario, Canada, 1978.
- (4) Elk, S. B. *J. Math. Chem.* **1990**, *4*, 55-68.
- (5) Gutman, I.; Ivic, A.; Elk, S. B. *J. Serb. Chem. Soc.*, **1993**, *58*, 193-201.
- (6) Elk, S. B. *MATCH* **1988**, *23*, 19-33.
- (7) Elk, S. B. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 11-16.