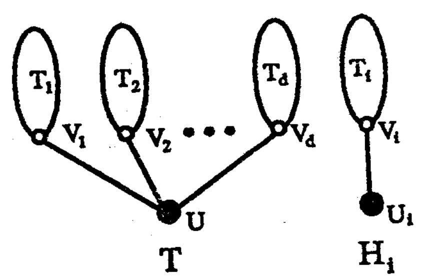
PUBLICATIONS DE L'INSTITUT MATHÉMATIQUE
Nouvelle série tome 53 (67), 1993, 17–22

DEDUCING PROPERTIES OF TREES FROM THEIR MATULA NUMBERS

Ivan Gutman and **Yeong-Nan Yeh**

Abstract. The Matula number is a natural number that uniquely characterizes a rooted tree. It is shown how a wariety of properties of a tree can be deduced directly from the Matula number, circumventing the reconstruction of the tree itself.

Introduction. In this paper we are concerned with trees and rooted trees, having finite numbers of vertices and being without loops and multiple edges. Let T be a tree rooted at its vertex u. Let the vertices adjacent to u be denoted by $v_1, \dots, v_d$, $d \geq 1$. Then T has the structure shown in the figure below. The subtrees $T_1, \dots, T_d$ are called the branches of T.



The vertices of a tree that differ from the root will be called non-root vertices.

Trees are bipartite graphs and their vertices can be colored by two colors (say, black and white), so that adjacent vertices are always differently colored. In what follows the color of the root vertex will always be chosen black.

In 1968 David Matula [**5**] pointed out the existence of a particular bijection between the set of natural numbers and the set of rooted trees. According to

AMS Subject Classification (1985): Primary 05 C 05

Matula, a number $n = n(T) = n(T, u)$ is associated to the tree T rooted at vertex u in the following recursive manner (cf.Figure):

$$n(T, u) = \prod_{i=1}^{d} P_{n(T_i, v_i)} \tag{1}$$

where p_t stands for the t-th prime number $(p_1 = 2, p_2 = 3, p_3 = 5, \dots)$ and where $n(T_0) = 1$ if T_0 consists of a single vertex. The number $n(T)$ is the Matula number of the rooted tree T.

Matula's result was for a long time considered just as a graph-theoretical curiosity. Only relatively recently it has been recognized that the Matula numbers offer a unique opportunity in computer-aided chemical information systems, because they enable the storage of the complete information on the structure of certain organic compounds by means of a single integer [**3**,**4**]. This potential practical application fostered further studies of the properties of the Matula numbers [**1**,**2**].

As already mentioned, a Matula number uniquely determines a rooted tree and vice versa [**5**]. Consequently, every property of a rooted tree can - in principle - be determined from its Matula number. The "brute-force" way of doing this would be to first reconstruct T from $n(T)$ and then establish the property $P(T)$ by examing T. Bearing in mind the possible applications of the Matula numbers in information science, it would be of particular interest to find procedures by which a property $P(T)$ could be deduced directly from $n(T)$, skipping the actual reconstruction of T. Until now such a procedure was known in only one case, namely when $P(T)$ is the number of vertices of T [**2**]. In this paper we establish results enabling for a number of other properties of (rooted) trees to be deduced directly from the respective Matula numbers.

Statement of results. In what follows we will assume that n has the following decomposition into prime numbers:

$$n = \prod_{i=1}^{d} p_{t_i}, \quad d \geq 1 \tag{2}$$

where, of course, some of the primes $p_{t_1}, \dots, p_{t_d}$ may be mutually equal. In eq. (1) it is assumed that $n > 1$. We now define

(a) $P_1(T)$ = number of vertices of T;

(b) $P_2^B(T), P_2^W(T)$ = number of black and white vertices of T, respectively; the root of T is colored black;

(c) $P_3(T)$ = width of T, i.e., the number of non-root vertices having degree one;

(d) $P_4^k(T)$ = number of vertices of T having degree $k, k \geq 1$;

(e) $P_5(T)$ = maximum degree of a vertex of T;

(f) $P_6(T)$ = minimum degree of a vertex of T, greater than unity;

(g) $P_7(T)$ = eccentricity of the root i.e. the greatest distance between the root and a vertex of T;

(h) $P_8(T)$ = smallest distance between the root of T and a vertex of degree one;

(i) $P_9(T)$ = diameter of T, i.e. the greatest distance between two vertices;

(j) $P_{10}^k(T)$ = number of vertices of T that are on distance k from the root, $k \leq 1$.

With the above notation our results may be formulated as the following:

THEOREM. *Let T be a rooted tree and $n, n > 1$, its Matula number whose decomposition into primes is given by eq.* (2). *Then the following identities are satisfied, relating the above properties with the Matula number:*

(a) $P_1(T) = f_1(n)+1$, *where* $f_1(n) = d + \sum_{i=1}^{d} f_1(t_i)$ *and the function* f_1 *satisfies the initial condition* $f_1(1) = 0$.

(b) $P_2^B(T) = f_2^B(n) + 1$ *and* $P_2^W(T) = f_2^W(n)$, *where*
$f_2^B(n) = \sum_{i=1}^{d} f_2^W(t_i)$; $\quad f_2^W(n) = d + \sum_{i=1}^{d} f_2^B(t_i)$
and where the initial conditions are $f_2^B(1) = f_2^W(1) = 0$.

(c) $P_3(T) = f_3(n)$, *where* $f_3(n) = \sum_{i=1}^{d} f_3(t_i)$ *and* $f_3(1) = 1$.

(d) $P_4^k(T) = f_4(n,k) + \delta_{d,k}$, *where* $f_4(n,k) = \sum_{i=1}^{d} \big(f_4(t_i,k) + \delta_{\Omega(t_i),k-1}\big)$ *and* $f_4(1,k) = \delta_{1,k}$. *By* $\Omega(t_i)$ *we denote the number of prime factors (not necessarily distinct) of* t_i. *The Kronecker* δ *symbol has its usual meaning:* $\delta_{u,v} = 1$ *if* $u = v$ *and* $\delta_{u,v} = 0$ *if* $u \neq v$.

(e) $P_5(T) = f_5(n)$, *where* $f_5(n) = \max\{d, f_5(2t_i) \,|\, i = 1, \ldots, d\}$ *and* $f_5(2) = 1$.

(f) $P_6(T) = f_6(n)$, *where*

$$f_6(n) = \begin{cases} \min\{d, f_6(2t_i) \,|\, i = 1, \ldots, d\}, & \textit{if } d \geq 2 \\ f_6(2t_1), & \textit{if } d = 1. \end{cases}$$

The function $f_6(n)$ *is undefined for* $n \leq 2$, *but it is consistent to formally set* $f_6(1) = f_6(2) = \infty$. *Then we have* $f_6(3) = 2$, $f_6(4) = 2$, *etc.*

(g) $P_7(T) = f_7(n)$, *where* $f_7(n) = 1 + \max\{f_7(t_i) \,|\, i = 1, \ldots, d\}$ *and* $f_7(1) = 0$.

(h) $P_8(T) = f_8(n)$, *where* $f_8(n) = 1 + \min\{f_8(t_i) \,|\, i = 1, \ldots, d\}$ *and* $f_8(1) = 0$.

(i) $P_9(T) = f_9(n)$, *where*

$$f_9(n) = \begin{cases} \max\{f_9(2t_i), g_{ij} \,|\, i = 1, \ldots, d,\ j = 1, \ldots, d,\ i \neq j\}, & \textit{if } d \geq 2 \\ f_9(2t_1), & \textit{if } d = 1. \end{cases}$$

The auxiliary quantities g_{ij} *are defined via the previously introduced function* f_7 *as:* $g_{ij} = f_7(t_i) + f_7(t_j) + 2$. *The initial condition is* $f_9(2) = 1$.

(j) $P_{10}^k(T) = f_{10}(n,k)$, *where* $f_{10}(n,k) = \sum_{i=1}^{d} f_{10}(t_i, k-1)$ *with initial conditions*

$$f_{10}(n,0) = 1, \qquad \textit{for } n \geq 1$$
$$f_{10}(1,k) = 0, \qquad \textit{for } k \geq 1.$$

Part (a) of the above results was previously reported [**2**]. We nevertheless include it into the theorem for the sake of completeness.

Proof of the theorem. The idea of the proofs of all the statements (a)-(j) is based on the fact that certain properties of a tree T can be deduced by examining one-by-one its branches $T_1, \ldots, T_d$ (see Figure). On the other hand, the Matula number $n(T, u)$ is also defined via the Matula numbers of $T_1, \ldots, T_d$ (cf. eq. (1)).

If the root is a vertex of degree one ($d = 1$ in eq. (1)), then the Matula number obeys $n(T, u) = p_{n(T_1, v_1)}$. Bearing this in mind we can rewrite eq. (1) as

$$n(T, u) = \prod_{i=1}^{d} n(H_i, u_i) \tag{3}$$

where the tree H_i is obtained by attaching a new vertex u_i to the vertex v_i of T_i (see Figure). If not stated otherwise, it is understood that u_i is the root of H_i and v_i the root of T_i.

The function f_1 counts, in fact, the non-root vertices of T. Their number is evidently equal to the sum of the vertex counts of all branches or, what is the same, to the number of non-root vertices of H_i, $i = 1, \ldots, d$. Therefore

$$f_1\big(n(T, u)\big) = \sum_{i=1}^{d} f_1\big(n(H_i, u_i)\big). \tag{4}$$

In view of eqs. (2) and (3) we may set $n(H_i, u_i) = p_{t_i}$. Then, however, $n(T_i, v_i) = t_i$.

The number of non-root vertices of H_i is by one greater than the number of non-root vertices of T_i (see Figure). This implies

$$f_1(p_{t_i}) = f_1(t_i) + 1. \tag{5}$$

Part (a) of the theorem follows now by combining (4) and (5). $\Box$

The proof of part (b) of the theorem is completely analogous: The function f_2^B and f_2^W count the black and white non-root vertices, respectively. We only have to observe that in the transformation $H_i \to T_i$ the coloring of the vertices has to be reversed. Evidently, $f_2^B + f_2^W = f_1$. $\Box$

The proof of the part (c) goes also along the same lines and is even simpler because here (by definition of the width) the root does not cause any difficulty. $\Box$

When counting the vertices of degree k we must distinguish between two cases. The root of T is either of degree k (then $k = d$) or not. The function f_4 counts the non-root vertices of degree k, and the term $\delta_{d,k}$ is the adjustment for the case when the root itself is of degree k. The trees H_i and T_i have the same numbers of non-root vertices of degree k, except if the vertex v_i in H_i is of degree k. If this occurs then v_i in T_i will be of degree $k - 1$ and, consequently, t_i (= the Matula number of T_i) will have exactly $k - 1$ prime factors. This correction is accounted for by the term $\delta_{\Omega(t_i), k-1}$.

The remaining reasoning leading to (d) is the same as in the previous three proofs. $\Box$

To determine the maximum vertex degree in T we again have to consider two cases. Either is this the degree of the root ($= d$) or the degree of some non-root vertex. Because the maximum vertex degree in H_i is $f_5(p_{t_i})$ we have

$$f_5(n) = \max\{d, f_5(p_{t_i}) \,|\, i = 1, \ldots, d\}. \tag{6}$$

Now, the value of the maximum vertex degree of H_i will not change if we chose the vertex v_i as the root (cf. Figure). If so, then the Matula number of H_i becomes equal to $2t_i$. Consequently, $f_5(p_{t_i}) = f_5(2t_i)$ and part (e) of the theorem follows. □

The smallest vertex degree in T is, trivially, equal to unity. The second smallest vertex degree can be obtained from the Matula number by means of a consideration that parallels the proof of (e). One only has to eliminate the term d from eq. (6) in the case when $d = 1$. □

It is evident that the vertex being at maximal distance from the root of T is of degree one. Suppose that this vertex is in the branch T_i and consider T_i as a tree rooted at the vertex v_i. Then the transformation $T \to T_i$ diminishes by one the eccentricity of the root vertex. This, together with the fact that $n(T_i, v_i) = t_i$, implies the part (g) of the theorem. □

In order to arrive at the result stated as part (h) of the theorem observe that the diameter of T is either the maximum distance between two vertices belonging to the same branch of T (the root inclusive), or between vertices belonging to different branches. In the former case the diameter is given by $\max\{f_9(p_{t_i}) \,|\, i = 1, \ldots, d\}$ because $f_9(p_{t_i})$ is the diameter of H_i. In the latter case the diameter is

$$\max\{g_{ij} \,|\, i = 1, \ldots, d,\ j = 1, \ldots, d,\ i \neq j\}$$

because g_{ij} is just the sum of the eccentricities of the roots of H_i and H_j. (This, of course, can occur only if $d \geq 2$.) The greatest of the above two maxima is $f_9(n)$.

As before, the fact that the change of the root of H_i (from u_i to v_i) will not affect the value of the diameter of H_i, infers the identity $f_9(p_{t_i}) = f_9(2t_i)$. This, combined with the above maxima leads to (h). □

To count the vertices that are on a distance k from the root of T it is sufficient to observe that their number is the sum of the numbers of vertices at distance $k - 1$ from the roots of the subtrees $T_1, \ldots, T_d$. Part (j) of the theorem follows immediately. □

By this we completed the proof of the theorem. □

Discussion. In certain cases the results stated in our Theorem can be somewhat simplified. This, in particular, happens when the Matula number is itself a prime, i.e. when $d = 1$:

COROLLARY 1. *If $n = p_t$ then the functions defined in the theorem conform* *to the following relations:*

(a) $f_1(n) = f_1(t) + 1$ (b) $f_2^B(n) = f_2^W(t)$; $f_2^W(n) = f_2^B(t) + 1$
(c) $f_3(n) = f_3(t)$ (d) $f_4(n,k) = f_4(t,k) + \delta_{\Omega(t),k-1}$
(e) $f_5(n) = f_5(2t)$ (f) $f_6(n) = f_6(2t)$
(g) $f_7(n) = f_7(t) + 1$ (h) $f_8(n) = f_8(t) + 1$
(i) $f_9(n) = f_9(2t)$ (j) $f_{10}(n,k) = f_{10}(t,k-1)$.

The calculation of the functions defined in the theorem 1 is also facilitated by the identities collected in the following two corollaries.

COROLLARY 2. *Let α and β be two positive integers (not necessarily primes). Then*

$$\begin{aligned} f_1(\alpha\beta) &= f_1(\alpha) + f_1(\beta) \\ f_2^B(\alpha\beta) &= f_2^B(\alpha) + f_2^B(\beta);\;\; f_2^W(\alpha\beta) = f_2^W(\alpha) + f_2^W(\beta) \\ f_4(\alpha\beta,k) &= f_4(\alpha,k) + f_4(\beta,k), \qquad k \geq 2 \\ f_7(\alpha\beta) &= \max\{f_7(\alpha), f_7(\beta)\} \\ f_{10}(\alpha\beta,k) &= f_{10}(\alpha,k) + f_{10}(\beta,k), \qquad k \geq 1. \end{aligned}$$

COROLLARY 3. *Let α and β be two integers greater than one (not necessarily primes). Then*

$$\begin{aligned} f_3(\alpha\beta) &= f_3(\alpha) + f_3(\beta) \\ f_4(\alpha\beta,k) &= f_4(\alpha,k) + f_4(\beta,k), \qquad k = 1 \\ f_8(\alpha\beta) &= \min\{f_8(\alpha), f_8(\beta)\}. \end{aligned}$$

Acknowledgement. One author (Yeh) in indebted to the National Science Council of the Republic of China for financial support under the grant NSC-82-0208-M001-042.

REFERENCES

[1] I. Gutman, A. Ivić, *Graphs with maximal and minimal Matula numbers*, Bull. Acad. Serbe Sci. Arts, to be published.

[2] I. Gutman, A. Ivic, S.B.Elk, *Matula numbers for coding chemical structures and some of their properties*, J. Serb. Chem. Soc. **58** (1983), 193-201.

[3] S.B. Elk, *A problem with the application of Matula's method of prime numbers and rooted trees for canonical numbering of alkanes*, Graph Theory Notes New York **18** (1989), 40-43.

[4] S.B. Elk, *A canonical ordering of polybenzenes and polyadamantanes using a prime number factorization technique*, J. Math. Chem. **4** (1990), 55-68.

[5] D.W. Matula, *A natural rooted tree enumeration by prime factorization*, SIAM Rev. **10** (1968), 273-273.

Faculty of Science
University of Kragujevac
P.O.Box 60
YU-34000 Kragujevac
Yugoslavia,

Institute of Mathematics
Academia Sinica
Taipei 11529
Taiwan, ROC

(Received 24 12 1992)
(Revised 23 03 1993)