



Forschungsschwerpunkt

Algorithmen und mathematische Modellierung



Maximizing the number of independent subsets over trees with bounded degree

Clemens Heuberger and Stephan G. Wagner

Project Area(s):

Analysis of Digital Expansions with Applications in Cryptography

The Hardy-Littlewood Method in the Analysis of Digit Problems and Enumerative Combinatorics

Institut für Optimierung und Diskrete Mathematik (Math B)

Report 2007-3, May 2007

MAXIMIZING THE NUMBER OF INDEPENDENT SUBSETS OVER TREES WITH BOUNDED DEGREE

CLEMENS HEUBERGER AND STEPHAN G. WAGNER

ABSTRACT. The number of independent vertex subsets is a graph parameter that is, apart from its purely mathematical importance, of interest in mathematical chemistry. In particular, the problem of maximizing or minimizing the number of independent vertex subsets within a given class of graphs has already been investigated by many authors. In view of the applications of this graph parameter, trees of restricted degree are of particular interest. In the current paper, we give a characterization of the trees with given maximum degree which maximize the number of independent subsets, and show that these trees also minimize the number of independent edge subsets. The structure of these trees is quite interesting and unexpected: it can be described by means of a novel digital system—in the case of maximum degree 3, we obtain a binary system using the digits 1 and 4. The proof mainly depends on an exchange lemma for branches of a tree.

1. INTRODUCTION

The parameter *number of independent subsets* of a graph first appears in the mathematical literature in a paper of Prodinger and Tichy [14] in 1982. They call it the *Fibonacci number* of a graph in view of the fact that the number of independent subsets of a path on n vertices is exactly the Fibonacci number F_{n+2} . Among other results, they show that the number of independent vertex subsets is maximal among all trees on n vertices for the star graph S_n and minimal for the path P_n . In subsequent papers, Kirschenhofer, Prodinger and Tichy study the asymptotic behavior of this parameter for complete t -ary trees and for families of simply generated trees [8, 9].

Independently, Merrifield and Simmons [13] introduced the number of independent vertex subsets (which they call the σ -*index*) to the chemical literature in 1989, showing connections between the σ -index of a molecular graph and physicochemical properties such as boiling points. Meanwhile, the number of independent subsets of a graph is called the *Merrifield-Simmons index* in mathematical chemistry, and there is already a substantial amount of literature on chemical applications as well as on graph-theoretical properties of this index (see [3, 19] and the references therein). It belongs to the group of so-called *topological indices*, which are used to give quantitative descriptions for the structure of a (molecular) graph. Another representative of this group is the so-called *Hosoya index* [5], defined as the number of independent edge subsets (which are also referred to as *matchings*).

One of the major graph-theoretical problems in connection with these indices is the question which graphs from a given class (typically, trees) maximize or minimize the index value [6, 11, 12, 20]. It is known that the extremal graphs for the number of independent vertex subsets resp. edge subsets usually coincide, where an extremal graph maximizes one of these indices while it minimizes the other, compare [20]. It is also conjectured that the trees which minimize the Hosoya index also minimize the energy (the sum of the absolute values of a graph's eigenvalues), see [1].

Date: May 2, 2007.

2000 Mathematics Subject Classification. 05C69; 05C05 11A63.

Key words and phrases. Independent sets, d -ary trees, maximum degree, digit expansion.

This paper was written while C. Heuberger was a visitor at the Center of Experimental Mathematics at the University of Stellenbosch. He thanks the center for its hospitality.

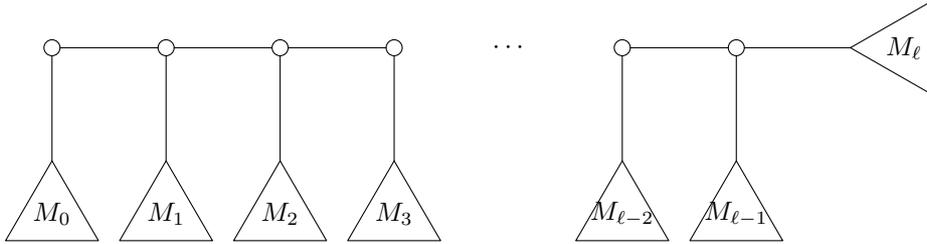
This work was also supported by the Austrian Science Foundation FWF, projects S9606 and S9611, that are part of the Austrian National Research Network "Analytic Combinatorics and Probabilistic Number Theory."

In view of the chemical applications, it is quite natural to consider trees of restricted degrees. For instance, trees of maximum degree 4 which minimize the Hosoya index have been determined by means of an extensive computer search in [1]; [18] provides a faster algorithm for this purpose. However, we believe that this question is also very interesting from a purely graph-theoretical point of view. Quite a lot of similar results are given in the graph-theoretic literature as well: for instance, Hedman [4] studies the (essentially equivalent) problem of maximizing the number of cliques in graphs with a given maximal clique size, and Wilf [21] gives the largest number of maximal independent vertex sets of a tree on n vertices, see also [15].

In the following, we fix the maximum degree $d + 1 \geq 3$ of our trees. In spite of the simple problem statement, the problem of determining the trees from this set which maximize the number of independent subsets is quite intricate and reveals an interesting and unexpected structure, which can be described by means of a novel digit system. For other graph parameters, namely the Wiener index (sum of all distances between pairs of vertices) and the number of subtrees, the extremal trees of given maximum degree are already known (see [2, 7, 16, 17])—basically, the solution is given by the complete d -ary trees.

Thus it was quite a surprise to find that the complete d -ary tree is not maximal with respect to the number of independent subsets. Yet, complete trees play a major role in the description of the extremal trees—in the case that $d = 2$, our main results read as follows:

Theorem. *Let n be a positive integer. Then there is a unique (up to isomorphism) tree T with n vertices and maximum degree ≤ 3 that simultaneously maximizes the number of independent vertex subsets and minimizes the number of independent edge subsets. It can be decomposed as*



with $M_k \in \{C_k, C_{k+2}\}$ for $0 \leq k < \ell$ and $M_{\ell} \in \{C_{\ell}, C_{\ell+1}, C_{\ell+2}\}$. Here, C_h denotes the complete binary tree of height $h - 1$ (and C_0 is the empty tree).

Furthermore, there are unique d_0, \dots, d_m such that $d_k \in \{1, 4\}$ for $0 \leq k < m$, $d_m \in \{1, 2, 4\}$, and

$$n + 1 = \sum_{k=0}^m d_k 2^k,$$

and they are given by $m = \ell$ and $d_k = 2^{\delta_k}$ for $0 \leq k \leq m$, where $\delta_k \in \{0, 1, 2\}$ is defined by $M_k = C_{k+\delta_k}$.

The paper is organized as follows: we start with basic definitions and a recursive characterization of the number of independent vertex subsets. Section 3 provides the necessary preliminary results. In Section 4, the main exchange lemma is formulated and used to prove the main result of this paper. In the final section, we show how the proof can be adapted to the case of independent edge subsets. If not stated otherwise, all trees are assumed to have maximum degree $\leq d + 1$ for some fixed integer $d \geq 2$.

2. NOTATIONS

Definition 2.1. (1) Let G be a graph. Then $\sigma(G)$ is defined to be the number of independent subsets of G .

(2) For a rooted tree T with root v , we also define $\sigma_0(T)$ to be the number of independent subsets of G not containing the root v and $\sigma_1(T)$ to be the number of independent subsets of G containing the root v . The ratio $\sigma_0(T)/\sigma(T)$ is abbreviated as $\rho(T)$.

Note that we do not mention the root v in the notations $\sigma_0(T)$, $\sigma_1(T)$, and $\rho(T)$ since the roots will usually be anonymous.

The empty set is always an independent subset of G , even if G is the empty graph. Therefore, $\sigma(G)$ is always positive.

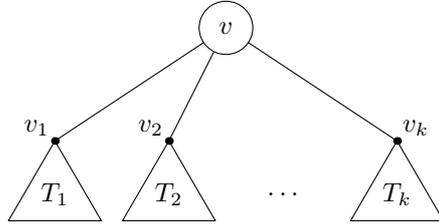


FIGURE 1. Rooted tree with branches

For a rooted tree T with root v , the connected components T_1, \dots, T_k of $T - v$ are called the branches of v , cf. Figure 1. Taking the neighbor v_j of v contained in T_j as root of T_j , T_j is again a rooted tree.

The following recursive formulæ will be used throughout the paper without further reference.

Lemma 2.2. *Let T be a rooted tree with root v and branches T_1, \dots, T_k . Then*

$$(1) \quad \sigma_0(T) = \prod_{j=1}^k \sigma(T_j),$$

$$(2) \quad \sigma_1(T) = \prod_{j=1}^k \sigma_0(T_j),$$

$$(3) \quad \rho(T) = \frac{1}{1 + \prod_{j=1}^k \rho(T_j)}.$$

Proof. The recursive formulæ (1) and (2) are easy to prove and well known, cf. for instance [3, 10].

Equation (3) is a consequence of (1) and (2). \square

3. ROOTED TREES

A *rooted d -ary tree* is a rooted tree where every vertex has 0 or d children. The (rooted) *complete d -ary tree* of height $n - 1$ is denoted by C_n , i.e., C_1 is a single vertex and C_n has d branches C_{n-1}, \dots, C_{n-1} , cf. Figure 2. It is convenient to define C_0 to be the empty graph with $\sigma_0(C_0) = 1$ and $\sigma_1(C_0) = 0$. Note that equations (1) and (2) remain valid with this setting.

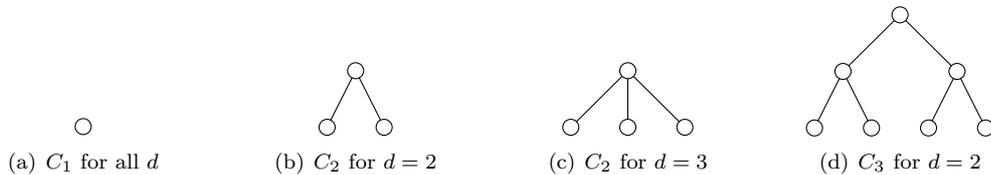


FIGURE 2. Complete d -ary trees

The number of vertices of C_n satisfies

$$|V(C_{n+1})| = 1 + d|V(C_n)| \text{ for } n \geq 0, \quad |V(C_0)| = 1$$

and therefore

$$|V(C_n)| = \frac{d^n - 1}{d - 1}$$

for $n \geq 0$.

The sequence $\rho(C_n)$, $n \geq 0$, has been studied by Kirschenhofer, Prodinger, and Tichy in [8]. By (3), we have

$$\rho(C_n) = \frac{1}{1 + \rho(C_{n-1})^d} \quad \text{for } n \geq 1, \quad \rho(C_0) = 1.$$

We list the first few values of $\rho(C_n)$ since they occur frequently:

$$\rho(C_0) = 1, \quad \rho(C_1) = \frac{1}{2}, \quad \rho(C_2) = 1 - \frac{1}{2^d + 1}.$$

Lemma 3.1 ([8]). *The subsequences of $\rho(C_n)$ of odd and even indices, respectively, are convergent to limits c_{odd} and c_{even} , respectively, where $c_{\text{odd}} \leq c_{\text{even}}$.*

The subsequence of $\rho(C_n)$ with odd indices is strictly increasing,

$$\frac{1}{2} = \rho(C_1) < \rho(C_3) < \cdots < \rho(C_{2k-1}) < \rho(C_{2k+1}) < \cdots < c_{\text{odd}},$$

whereas the subsequence with even indices is strictly decreasing,

$$1 = \rho(C_0) > \rho(C_2) > \cdots > \rho(C_{2k}) > \rho(C_{2k+2}) > \cdots > c_{\text{even}}.$$

We note a first rough estimate of $\rho(T)$ for rooted trees T .

Lemma 3.2. *Let T be a rooted tree. Then*

$$\frac{1}{2} \leq \rho(T) \leq 1 - \frac{1}{2^d + 1},$$

unless T is empty, where $\rho(T) = 1$.

Proof. Since $\rho(C_0) = 1$, we can assume that T has at least one vertex and proceed by induction on the number of vertices. Denote the branches of T by T_1, \dots, T_d (allowing T_j to be the empty graph), where each T_j satisfies $1/2 \leq \rho(T_j) \leq 1$ by the induction hypothesis. We obtain

$$\frac{1}{2} = \frac{1}{1+1} \leq \rho(T) = \frac{1}{1 + \rho(T_1) \cdots \rho(T_d)} \leq \frac{1}{1 + \frac{1}{2} \cdots \frac{1}{2}} = 1 - \frac{1}{2^d + 1}.$$

□

Definition 3.3. Let T be a possibly rooted tree. Then we construct the *outline graph* of T by replacing all maximal subtrees isomorphic to some C_k , $k \geq 0$, by a special leaf C_k . In this process, we attach $(d+1-r)$ leaves C_0 to internal nodes (non-leaves and non-root) of degree r with $2 \leq r \leq d$. If T is a rooted tree with a root of degree r ($1 \leq r \leq d$), then we also attach $d-r$ leaves C_0 to it.

The construction ensures that the outline graph of a rooted tree is a rooted d -ary tree, and that the outline graph of an arbitrary tree of maximum degree $\leq d+1$ has only vertices of degree 1 and $d+1$. An example is shown in Figure 3. The outline of a rooted tree C_k is just the rooted tree consisting of the single leaf C_k .

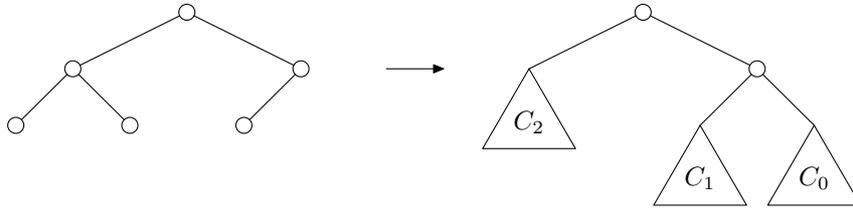


FIGURE 3. Reduction to the outline graph ($d = 2$)

If enough information on the outline of a rooted tree is available, we can determine it from its ρ -value.

Lemma 3.4. *Let $j \geq 0$ be an integer and T be a rooted tree whose outline does not contain any C_k for $0 \leq k \leq j - 3$. If*

$$j \text{ is even and } \rho(C_j) \leq \rho(T)$$

or

$$j \text{ is odd and } \rho(T) \leq \rho(C_j),$$

then $T \in \{C_{j-2}, C_j\}$.

For clarity, we remark that for $j < 2$, we assert that $T = C_j$.

Proof. We first assume that $T = C_\ell$ for some ℓ . Since the outline of T does not contain a C_k for $k \leq j - 3$, we conclude that $\ell \geq j - 2$. The monotonicity properties of $\rho(C_n)$ in Lemma 3.1 imply that $\ell \in \{j - 2, j\}$ and we are done. Thus we may assume that the outline of T is a d -ary rooted tree with more than one vertex.

We proceed by induction on j . The case of $j = 0$ has been shown in Lemma 3.2. Assume that $j \geq 1$. By our assumption, T is not empty and has (possibly empty) branches T_1, T_2, \dots, T_d . Furthermore, there is no C_k in the outlines of the T_i for $k \leq j - 3$.

Assume that j is odd. Then

$$\frac{1}{1 + \rho(T_1) \dots \rho(T_d)} = \rho(T) \leq \rho(C_j) = \frac{1}{1 + \rho(C_{j-1})^d},$$

which is equivalent to

$$(4) \quad \rho(C_{j-1})^d \leq \rho(T_1) \dots \rho(T_d).$$

Without loss of generality, we may assume $\rho(T_1) \geq \rho(T_2) \geq \dots \geq \rho(T_d)$.

We claim that $\rho(T_i) = C_{j-1}$ for $1 \leq i \leq d$. We proceed by induction on i . Assume that the claim is proved up to $i - 1$ for some $i \geq 1$. Then dividing (4) by $\rho(C_{j-1})^{i-1}$ yields

$$\rho(C_{j-1})^{d-i+1} \leq \rho(T_i) \dots \rho(T_d) \leq \rho(T_i)^{d-i+1},$$

thus $\rho(C_{j-1}) \leq \rho(T_i)$. By the (outer) induction hypothesis, $T_i \in \{C_{j-1}, C_{j-3}\}$. Since the outline of T_i does not contain any C_k for $k \leq j - 3$, we clearly have $T_i \neq C_{j-3}$ and therefore $T_i = C_{j-1}$, as required.

Thus $T_1 = \dots = T_i = C_{j-1}$ and therefore $T = C_j$, which has been handled earlier.

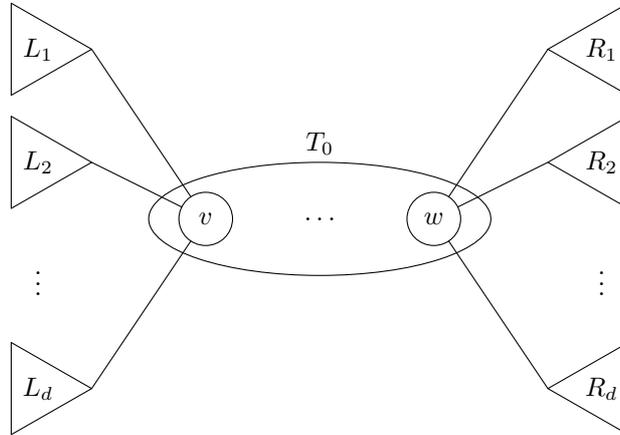
The proof for even j follows by exchanging ‘‘odd’’ and ‘‘even’’ and reversing all inequality signs. \square

4. OPTIMAL TREES

A tree T is called maximal, if $\sigma(T) \geq \sigma(T')$ for all trees T' with the same number of vertices as T (and maximum degree $\leq d + 1$).

Our key tool is the following local exchange lemma.

Lemma 4.1. *Let T be a maximal tree. If there are (possibly empty) rooted trees $L_1, \dots, L_d, R_1, \dots, R_d$ and a tree T_0 such that T can be decomposed as*



and such that $\rho(L_1) < \rho(R_1)$, then

$$\max\{\rho(L_i) : 1 \leq i \leq d\} \leq \min\{\rho(R_i) : 1 \leq i \leq d\}.$$

Proof. We need the following auxiliary quantities:

- $\sigma_{00}(T_0)$: number of independent subsets of T_0 containing neither v nor w .
- $\sigma_{10}(T_0)$: number of independent subsets of T_0 containing v , but not w .
- $\sigma_{01}(T_0)$: number of independent subsets of T_0 containing w , but not v .
- $\sigma_{11}(T_0)$: number of independent subsets of T_0 containing both v and w .

Define

$$F(L_1, \dots, L_d, R_1, \dots, R_d) := \sigma_{00}(T_0) + \sigma_{10}(T_0) \prod_{i=1}^d \rho(L_i) + \sigma_{01}(T_0) \prod_{i=1}^d \rho(R_i) + \sigma_{11}(T_0) \prod_{i=1}^d \rho(L_i) \prod_{i=1}^d \rho(R_i).$$

Then it is easily seen that

$$\sigma(T) = F(L_1, \dots, L_d, R_1, \dots, R_d) \prod_{i=1}^d \sigma(L_i) \prod_{i=1}^d \sigma(R_i).$$

Since T is maximal, we must have

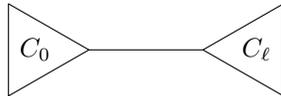
$$F(L_1, \dots, L_d, R_1, \dots, R_d) \geq F(\pi(L_1), \dots, \pi(L_d), \pi(R_1), \dots, \pi(R_d))$$

for all permutations π of $\{L_1, \dots, L_d, R_1, \dots, R_d\}$.

The first and the fourth summand of $F(\pi(L_1), \dots, \pi(L_d), \pi(R_1), \dots, \pi(R_d))$ are independent of the permutation π , so the sum of the second and the third summand is maximal. By the rearrangement inequality (note that all quantities are positive), we conclude that $\rho(L_i) \leq \rho(R_i)$ for $1 \leq i \leq d$ and $\sigma_{10}(T_0) \leq \sigma_{01}(T_0)$. Exchanging R_i and R_j for some $i \neq j$ does not change anything, but applying the rearrangement inequality in that situation, we obtain $\rho(L_i) \leq \rho(R_j)$ and $\rho(L_j) \leq \rho(R_i)$. \square

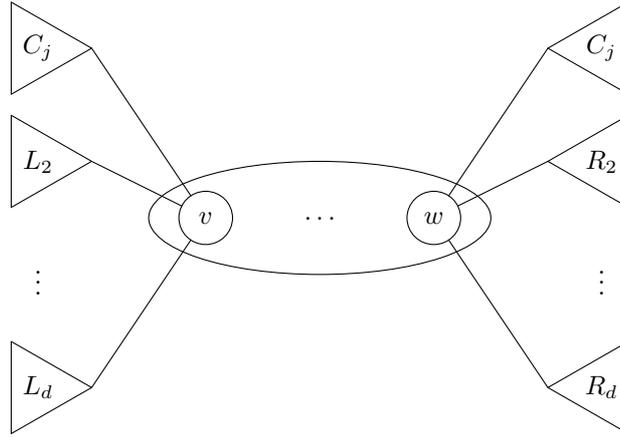
Remark 4.2. It is a consequence of Lemma 4.1 that a maximal tree T contains at most one non-leaf vertex of degree $< d+1$: otherwise, let v, w be two vertices of degrees D_1, D_2 ($1 < D_1, D_2 < d+1$); then we have $L_{D_1} = \dots = L_d = C_0$, $R_{D_2} = \dots = R_d = C_0$ (and $L_1, \dots, L_{D_1-1}, R_1, \dots, R_{D_2-1} \neq C_0$) in Lemma 4.1, which is applicable since $\rho(L_1) < \rho(R_d) = \rho(C_0) = 1$ by Lemma 3.2. We obtain $1 = \rho(C_0) = \rho(L_d) \leq \rho(R_1)$, i.e., $R_1 = C_0$ by Lemma 3.2, a contradiction. This result has already been proved in [18].

For our purposes, we need a stronger version of Remark 4.2. Here, we use the outline graph of the maximal tree T . Note that since we do not assume it to be a rooted tree, some copies of C_0 are attached to all vertices of degree r with $2 \leq r \leq d$. For instance, if $T = C_\ell$ for some $\ell \geq 2$, its outline graph is:



Lemma 4.3. *Let T be a maximal tree and let j be the least nonnegative integer such that the outline graph of T contains a C_j . Then the outline graph of T contains C_j at most $(d-1)$ times and there is a vertex v of the outline graph of T which is adjacent to all copies of C_j in the outline graph of T .*

Proof. Assume that there are two copies of C_j in the outline graph of T that are adjacent to v and to w , respectively, for different vertices v and w of the outline graph of T . For suitable rooted trees L_2, \dots, L_d and R_2, \dots, R_d , the outline of T can be decomposed as follows:



By our assumption, the outlines of these rooted trees do not contain a C_ℓ for $\ell < j$.

Assume that j is odd. We claim that there is a L_i with $2 \leq i \leq d$ such that $\rho(L_i) > \rho(C_j)$. Otherwise, all L_i satisfy $\rho(L_i) \leq \rho(C_j)$ and do not contain a C_ℓ for $\ell < j$, which implies that $L_i = C_j$ by Lemma 3.4 for all $2 \leq i \leq d$. Thus C_j, L_2, \dots, L_d could be merged to a C_{j+1} , which is a contradiction to the assumption that the outline graph of T has the form given above.

Therefore, there is a L_i with $\rho(L_i) > \rho(C_j)$. By Lemma 4.1, we conclude that $\rho(R_i) \leq \rho(C_j)$ for all $2 \leq i \leq d$. As above, Lemma 3.4 implies that $R_2 = \dots = R_d = C_j$, again a contradiction.

Hence all copies of C_j in the outline of T are adjacent to the same vertex v . There cannot be d copies of C_j , because they would be merged to a C_{j+1} . Therefore, there are at most $d - 1$ copies of C_j .

Once again, the proof for even j follows by exchanging “odd” and “even” and reversing all inequality signs. \square

The following lemma performs one step of the decomposition of a maximal tree T .

Lemma 4.4. *Let T be a maximal tree, k be a nonnegative integer and assume that the outline graph of T can be decomposed as*



for some rooted trees L_k (possibly empty) and R_k with

$$(5a) \quad k \text{ is odd and } \rho(C_k) < \rho(L_k) < \rho(C_{k+2})$$

or

$$(5b) \quad k \text{ is even and } \rho(C_{k+2}) < \rho(L_k) < \rho(C_k)$$

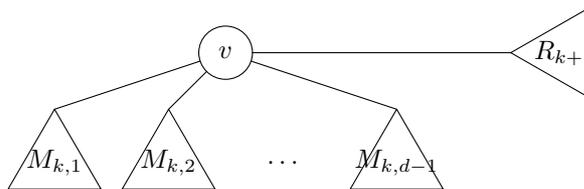
or

$$L_k = C_k.$$

Assume that R_k is non-empty and the outline of R_k does not contain any C_ℓ with $\ell < k$.

Then exactly one of the following assertions is true:

- (1) $R_k \in \{C_k, C_{k+1}, C_{k+3}\}$,
- (2) R_k consists of d branches $C_{k+1}, C_{k+1}, C_{\ell_3}, \dots, C_{\ell_d}$ with $\ell_i \in \{k, k+1, k+2\}$ for $3 \leq i \leq d$,
- (3) the outline of R_k can be decomposed as



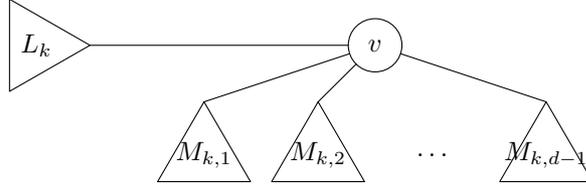
for $M_{k,1}, \dots, M_{k,d-1} \in \{C_k, C_{k+2}\}$ and a non-empty rooted tree R_{k+1} whose outline does not contain any C_ℓ for $\ell \leq k$. Furthermore,

$$(6a) \quad k \text{ is odd and } \rho(C_{k+3}) < \rho(L_{k+1}) < \rho(C_{k+1})$$

or

$$(6b) \quad k \text{ is even and } \rho(C_{k+1}) < \rho(L_{k+1}) < \rho(C_{k+3})$$

where L_{k+1} is defined as follows:



Before we prove the lemma it is worth pointing out that its assumptions are satisfied for $k = j$ and $L_k = C_j$ by Lemma 4.3.

Proof. We first note that the three assertions are indeed mutually exclusive. In the second case, R_k has at least two branches C_{k+1} . Thus R_k contains more vertices than a C_k or a C_{k+1} , but we also have $R_k \neq C_{k+3}$; so we are not in the first case. In the third case, R_k has at most one branch C_{k+1} , so the second and the third case are mutually exclusive. Finally, the third and the first case are mutually exclusive since the outline of R_k can be decomposed in the third case, but this cannot be done in the first case.

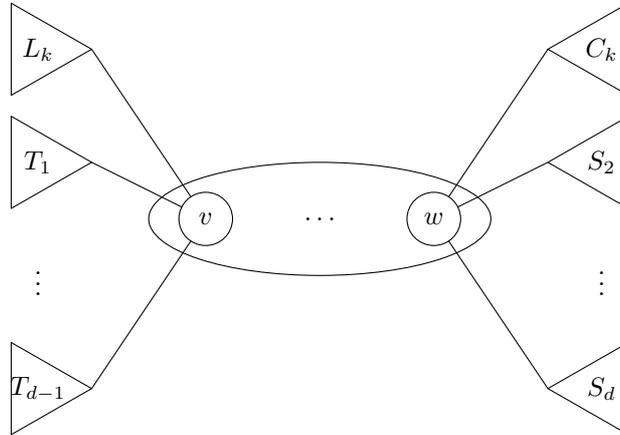
If $R_k \in \{C_k, C_{k+1}, C_{k+3}\}$, there is nothing to prove. Thus we assume that this is not the case. We consider the case of odd k , the other case follows by reversing the signs of the inequalities.

Claim 1. $R_k \neq C_1$. The outline of R_k does not contain a C_1 for $k \geq 2$, thus $R_k \neq C_1$ for $k \geq 2$. For $k \in \{0, 1\}$ the case $R_k = C_1$ has already been excluded from our considerations. Therefore, R_k has at least one non-empty branch. We denote the branches of R_k by T_1, \dots, T_d , where some of them are allowed to be empty.

We observe that the outline graphs of T_1, \dots, T_d are the branches of the outline graph of R_k unless $R_k = C_\ell$ for some ℓ .

Claim 2. The outlines of T_i , $1 \leq i \leq d$, do not contain any C_ℓ with $\ell < k$. By the above observation and the assumptions of the lemma, this could only occur if $T_1 = \dots = T_d = C_{k-1}$ and $R_k = C_k$, and this has already been ruled out.

Claim 3. If the outline of T_i contains C_k for some $i \in \{1, 2, \dots, d\}$, then $T_i = C_k$. Consider first the case that $L_k = C_k$. In that case, all C_k in the outline graph of T are adjacent to the same vertex v by Lemma 4.3, i.e., they are branches of R_k . Thus the outline graph of T_i can only contain a C_k if $T_i = C_k$. So we may assume $L_k \neq C_k$ and that the outline of T_d , say, contains a C_k with $T_d \neq C_k$. We can decompose the outline of T_d such that we get a decomposition of T as



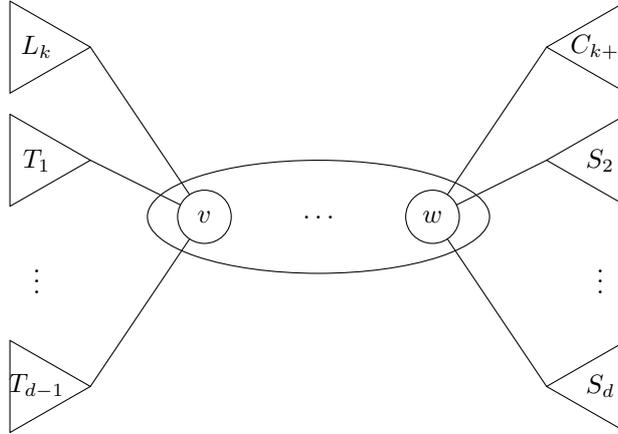
for suitable (possibly empty) rooted trees S_2, \dots, S_d . The outline of S_i is a subtree of the outline of T_d , therefore, it does not contain a C_ℓ for $\ell < k$. We have $\rho(C_k) < \rho(L_k)$ by (5a) and therefore, by Lemma 4.1, $\rho(S_i) \leq \rho(L_k) < \rho(C_{k+2})$ for $2 \leq i \leq d$. From Lemma 3.4, we conclude that $S_2 = \dots = S_d = C_k$ (the other possibility cannot occur due to the strict inequality). But this means that the d copies of C_k would have been merged to a C_{k+1} in the outline of T_d , contradiction. Our claim is proved.

Claim 4. If $T_1, \dots, T_{d-1} \in \{C_k, C_{k+2}\}$, then we are done. If $T_1 = \dots = T_{d-1} = T_d$, then $R_k \in \{C_{k+1}, C_{k+3}\}$, which has been excluded. Otherwise, the outline of R_k decomposes into the outlines of T_1, \dots, T_d . If $T_d = C_k$, we exchange T_d and some $T_i \neq C_k$ (which exists since $R_k \neq C_{k+1}$). Thus we may assume that $T_d \neq C_k$. In particular, T_d is not empty, since the outline of R_k is known not to contain any C_ℓ for $\ell < k$. We set $M_{k,i} = T_i$ for $1 \leq i \leq d-1$ and $R_{k+1} = T_d$. By our above claim, we know that the outline of R_{k+1} does not contain a C_ℓ for $\ell \leq k$. The inequality

$$\rho(C_{k+3}) = \frac{1}{1 + \rho(C_{k+2})^d} < \rho(L_{k+1}) = \frac{1}{1 + \rho(L_k)\rho(M_{k,1}) \cdots \rho(M_{k,d-1})} \leq \frac{1}{1 + \rho(C_k)^d} = \rho(C_{k+1})$$

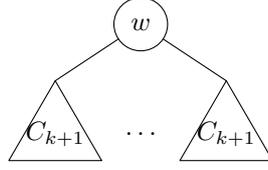
holds by (5a) and since $\rho(C_k) \leq M_{k,1} \leq \rho(C_{k+2})$ for all $1 \leq i \leq d-1$. To show that the right inequality is strict, we note that $\rho(L_k) = \rho(M_{k,1}) = \rho(M_{k,d-1}) = \rho(C_k)$ implies $L_k = M_{k,1} = \dots = M_{k,d} = C_k$, but then we have d copies of C_k adjacent to the same vertex, which could therefore be merged to a C_{k+1} . This is a contradiction to our assumption that the original decomposition of T into L_k and R_k is a decomposition of its outline graph. Thus we showed (6a) and the claim is proved.

Claim 5. If one of the T_i , say T_d , contains a C_{k+2} as a proper subtree, then we are done. For clarity we emphasize that C_{k+2} is just assumed to be a subtree of T_d , we neither require it to be contained in the outline of T_d nor to be a branch of T_d . We can decompose T as



for suitable (possibly empty) rooted trees S_2, \dots, S_d . By (5a), we have $\rho(L_k) < \rho(C_{k+2})$ and we apply Lemma 4.1 to deduce $\rho(T_i) \leq \rho(C_{k+2})$ for $1 \leq i \leq d-1$. Since the outline of T_i is known to contain no C_ℓ for $\ell < k$, we obtain $T_i \in \{C_k, C_{k+2}\}$ from Lemma 3.4 for $1 \leq i \leq d-1$. We are done by the previous claim.

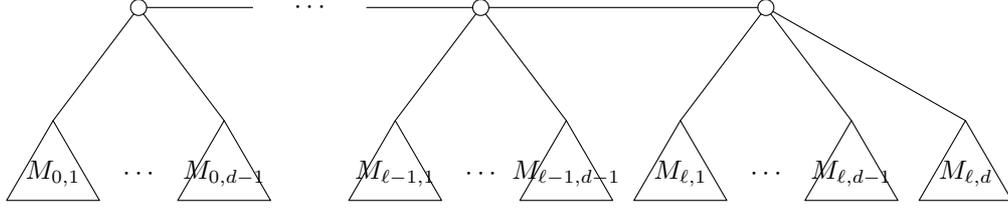
We may now assume that none of the T_i contains a C_{k+2} as a proper subtree, and claim that this implies that $T_i \in \{C_k, C_{k+1}, C_{k+2}\}$ for all $1 \leq i \leq d$. Assume that $T_i \notin \{C_k, C_{k+1}, C_{k+2}\}$. A leaf of the outline of T_i is of the form C_ℓ for some ℓ by the definition of an outline graph. However, the range for ℓ is already very restricted: We have $\ell \geq k+1$ since the outline graph of T_i has been proved to contain no smaller C_ℓ . On the other hand, we have $\ell \leq k+1$, since T_i does not contain C_{k+2} as a proper subtree. In short, all leaves of T_i equal C_{k+1} . Consider a leaf C_{k+1} of maximum height (distance from the root). This leaf C_{k+1} has a parent w , and all of the children of w have to be leaves since our original leaf was assumed to be of maximum height. Thus we found the subgraph



of the outline graph of T_i , this would have been contracted to a C_{k+2} , contradiction.

We are now in the situation that all $T_i \in \{C_k, C_{k+1}, C_{k+2}\}$. If there are at least two copies of C_{k+1} amongst the T_i , we are in the second case. Otherwise, we have $T_1, \dots, T_{d-1} \in \{C_k, C_{k+2}\}$ (after renumbering the branches), and we are done by one of the above claims. \square

Theorem 1. *Let n be a positive integer. Then there is a unique (up to isomorphism) maximal tree T with n vertices and maximum degree $\leq d+1$. It can be decomposed as*



with $M_{k,1}, \dots, M_{k,d-1} \in \{C_k, C_{k+2}\}$ for $0 \leq k < \ell$ and either $M_{\ell,1} = \dots = M_{\ell,d} = C_{\ell-1}$ or $M_{\ell,1} = \dots = M_{\ell,d} = C_\ell$ or $M_{\ell,1}, \dots, M_{\ell,d} \in \{C_\ell, C_{\ell+1}, C_{\ell+2}\}$, where at least two of $M_{\ell,1}, \dots, M_{\ell,d}$ equal $C_{\ell+1}$.

Proof. Let T be a maximal tree with n vertices, and let j be the least integer such that the outline graph of T contains C_j . Now, we apply Lemma 4.4 inductively, starting with $k = j$ and $L_j = C_j$, until we reach a point where either assertion (1) or assertion (2) holds in Lemma 4.4. Now, we expand L_j artificially: for $j \geq r \geq 1$, we further decompose $L_r = C_r$ into its branches $L_{r-1} = C_{r-1}$ and $M_{r-1,1} = \dots = M_{r-1,d-1} = C_{r-1}$. At the end, $L_0 = C_0$ is discarded, but the $M_{0,i}$ are retained even if they equal C_0 . Finally,

- if $R_k = C_k$, we set $\ell = k$, $M_{\ell,1} = \dots = M_{\ell,d} = C_{\ell-1}$,
- if $R_k = C_{k+1}$, we set $\ell = k$, $M_{\ell,1} = \dots = M_{\ell,d} = C_\ell$,
- if $R_k = C_{k+3}$, we set $\ell = k+1$, $M_{\ell-1,1} = \dots = M_{\ell-1,d-1} = C_{\ell+1}$ and $M_{\ell,1} = \dots = M_{\ell,d} = C_\ell$,
- if (2) holds in Lemma 4.4, we set $\ell = k$ and use the branches of R_k as our $M_{\ell,1}, \dots, M_{\ell,d}$.

Now it can be seen that T has exactly the form that is described in the statement of the theorem, we only have to check that T is uniquely determined by this description. For this purpose, we count the number of vertices in the representation given in the theorem. Let r_k be the number of $M_{k,i}$ which are equal to C_{k+2} . Then the number of vertices in all $M_{k,i}$, together with the vertex they are all attached to, equals

$$(d-1-r_k) \cdot \frac{d^k-1}{d-1} + r_k \cdot \frac{d^{k+2}-1}{d-1} + 1 = d^k(1+(d+1)r_k)$$

for $1 \leq k < \ell$. Furthermore, let q_ℓ, r_ℓ be the number of branches $M_{\ell,i}$ which are equal to $C_{\ell+1}$ and $C_{\ell+2}$ respectively. Then the number of vertices in all $M_{\ell,i}$, together with the vertex they are all attached to, equals either $\frac{d^\ell-1}{d-1}$ (if all $M_{\ell,i} = C_{\ell-1}$) or $\frac{d^{\ell+1}-1}{d-1}$ (if all $M_{\ell,i} = C_\ell$) or

$$(d-q_\ell-r_\ell) \cdot \frac{d^\ell-1}{d-1} + q_\ell \cdot \frac{d^{\ell+1}-1}{d-1} + r_\ell \cdot \frac{d^{\ell+2}-1}{d-1} + 1 = \frac{d^{\ell+1}-1}{d-1} + q_\ell d^\ell + r_\ell (d+1)d^\ell.$$

It follows that

$$(7) \quad (d-1)n+1 = \sum_{k=0}^{\ell} a_k d^k,$$

where $a_k = (d-1)(1+(d+1)r_k)$ with $0 \leq r_k \leq d-1$ for $k < \ell$ and $a_\ell = 1$ or $a_\ell = d$ or $a_\ell = d+(d-1)q_\ell+(d^2-1)r_\ell$ with $q_\ell \geq 2$ and $q_\ell+r_\ell \leq d$.

Now assume that there is another expansion

$$(8) \quad (d-1)n+1 = \sum_{k=0}^{\ell'} a'_k d^k$$

that satisfies the same conditions; in particular, r'_k, q'_ℓ are given analogously. Let h be the least integer such that $d_h \neq d'_h$. Then (7) and (8) yield

$$(9) \quad \sum_{k=h}^{\ell} a_k d^{k-h} = \sum_{k=h}^{\ell'} a'_k d^{k-h}.$$

If $h < \min\{\ell, \ell'\}$, we consider (9) modulo d and see that $a_h \equiv a'_h \pmod{d}$ or $r_h \equiv r'_h \pmod{d}$, which is impossible, since $0 \leq r_h, r'_h < d$ and $r_h \neq r'_h$, a contradiction. Hence, assume without loss of generality that $h = \ell'$. If $\ell = \ell' = h$, it follows that $d_\ell = d'_\ell$, yielding a contradiction once again. Thus $\ell > \ell' = h$. Now, using (9) modulo d once again, we see that

$$a_h = (d-1)(1 + (d+1)r_h) \equiv d + (d-1)q'_h + (d^2-1)r'_h = a'_h \pmod{d}$$

if $a'_h \notin \{1, d\}$, which simplifies to $r_h + 1 \equiv q'_h + r'_h \pmod{d}$. Since $1 \leq r_h + 1 \leq d$ and $1 \leq q'_h + r'_h \leq d$, it follows that $r_h = q'_h + r'_h - 1$. Now, using the condition $q'_h \geq 2$, we obtain the inequality

$$\begin{aligned} a'_h &= d + (d-1)q'_h + (d^2-1)r'_h \leq d + 2(d-1) + (d^2-1)(q'_h + r'_h - 2) \\ &= 3d - 2 + (d^2-1)(r_h - 1) = -d^2 + 3d - 1 + (d^2-1)r_h \leq d - 1 + (d^2-1)r_h = a_h. \end{aligned}$$

This estimate also holds if $a'_h = 1$ (trivially) and if $a'_h = d$. In the latter case, $a'_h \equiv a_h \pmod{d}$ implies $r_h = d - 1$ and $a_h = (d-1)d^2 > a'_h$. But now it follows that the right hand side of (9) is smaller than its left hand side, contradiction. Thus the uniqueness of the representation (7) is shown, which means that the maximal tree T is uniquely characterized up to isomorphism. \square

Remark 4.5. From the construction, it is clear that a representation of $(d-1)n+1$ of the form (7) always exists. Of course, this can also be shown directly. The numbers a_k can be interpreted as digits, thus yielding a novel digital system that seems to be of interest on its own right. In the special case $d = 2$, things simplify a lot, and one obtains a representation of the form

$$n+1 = \sum_{k=0}^m d_k 2^k,$$

with $d_k \in \{1, 4\}$ and $d_m \in \{1, 2, 4\}$, as it has been presented in the introduction.

5. INDEPENDENT EDGE SUBSETS

In this section, we will give a proof of an analogous theorem for independent edge subsets that will basically follow the same lines. We start with the appropriate definitions:

- Definition 5.1.**
- (1) Let G be a graph. Then $z(G)$ is defined to be the number of independent edge subsets of G .
 - (2) For a rooted tree T with root v , we also define $z_0(T)$ to be the number of independent edge subsets of G not containing an edge incident with the root v and $z_1(T)$ to be the number of independent edge subsets of G containing an edge incident with the root v . The ratio $z_0(T)/z(T)$ is abbreviated as $\tau(T)$.

Again, it is possible to give simple recursive formulæ for z , z_0 and z_1 :

Lemma 5.2. *Let T be a rooted tree with root v and branches T_1, \dots, T_k . Then*

$$(10) \quad z_0(T) = \prod_{j=1}^k z(T_j),$$

$$(11) \quad z_1(T) = \sum_{j=1}^k \frac{z_0(T_j)}{z(T_j)} \prod_{i=1}^k z(T_i),$$

$$(12) \quad \tau(T) = \frac{1}{1 + \sum_{j=1}^k \tau(T_j)}.$$

Proof. The formulæ for z_0 and z_1 are easy to prove and can be found in [3, 10] again, and the identity for $\tau(T)$ follows at once. \square

From formula (12), we obtain a simple recursion for $\tau(C_n)$. We define $z_0(C_0) = 0$, $z_1(C_0) = 1$ (so that (10) and (11) remain valid) and have, by (12),

$$\tau(C_n) = \frac{1}{1 + d\tau(C_{n-1})} \quad \text{for } n \geq 1, \quad \tau(C_0) = 0.$$

The first values are given by

$$\tau(C_0) = 0, \quad \tau(C_1) = 1, \quad \tau(C_2) = \frac{1}{d+1}.$$

It is an easy exercise to prove the following lemma:

Lemma 5.3. *The sequence $\tau(C_n)$, $n \geq 0$, is convergent with limit $c = \frac{\sqrt{4d+1}-1}{2d}$.*

The subsequence with odd indices is strictly decreasing,

$$1 = \tau(C_1) > \tau(C_3) > \dots > \tau(C_{2k-1}) > \tau(C_{2k+1}) > \dots > c,$$

whereas the subsequence with even indices is strictly increasing,

$$0 = \tau(C_0) < \tau(C_2) < \dots < \tau(C_{2k}) < \tau(C_{2k+2}) < \dots < c.$$

Furthermore, we have the following analogue of Lemma 3.2 (the proof is basically the same):

Lemma 5.4. *Let T be a rooted tree. Then*

$$\frac{1}{d+1} \leq \tau(T) \leq 1,$$

unless T is empty, where $\tau(T) = 0$.

Note that it doesn't make a difference in the proof of Lemma 3.4 whether we have a product or a sum in the denominator of formula (3). Hence we obtain the following analogue of Lemma 3.4 (we also remark that the roles of "odd" and "even" are interchanged in view of Lemma 5.3):

Lemma 5.5. *Let $j \geq 0$ be an integer and T be a rooted tree whose outline does not contain any C_k for $0 \leq k \leq j-3$. If*

$$j \text{ is odd and } \tau(C_j) \leq \tau(T)$$

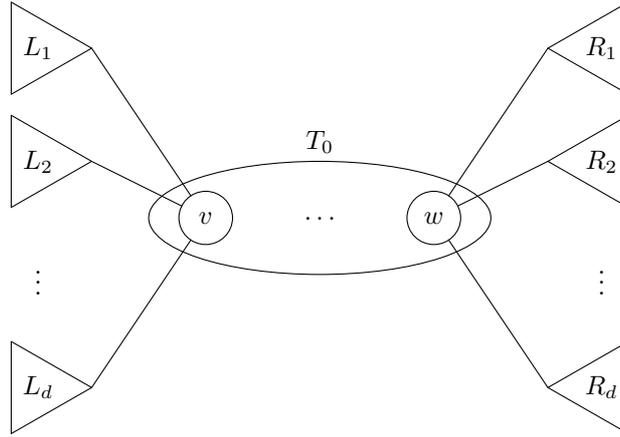
or

$$j \text{ is even and } \tau(T) \leq \tau(C_j),$$

then $T \in \{C_{j-2}, C_j\}$.

Now, it only remains to prove an analogue of the key lemma in our proof:

Lemma 5.6. *Let T be a minimal tree in the sense that it minimizes $z(T)$ among all trees of maximum degree $\leq d+1$ on n vertices. If there are (possibly empty) rooted trees $L_1, \dots, L_d, R_1, \dots, R_d$ and a tree T_0 such that T can be decomposed as*



and such that $\tau(L_1) < \tau(R_1)$, then

$$\max\{\tau(L_i) : 1 \leq i \leq d\} \leq \min\{\tau(R_i) : 1 \leq i \leq d\}.$$

Proof. Again, we need four auxiliary quantities:

- $z_{00}(T_0)$: number of independent edge subsets of T_0 not containing an edge that is incident with either v or w .
- $z_{10}(T_0)$: number of independent edge subsets of T_0 containing an edge incident with v , but no edge incident with w .
- $z_{01}(T_0)$: number of independent edge subsets of T_0 containing an edge incident with w , but no edge incident with v .
- $z_{11}(T_0)$: number of independent edge subsets of T_0 containing an edge incident with v and containing an edge incident with w .

Define

$$\begin{aligned} G(L_1, \dots, L_d, R_1, \dots, R_d) &:= z_{00}(T_0) \left(1 + \sum_{i=1}^d \tau(L_i)\right) \left(1 + \sum_{i=1}^d \tau(R_i)\right) \\ &\quad + z_{10}(T_0) \left(1 + \sum_{i=1}^d \tau(R_i)\right) + z_{01}(T_0) \left(1 + \sum_{i=1}^d \tau(L_i)\right) + z_{11}(T_0). \end{aligned}$$

Then it is easily seen that

$$z(T) = G(L_1, \dots, L_d, R_1, \dots, R_d) \prod_{i=1}^d z(L_i) \prod_{i=1}^d z(R_i).$$

In view of the minimality of $z(T)$, we must have

$$G(L_1, \dots, L_d, R_1, \dots, R_d) \leq G(\pi(L_1), \dots, \pi(L_d), \pi(R_1), \dots, \pi(R_d))$$

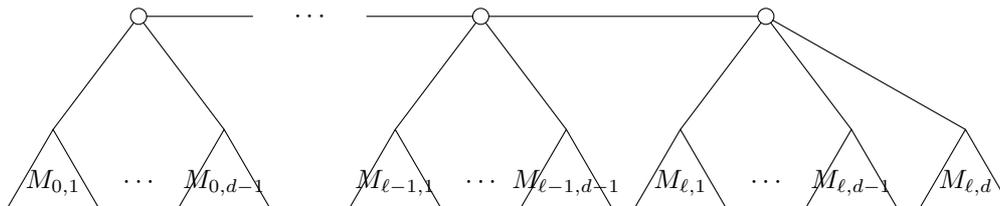
for all permutations π of $\{L_1, \dots, L_d, R_1, \dots, R_d\}$. Ignoring the assumption $\tau(L_1) < \tau(R_1)$ for the moment, we see that the minimum of the first summand among all possible permutations is attained if

$$(13) \quad \max\{\tau(L_i) : i = 1, \dots, d\} \leq \min\{\tau(R_i) : i = 1, \dots, d\} \text{ or} \\ \min\{\tau(L_i) : i = 1, \dots, d\} \geq \max\{\tau(R_i) : i = 1, \dots, d\}$$

by standard arguments (note that the sum of the two factors does not depend on the permutation). The sum of the second and third summand is minimized if $\max\{\tau(L_i) : i = 1, \dots, d\} \leq \min\{\tau(R_i) : i = 1, \dots, d\}$ in the case $z_{10}(T_0) \leq z_{01}(T_0)$ and is minimized if $\min\{\tau(L_i) : i = 1, \dots, d\} \geq \max\{\tau(R_i) : i = 1, \dots, d\}$ in the case that $z_{10}(T_0) \geq z_{01}(T_0)$. Therefore, the minimality of G yields (13). The assumption $\tau(L_1) < \tau(R_1)$ implies the first possibility. \square

From this point on, the remaining steps are literally the same as in the case of independent vertex subsets (note again that the only slight difference lies in the fact that we have a sum in the denominator of formula (12), as opposed to the product in formula (3), but this doesn't alter the argument). Hence we just state the resulting theorem—as it has been stated in the introduction, the trees which maximize the number of independent vertex subsets among all trees of maximum degree $\leq d + 1$ are also the ones which minimize the number of independent edge subsets.

Theorem 2. *Let n be a positive integer. Then there is a unique (up to isomorphism) tree T with n vertices and maximum degree $\leq d + 1$ that minimizes $z(T)$. It can be decomposed as*



with the same M_k as in Theorem 1.

REFERENCES

- [1] M. Fischermann, I. Gutman, A. Hoffmann, D. Rautenbach, D. Vidović, and L. Volkmann. Extremal Chemical Trees. *Z. Naturforsch.*, 57a:49–52, 2002.
- [2] M. Fischermann, A. Hoffmann, D. Rautenbach, L. Székely, and L. Volkmann. Wiener index versus maximum degree in trees. *Discrete Appl. Math.*, 122(1-3):127–137, 2002.
- [3] I. Gutman and O. E. Polansky. *Mathematical concepts in organic chemistry*. Springer-Verlag, Berlin, 1986.
- [4] B. Hedman. Another extremal problem for Turán graphs. *Discrete Math.*, 65(2):173–176, 1987.
- [5] H. Hosoya. Topological index as a common tool for quantum chemistry, statistical mechanics, and graph theory. In *Mathematical and computational concepts in chemistry (Dubrovnik, 1985)*, Ellis Horwood Ser. Math. Appl., pages 110–123. Horwood, Chichester, 1986.
- [6] Y. Hou. On acyclic systems with minimal Hosoya index. *Discrete Appl. Math.*, 119(3):251–257, 2002.
- [7] F. Jelen and E. Triesch. Superdominance order and distance of trees with bounded maximum degree. *Discrete Appl. Math.*, 125(2-3):225–233, 2003.
- [8] P. Kirschenhofer, H. Prodinger, and R. F. Tichy. Fibonacci numbers of graphs. II. *Fibonacci Quart.*, 21(3):219–229, 1983.
- [9] P. Kirschenhofer, H. Prodinger, and R. F. Tichy. Fibonacci numbers of graphs. III. Planted plane trees. In *Fibonacci numbers and their applications (Patras, 1984)*, volume 28 of *Math. Appl.*, pages 105–120. Reidel, Dordrecht, 1986.
- [10] X. Li, Z. Li, and L. Wang. The inverse problems for some topological indices in combinatorial chemistry. *J. Comput. Biol.*, 10:47–55, 2003.
- [11] X. Li, H. Zhao, and I. Gutman. On the Merrifield-Simmons index of trees. *MATCH Commun. Math. Comput. Chem.*, 54(2):389–402, 2005.
- [12] S. B. Lin and C. Lin. Trees and forests with large and small independent indices. *Chinese J. Math.*, 23(3):199–210, 1995.
- [13] R. E. Merrifield and H. E. Simmons. *Topological Methods in Chemistry*. Wiley, New York, 1989.
- [14] H. Prodinger and R. F. Tichy. Fibonacci numbers of graphs. *Fibonacci Quart.*, 20(1):16–21, 1982.
- [15] B. E. Sagan. A note on independent sets in trees. *SIAM J. Discrete Math.*, 1(1):105–108, 1988.
- [16] L. A. Székely and H. Wang. On subtrees of trees. *Adv. in Appl. Math.*, 34(1):138–155, 2005.
- [17] L. A. Székely and H. Wang. Binary trees with the largest number of subtrees. *Discrete Appl. Math.*, 155(3):374–385, 2007.
- [18] R. F. Tichy and S. G. Wagner. Algorithmic generation of molecular graphs with large Merrifield-Simmons index. Preprint.
- [19] N. Trinajstić. *Chemical graph theory*. CRC Press, Boca Raton, FL., 1992.
- [20] S. Wagner. Extremal trees with respect to Hosoya Index and Merrifield-Simmons Index. *MATCH Commun. Math. Comput. Chem.*, 57(1):221–233, 2007.
- [21] H. S. Wilf. The number of maximal independent sets in a tree. *SIAM J. Algebraic Discrete Methods*, 7(1):125–130, 1986.

INSTITUT FÜR MATHEMATIK B, TECHNISCHE UNIVERSITÄT GRAZ, AUSTRIA
E-mail address: clemens.heuberger@tugraz.at

DEPARTMENT OF MATHEMATICAL SCIENCES, UNIVERSITY OF STELLENBOSCH, SOUTH AFRICA
E-mail address: swagner@sun.ac.za