THE TOPOLOGICAL TREES WITH EXTREMAL MATULA NUMBERS

AUDACE A. V. DOSSOU-OLORY

ABSTRACT. Let T be a rooted tree with branches $T_1, T_2, \ldots, T_r$ and p_m the m-th prime number ($p_1 = 2, p_2 = 3, p_3 = 5, \ldots$). The Matula number $M(T)$ of T is $p_{M(T_1)} \cdot p_{M(T_2)} \cdot \ldots \cdot p_{M(T_r)}$, starting with $M(\bullet) = 1$. It was put forward half a century ago by the American mathematician David Matula. In this paper, we prove that the star (consisting of a root and leaves attached to it) and the binary caterpillar (a binary tree whose internal vertices form a path starting at the root) have the smallest and greatest Matula number, respectively, over all topological trees (rooted trees without vertices of outdegree 1) with a prescribed number of leaves – the extremal values are also derived.

1. INTRODUCTION

Fifty years ago, the American mathematician David Matula gave an explicit bijection between the set of all rooted trees and the set of all positive integers [1]. The bijection is described by means of prime numbers. Throughout, p_m always means the m-th prime number (in ascending order); for example,

$$p_1 = 2, \ p_2 = 3, \ p_3 = 5, \ p_4 = 7, \ p_5 = 11, \ p_6 = 13, \ldots .$$

The Matula number of the tree that has only one vertex is defined to be 1, and if T is a rooted tree with branches (the connected components that remain after deleting the root and all edges incident to it) $T_1, T_2, \ldots, T_r$, then the Matula number of T – henceforth denoted by $M(T)$ – is given recursively by

$$M(T) = p_{M(T_1)} \cdot p_{M(T_2)} \cdot \ldots \cdot p_{M(T_r)} .$$

For example, consider the rooted tree T shown in Figure 1; T has three branches T_1, T_2, T_3. We have

$$M(T_1) = p_{M(\bullet)} = p_1 = 2, \ M(T_2) = p_{M(\bullet)} \cdot p_{M(\bullet)} = p_1^2 = 4, \ M(T_3) = M(\bullet) = 1 ,$$

which implies that

$$M(T) = p_{M(T_1)} \cdot p_{M(T_2)} \cdot p_{M(T_3)} = p_2 \cdot p_4 \cdot p_1 = 42 .$$
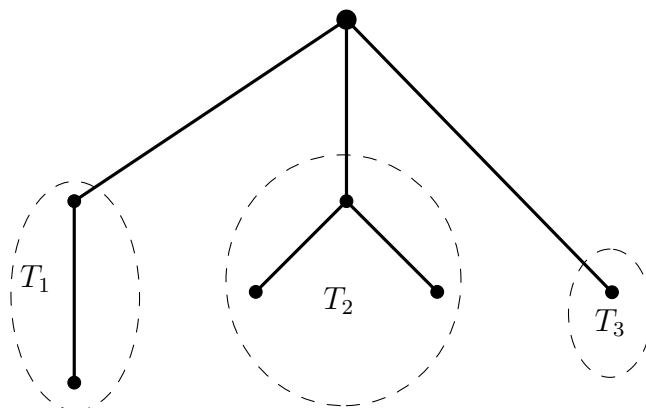
FIGURE 1. A rooted tree T with three branches T_1, T_2, T_3.

The definition of the Matula numbers also suggests that for every positive integer n, there is a unique rooted tree T whose Matula number is n. This observation follows by induction on n.

Indeed, if $n > 1$, then write n in its (unique) prime decomposition, say $n = t_1 \cdot t_2 \cdot \ldots \cdot t_l$, where $t_1, t_2, \ldots, t_l$ are all primes (not necessarily distinct!). Let m_i be the unique positive integer such that $t_i = p_{m_i}$; form a rooted tree T by joining the roots of the trees with Matula numbers $m_1, m_2, \ldots, m_l$ to a new common vertex (the root of T). The Matula number of this tree is n. Hence, there is a bijection between the set of all rooted trees and the set of all positive integers.

As mentioned by Ivan Gutman and Aleksandar Ivić in their 1994 paper [2], at the time when the bijection was found by Matula, the mathematics community had considered this discovery as 'just' a curiosity driven observation. One had to wait until two decades later when the American chemist Seymour Elk proved in subsequent papers [3, 4, 6, 7] that Matula numbers can be useful for canonical nomenclatures of alkanes and all potential combinations of polybenzene or polymantane modules. In 1992, Gutman, Ivić and Elk [5] obtained further results on how the molecular structures of certain organic compounds can be encoded by means of Matula numbers.

Since the Matula number also appears to be a natural (rooted) tree invariant, Gutman and Yeh listed in their 1993 paper [8] ten different parameters (such as the number of vertices, number of leaves, minimum vertex degree, maximum vertex degree, diameter, etc.) of a rooted tree that can be obtained directly from the associated Matula number. Three years later, Gutman and Ivić [9] proved that for $n \geq 5$, the rooted tree obtained by taking a root path (rooted at one of its endvertices) on $n - 3$ vertices and attaching three leaves to the other endvertex of the path, is the one that has the maximal Matula number over the set of all rooted trees with n vertices. For the minimum, they showed that for $n \geq 3$ and depending on the residue class of n modulo 3, the rooted tree depicted in Figure 2 is minimal among all n-vertex rooted trees.
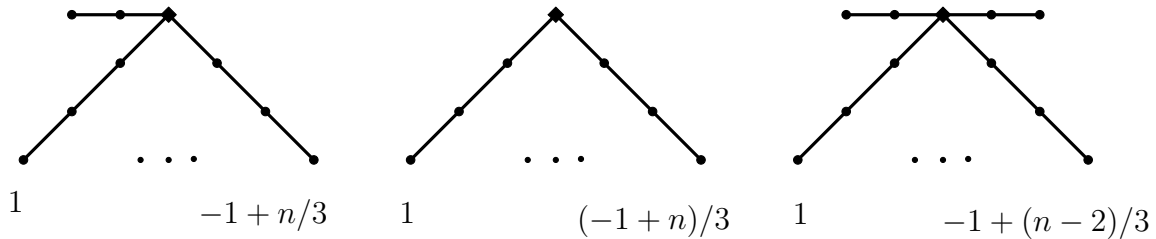
FIGURE 2. The minimal trees among n-vertex rooted trees according to the residue class of n modulo 3 (the root is the square vertex on top) [9].

In 2012, Deutsch [10] showed how to determine further properties (mostly distance-based) of a rooted tree directly from the corresponding Matula number. As such, we have the path length, number of subtrees, Wiener index, terminal Wiener index, Wiener polynomial, first Zagreb index, second Zagreb index, just to mention a few.

In this short note, we are concerned with the Matula extremal trees (and thus the Matula extremal numbers) among so-called topological trees, given the number of leaves.

In [12], a topological tree is defined as a tree without vertices of degree 2. In this note (see also [11]), a rooted tree will be called a *topological* tree if it does not have a vertex of outdegree 1. The purpose of avoiding vertices of outdegree 1 in the tree is rather natural: we are collecting trees according to the number of leaves, and there are infinitely many trees with the same number of leaves (e.g. all paths; more generally, one can subdivide any set of edges in a given tree).

We mention that trees without vertices of degree 2 (also known as *series-reduced* or *homeomorphically irreducible* trees) have already been object of study in the past by mathematicians and theoretical biologists; see e.g. Jamison [17], Bergeron et al. [12], Allman and Rhodes [13], or the sequence A000669 in [16].

2. GETTING TO THE EXTREMAL TREES

- A *binary* tree is a rooted tree in which every vertex has outdegree exactly 0 or 2;
- A *star* is a rooted tree in which all leaves are adjacent to the root (see Figure 3);
- A binary *caterpillar* is a binary tree whose non-leaf vertices lie on a single path starting at the root (see Figure 3).

We denote the star with n leaves by C_n and the n-leaf binary caterpillar by F_n. We shall prove that the star C_n minimises the Matula number and the binary caterpillar F_n maximises the Matula number over all topological trees with n leaves. We also derive the extremal values (i.e., $M(C_n)$ and $M(F_n)$) for every n.

Note the following observation, which is analogous to a transformation used by the authors of [9] in their context of trees with a given number of vertices. Let us be given two arbitrary topological trees T_1 and T_2. Denote by $S_{1,2}$ the topological tree with the two
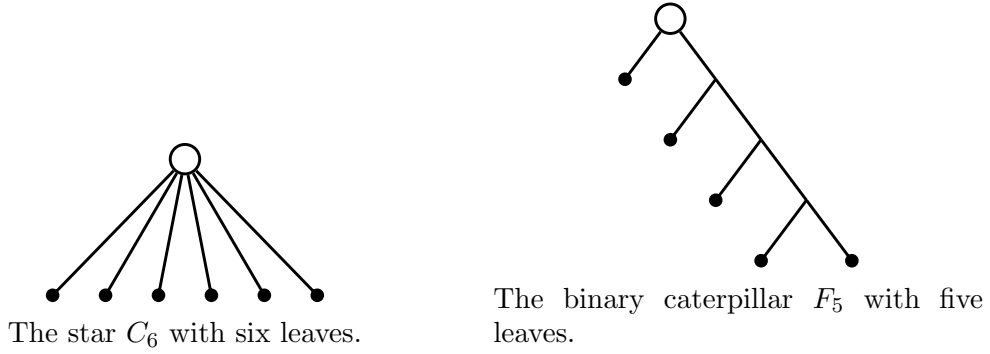
FIGURE 3. A star and a binary caterpillar.

branches T_1 and T_2. Then we have

$$p_{M(T_1)} \cdot p_{M(T_2)} < p_{M(S_{1,2})} \,.$$

To see this, simply note that $M(S_{1,2}) = p_{M(T_1)} \cdot p_{M(T_2)}$ by definition, and so

$$p_{M(S_{1,2})} = p_{p_{M(T_1)} \cdot p_{M(T_2)}} > p_{M(T_1)} \cdot p_{M(T_2)}$$

as p_m (the m-th prime number) is greater than m for all m.

Let T be a topological tree with branches $T_1, T_2, \ldots, T_r$ such that $r \geq 3$. Again, denote by $S_{1,2}$ the topological tree with the two branches T_1 and T_2. Further, denote by $F(T)$ the tree whose branches are $S_{1,2}, T_3, \ldots, T_r$; see Figure 4 for a picture. Obviously, T and $F(T)$ have the same number of leaves.
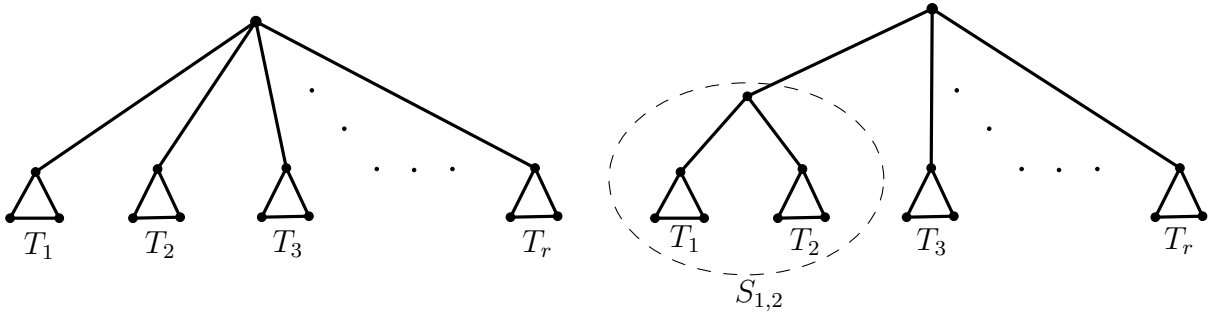
FIGURE 4. Illustration of the tree-transformation F: A topological tree T (left) and the corresponding tree $F(T)$ (right).

Then it follows immediately from the previous observation and the definition of the Matula number that

$$M(F(T)) > M(T) \,.$$

Consequently, by repeatedly applying the tree-transformation F, we then obtain from T a new topological tree, say T', with exactly two branches and having the property that

$M(T') > M(T)$. In the same way, repeated application of the same tree-transformation F to the branches of T' yields a binary tree.

Hence, the tree that has the maximal Matula number among all topological trees with a prescribed number of leaves must be a binary tree. On the other hand, it also follows from our discussion that the tree that minimises the Matula number among all topological trees with $n \geq 2$ leaves must have n branches. We have then proved the following theorem:

Theorem 1. *The star C_n (consisting of a root with n leaves attached to it) is the topological tree that has the smallest Matula number among all n-leaf topological trees. Moreover, for $n > 1$, we have $M(C_n) = 2^n$.*

3. FINDING THE MAXIMAL TOPOLOGICAL TREE

We begin with a lemma:

Lemma 2. *Let the sequence $(q_k)_{k \geq 1}$ of positive integers be defined recursively by*

$$q_1 = 1 \text{ and } q_k = 2p_{q_{k-1}} \text{ for } k > 1.$$

Then we have

$$p_{q_{k_1}} \cdot p_{q_{k_2}} \leq q_{k_1 + k_2}$$

for all pairs (k_1, k_2) of positive integers.

The first few values of the sequence $(q_k)_{k \geq 1}$ are

$$q_1 = 1, \ q_2 = 2p_1 = 4, \ q_3 = 2p_4 = 14, \ q_4 = 2p_{14} = 86, \ q_5 = 2p_{86} = 886, \ q_6 = 2p_{886} = 13766$$

as $p_{14} = 43$, $p_{86} = 443$ and $p_{886} = 6883$.

To prove the lemma, we shall need the following two theorems concerning primes:

Theorem 3 (Robin [15])**.** *We have*

$$p_m \geq m \cdot \big(\ln(m) + \ln(\ln(m)) - 1.0072629 \big)$$

for all $m \geq 2$.

Theorem 4 (Rosser-Schoenfeld [14])**.** *We have*

$$p_m \leq m \cdot \big(\ln(m) + \ln(\ln(m)) - 0.5 \big)$$

for all $m \geq 20$.

Proof of Lemma 2. We may assume $k_1 \leq k_2$. Set $k = k_1 + k_2$. For $k = 2$, we have $k_1 = k_2 = 1$ and $p_{q_{k_1}} \cdot p_{q_{k_2}} = p_{q_1}^2 = p_1^2 = 4 = 2p_1 = 2p_{q_1} = q_2 = q_{k_1 + k_2}$. For $k = 3$, we have $k_1 = 1, k_2 = 2$ and $p_{q_{k_1}} \cdot p_{q_{k_2}} = p_{q_1} \cdot p_{q_2} = p_1 \cdot p_4 = 14 = 2p_4 = 2p_{q_2} = q_3 = q_{k_1 + k_2}$. So the inequality holds for $k \in \{2, 3\}$. In fact, for $k \leq 6$, the inequality is easily verified (see Table 1). We then assume $k > 6$ and continue the proof of the inequality by induction on k.

TABLE 1. $k_1 + k_2 \in \{4, 5, 6\}$

(k_1, k_2)	(1,3)	(2,2)	(1,4)	(2,3)	(1,5)	(2,4)	(3,3)
$p_{q_{k_1}} \cdot p_{q_{k_2}}$	86	49	886	301	13766	3101	1849

Let k_1, k_2 be two arbitrary positive integers such that $k_1 \leq k_2$ and $k_1 + k_2 = k$. We may assume $k_1 \geq 2$ because for $k_1 = 1$, the lemma clearly holds with equality. We may also assume $k_2 \geq 4$ because $k > 6$ by assumption.

Now note that $p_{q_{k_1}} \geq 4$ (as $k_1 \geq 2$ implies that $q_{k_1} \geq 4$) and thus

$$\ln(p_{q_{k_1}} \cdot p_{q_{k_2-1}}) + \ln(\ln(p_{q_{k_1}} \cdot p_{q_{k_2-1}})) - 1.0072629$$
$$\geq \ln(2) + \ln(2p_{q_{k_2-1}}) + \ln(\ln(2p_{q_{k_2-1}})) - 1.0072629$$
$$\geq \ln(2p_{q_{k_2-1}}) + \ln(\ln(2p_{q_{k_2-1}})) - 0.5$$
$$= \ln(q_{k_2}) + \ln(\ln(q_{k_2})) - 0.5$$

as $\ln(2) - 1.0072629 = -0.314116\ldots$. Thus, we have

$$2p_{q_{k_2-1}}\Big(\ln(p_{q_{k_1}} \cdot p_{q_{k_2-1}}) + \ln(\ln(p_{q_{k_1}} \cdot p_{q_{k_2-1}})) - 1.0072629 \Big)$$
$$\geq q_{k_2}\Big(\ln(q_{k_2}) + \ln(\ln(q_{k_2})) - 0.5 \Big)$$
$$\geq p_{q_{k_2}},$$

where the inequality in the last step follows from Theorem 4 as $q_{k_2} \geq 20$ (since $k_2 \geq 4$).

It follows that

$$2p_{q_{k_1}} \cdot p_{q_{k_2-1}}\Big(\ln(p_{q_{k_1}} \cdot p_{q_{k_2-1}}) + \ln(\ln(p_{q_{k_1}} \cdot p_{q_{k_2-1}})) - 1.0072629 \Big) \geq p_{q_{k_1}} \cdot p_{q_{k_2}},$$

so that Theorem 3 together with the induction hypothesis gives us

$$q_k = 2p_{q_{k-1}} \geq 2q_{k-1}\Big(\ln(q_{k-1}) + \ln(\ln(q_{k-1})) - 1.0072629 \Big)$$
$$\geq 2p_{q_{k_1}} \cdot p_{q_{k_2-1}}\Big(\ln(p_{q_{k_1}} \cdot p_{q_{k_2-1}}) + \ln(\ln(p_{q_{k_1}} \cdot p_{q_{k_2-1}})) - 1.0072629 \Big)$$
$$\geq p_{q_{k_1}} \cdot p_{q_{k_2}}$$

as $q_{k-1} \geq 2$ (since $k > 6$) and $k_1 + k_2 = k$. Therefore, we obtain

$$q_k = q_{k_1+k_2} \geq p_{q_{k_1}} \cdot p_{q_{k_2}}$$

for all pairs (k_1, k_2) of positive integers. This completes the induction step and the proof of the lemma. $\square$

We can now state and prove our next theorem:

Theorem 5. *Among all topological trees with n leaves, the binary caterpillar F_n has the greatest Matula number. Furthermore, we have $M(F_n) = q_n$, where q_n is the positive integer defined recursively in Lemma 2.*

Proof. First of all, note that for $k > 1$, we have $M(F_k) = p_1 \cdot p_{M(F_{k-1})}$ and by iteration on k, we obtain $M(F_k) = q_k$ for every k.

It is shown in Section 2 that the topological tree that maximises the Matula number among all topological trees with a prescribed number of leaves must be a binary tree. Now, let us use induction on k to prove that $M(B) \leq M(F_k)$ for every binary tree B with k leaves.

The statement of the theorem is trivial for $k \leq 3$ as in this case there is only one possibility for the shape of the binary tree. Assume the statement holds for binary trees with at most $k - 1 \geq 3$ leaves and consider a binary tree B with k leaves. Denote by B_1 and B_2 the two branches of B. Since $M(B) = p_{M(B_1)} \cdot p_{M(B_2)}$, we obtain $M(B) \leq p_{M(F_{|B_1|})} \cdot p_{M(F_{|B_2|})}$ by the induction hypothesis. Consequently, invoking Lemma 2, one obtains

$$M(B) \leq p_{M(F_{|B_1|})} \cdot p_{M(F_{|B_2|})} \leq M(F_{|B|})$$

as $M(F_k) = q_k$. $\qquad\square$

4. Conclusion

The fact that the binary caterpillar F_n is maximal in the set of all n-leaf topological trees with respect to the Matula number, also shows that F_n is maximal in the set of all n-leaf binary trees with respect to the Matula number. Probably, the next natural problem for a future study will be to characterise the tree that minimises the Matula number over all binary trees (rooted trees in which every vertex has outdegree exactly 0 or 2) with a given number of leaves.

Given a positive integer $k > 1$, let s be the unique nonnegative integer satisfying $2^{s+1} \leq k < 2^{s+2}$. Then write $k = r + 2^{1+s}$ with r the residue of k modulo 2^{1+s}. Define the sequence $(l_k)_{k \geq 1}$ of positive integers recursively by

$$l_k = l_{r+2^{1+s}} = \begin{cases} p_{l_{2^s}} \cdot p_{l_{r+2^s}} & \text{if } r \leq 2^s \\ p_{l_r} \cdot p_{l_{2^{1+s}}} & \text{if } r > 2^s \,, \end{cases}$$

starting with $l_1 = 1$. The sequence begins

$$1, 4, 14, 49, 301, 1589, 9761, 51529, 452411, 3041573, 23140153, \dots .$$

Calculations suggest that $M(B) \geq l_k$ for every binary tree B with $k \leq 18$ leaves. The k-leaf minimal binary tree S_k in this case is also obtained in the same recursive way: the branches of $S_k = S_{r+2^{1+s}}$ are the binary trees S_{2^s} and S_{r+2^s} if $r \leq 2^s$, and the binary trees S_r and $S_{2^{1+s}}$ if $r > 2^s$, the starting tree being the tree that has only one vertex. For example, we show in Figure 5 the minimal binary trees S_6 and S_{13}.

Note that the numbers l_k soon get really large; for instance, the Matula number of S_{18} is 3078140605053. So it is difficult to check our evidence for more values of k; although

$$1.07555 \times 10^{15} \leq p_{32078140605053} \leq 1.09182 \times 10^{15} \,,$$

a 'standard' computer already fails to produce the prime number $p_{32078140605053}$.
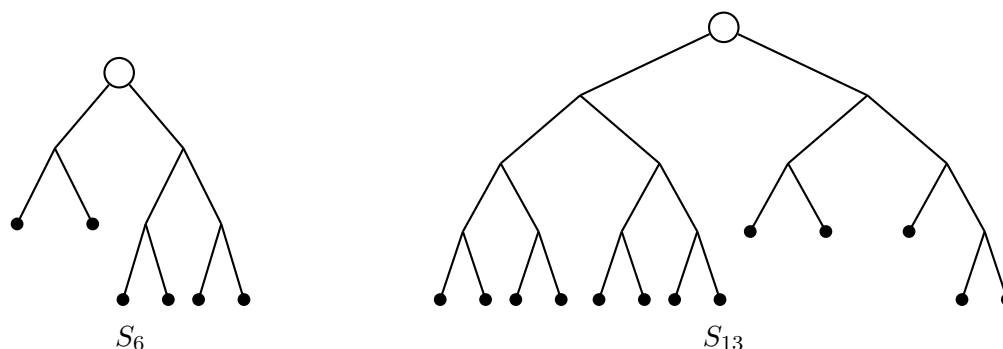
FIGURE 5. The minimal binary trees S_6 and S_{13}.

On the other hand, since "weak" binary trees (every vertex has degree at most 3) are more realistic molecular graphs, the problem of finding the extremal tree structures among these trees, given the number of vertices, may also be of interest. Recall that for the more general case of rooted trees, the problem has already been solved by Gutman and Ivić [9].

REFERENCES

[1] D. W. Matula. A natural rooted tree enumeration by prime factorization. *SIAM Review*, 10(2):273, 1968.

[2] Ivan Gutman and Aleksandar Ivić. Graphs with maximal and minimal Matula numbers. *Bull. Acad. Serbe Sci. Arts. Cl. Sci. Math. Natur.*, 19:65–74, 1994.

[3] Seymour B. Elk. A problem with the application of Matulas method of prime numbers and rooted trees for canonical nomenclatures of alkanes. *Graph theory notes (New York)*, 18:40–43, 1989.

[4] Seymour B. Elk. A canonical ordering of polybenzenes and polyadamantanes using a prime technique. *J. Math. Chem.*, 4:55–68, 1990.

[5] Ivan Gutman, Aleksandar Ivić and Seymour B. Elk. Matula numbers for coding chemical structure of their properties. *J. Serb, Chem. Soc.*, 58:193–201, 1992.

[6] Seymour B. Elk. Expansion of Matula numbers to heteroatoms and to ring compounds. *J. Chem. Inform. and Comp. Sciences*, 35(2):233-236, 1995.

[7] Seymour B. Elk. Further properties derivable from the Matula number of an alkane. *J. Chem. Inform. and Comp. Sciences*, 34(2):331–334, 1994.

[8] Ivan Gutman and Yeong-Nan Yeh. Deducing properties of trees from their Matula numbers. *Publ. Inst. Math. Belgrade*, 53:17–22, 1993.

[9] Ivan Gutman and Aleksandar Ivić. On Matula numbers. *Discrete Mathematics*, 150(1-3):131–142, 1996.

[10] Emeric Deutsch. Rooted tree statistics from Matula numbers. *Discrete Applied Mathematics*, 160(15):2314–2322, 2012.

[11] Audace A. V. Dossou-Olory and Stephan Wagner. Inducibility of topological trees. Accepted for publication in Quaestiones Mathematicae, 2018. https://arxiv.org/abs/1802.06696

[12] F. Bergeron, G. Labelle, and P. Leroux. Combinatorial species and tree-like structures. *Cambridge Univ. Press*, 67, 1998.

[13] E. S. Allman and J. A. Rhodes. Mathematical models in biology: an introduction. *Cambridge Univ. Press*, 2004.

[14] J Barkley Rosser, Lowell Schoenfeld, and others. Approximate formulas for some functions of prime numbers. *Illinois J. Mathematics*, 6(1):64–94, 1962.

[15] Guy Robin. Estimation de la fonction de Tchebychef θ sur le k-ième nombre premier et grandes valeurs de la fonction $\omega(n)$ nombre de diviseurs premiers de n. *Acta Arithmetica*, 42(4):367–389, 1983.

[16] The On-Line Encyclopedia of Integer Sequences. Published electronically at https://oeis.org, 2018.

[17] Robert E. Jamison. On the average number of nodes in a subtree of a tree. *J. Combinatorial Theory, Series B*, 35(3):207–223, 1983.

AUDACE A. V. DOSSOU-OLORY, DEPARTMENT OF MATHEMATICAL SCIENCES, STELLENBOSCH UNIVERSITY, PRIVATE BAG X1, MATIELAND 7602, SOUTH AFRICA

E-mail address: `audaced@sun.ac.za`