Generation, Ranking and Unranking of Ordered Trees with Degree Bounds

Mahdi Amani
Dipartimento di Informatica
Università di Pisa, Pisa, Italy.
m_amani@di.unipi.it

Abbas Nowzari-Dalini
School of Mathematics, Statistics, and Computer Science
Colleague of Science, University of Tehran, Tehran, Iran.
nowzari@ut.ac.ir

We study the problem of generating, ranking and unranking of unlabeled ordered trees whose nodes have maximum degree of Δ. This class of trees represents a generalization of chemical trees. A chemical tree is an unlabeled tree in which no node has degree greater than 4. By allowing up to Δ children for each node of chemical tree instead of 4, we will have a generalization of chemical trees. Here, we introduce a new encoding over an alphabet of size 4 for representing unlabeled ordered trees with maximum degree of Δ. We use this encoding for generating these trees in A-order with constant average time and $O(n)$ worst case time. Due to the given encoding, with a precomputation of size and time $O(n^2)$ (assuming Δ is constant), both ranking and unranking algorithms are also designed taking $O(n)$ and $O(n\log n)$ time complexities.

1 Introduction

A *labeled tree* is a tree in which each node is given a unique label. A *rooted tree* is a tree in which one of the nodes is distinguished from the others as the *root*. An *ordered tree* or *plane tree* is a rooted tree for which an ordering is specified for the children of each node. Studying combinatorial properties of restricted graphs or graphs with configurations has many applications in various fields such as machine learning and chemoinformatics. Studying combinatorial properties of restricted trees and outerplanar graphs (*e.g.* ordered trees with bounded degrees) can be used for many purposes including virtual exploration of chemical universe, reconstruction of molecular structures from their signatures, and the inference of structures of chemical compounds [5, 11, 14, 15, 29, 12, 40]. In this paper we study the generation of unlabeled ordered trees whose nodes have maximum degree of Δ and for the sake of simplicity, we denote it by T^{Δ} *tree*, also we use T_n^{Δ} to denote the class of T^{Δ} trees with n nodes.

Chemical trees are the most similar trees to T^{Δ} trees. Chemical trees are the graph representations of *alkanes*, or more precisely, the carbon atom skeleton of the molecules of alkanes [8, 9, 13, 14, 20, 12]. The alkane molecular family is partitioned into *classes of homologous molecules*, that is molecules with the same numbers of carbonium and hydrogen atoms; the n^{th} class of alkane molecular family is characterized by the formula C_nH_{2n+2}, $n = 1,2,...$ [5] with the same numbers of carbonium and hydrogen atoms. They are usually represented by indicating the carbonium atoms and their links, omitting to represent hydrogen atoms [5]; therefore, all the nodes would have the same label; carbon (*i.e.*, the tree is unlabeled), as shown in Figure 1 for $n = 3$ and $n = 4$. A *chemical tree* is defined as a tree in which no node has degree greater than 4 [12, 14, 13, 9, 8, 20], chemical trees are also considered to be unlabeled [13, 9, 8, 20]. Therefore, T^{Δ} tree can be considered as a generalization of chemical trees to unlabeled ordered trees whose nodes have maximum degree of Δ instead of 4.

Generation, ranking and unranking of trees are very basic problems in computer science and discrete mathematics [39]. In general, for any combinatorial object of size n, we can define a variety of orderings.

C.A. Muñoz and J. A. Pérez (Eds.) :
Developments in Computational Models
EPTCS 204, 2016, pp. 31–45, doi:10.4204/EPTCS.204.4

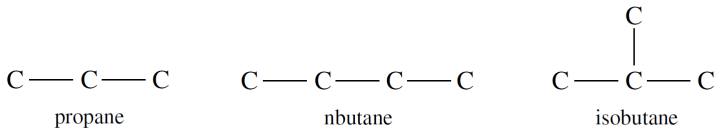

Figure 1: Left: C_3H_8 propane, middle and right: C_4H_{10} butanes.

Since we study trees, let's consider an arbitrary class of trees of size n (n nodes) showed by $\mathbb{T}_n$ (*e.g.* semi-chemical trees of size n), the elements of this set ($\mathbb{T}_n$) can be listed based on any defined ordering. Two such orderings for the trees are *A-order* and *B-order*, which will be explained in the next section. By having $\mathbb{T}_n$ and an ordering (*e.g.* A-order or B-order), *next*, *rank*, and *unrank* functions are defined as follows.

For a given tree $T \in \mathbb{T}_n$, the *next function* gives the successor tree of T with respect to the defined ordering, the 'position' of tree T in $\mathbb{T}_n$ is called *rank*, the *rank function* determines the rank of T; the inverse operation of ranking is *unranking*, for a position r, the *unrank function* gives the tree T corresponding to this position. A *generation algorithm*, generates all the elements in $\mathbb{T}_n$ with respect to the given ordering (starting from the first tree, then repeating the next function until producing the last tree) [39].

In most of the tree generation algorithms, a tree is represented by an integer or an alphabet sequence called *codeword*, hence all possible sequences of this representation are generated. This operation is called *tree encoding*. Basically, the uniqueness of the encoding, the length of the encoding, and the capability of constructing the tree from its representation, which is called *decoding*, are essential considerations in the design of the tree encoding schema [39].

Many papers have been published earlier in the literature for generating different classes of trees. For example we can mention the generation of binary trees in [26, 22, 30, 2, 3, 4, 36], k-ary trees in [27, 10, 31, 19, 37, 18, 17, 1, 35], rooted trees in [7, 32], trees with n nodes and m leaves in [24, 28], neuronal trees in [25], and AVL trees in [21]. On the other hand, many papers have thoroughly investigated basic combinatorial features of chemical trees [12, 14, 13, 9, 8, 20, 33].

More related to our work, in [15] a coding for chemical trees without the generation algorithm, and in [5] the enumeration of chemical trees and in [11, 29] the enumeration of tree-like chemical graphs have been presented. Hendrickson and Parks in [16] investigated the enumeration and the generation of carbon skeletons which can have cycles and are not necessarily trees. The work most related to our paper is an algorithm for the generation of certain classes of trees such as chemical trees in [6] with no ranking or unranking algorithm. In that paper, all chemical trees with n nodes are generated from the complete set of chemical trees with $n-1$ nodes, the redundant generations are possible and they needed to minimize the possible redundancy.

The problem of enumeration of ordered trees were also studied in [38] and the generation of different ordered trees (with no bounds on the degrees of the nodes) were studied in [39]. In [40], a generation algorithm with constant average delay time but with no ranking or unranking algorithms was given for all unrooted trees of n nodes and a diameter at least d such that the degree of each node with distance k from the center of the tree is bounded by a given function. In [34] all unrooted unlabeled trees have been generated in constant average time with no ranking or unranking algorithms. Nakano and Uno in [23] gave an algorithm to generate all rooted unordered trees with exactly n nodes and diameter d in constant delay time. Up to now, to our knowledge, no efficient generation, ranking or unranking algorithms are known for either 'chemical trees' or 'ordered trees with bounded degrees'.

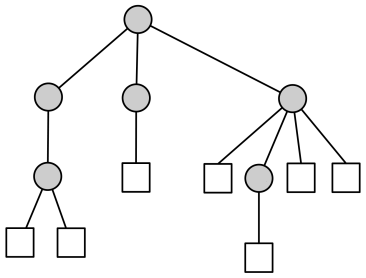

Figure 2: A T^{Δ} tree with 13 nodes.

The remaining of the paper is organized as follows. Section 2 introduces the definitions and notions that are used further. Our new encoding for T_n^{Δ} trees is presented in Section 3. The size of our encoding is n while the alphabet size is always 4. Based on the presented encoding, a new generation algorithm with constant average time and $O(n)$ worst case time is given Section 4. In this algorithm, T_n^{Δ} trees are generated in A-order. Ranking and unranking algorithms are also designed in Section 5 with $O(n)$ and $O(n\log n)$ time complexities, respectively. The presented ranking and unranking algorithms need a precomputation of size and time $O(n^2)$ (assuming Δ is constant).

2 Definition

As mentioned before, the n^{th} class of alkane molecular family is characterized by the formula C_nH_{2n+2}, $n = 1,2,...$ [5] with the same numbers of carbonium and hydrogen atoms. They are usually represented by indicating the carbonium atoms and their links, omitting to represent hydrogen atoms. Therefore, a chemical tree is a tree in which no node has degree greater than 4 [12, 8, 20]. We study the class of ordered trees whose nodes have maximum degree of Δ and for the sake of simplicity, we denote it by T^{Δ}. T^{Δ} can be considered as a generalized version of chemical trees.

Formally, a T^{Δ} tree T is defined as a finite set of nodes such that T has a root r, and if T has more than one node, r is connected to $j \leq \Delta$ subtrees $T_1, T_2, \ldots, T_j$ which each one of them is also recursively a T^{Δ} tree and by T_n^{Δ} we represent the class of T^{Δ} trees with n nodes. An example of a T^{Δ} tree is shown in Figure 2.

As mentioned earlier, in most of the tree generation algorithms, a tree is represented by an integer or an alphabet sequence called *codeword*, hence, all possible sequences of this representation are generated. In general, on any class of trees, we can define a variety of ordering for the set of trees. Classical orderings on trees are *A-order* and *B-order* which are defined as follows [30, 39].

Definition 1 *Let T and T' be two trees in T^{Δ} and $k = max\{deg(T), deg(T')\}$, we say that T is less than T' in A-order ($T \prec_A T'$), iff*

- $|T| < |T'|$, *or*
- $|T| = |T'|$ *and for some* $1 \leq i \leq k$, $T_j =_A T'_j$ *for all* $j = 1,2,\ldots,i-1$ *and* $T_i \prec_A T'_i$;

where $|T|$ is the number of nodes in the tree T and $deg(T)$ is the degree of the root of T.

Definition 2 *Let T and T' be two trees in T^{Δ} and $k = max\{deg(T), deg(T')\}$, we say that T is less than T' in B-order ($T \prec_B T'$), iff*

- $deg(T) < deg(T')$, *or*

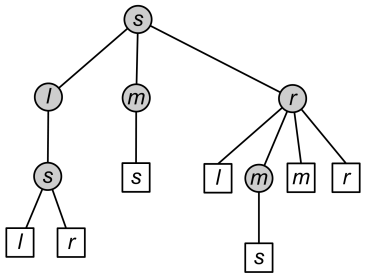

Figure 3: An example of a tree $T \in T_n^\Delta$ (for $\Delta \geq 4$). Its codeword is "$s\ell s\ell rmsr\ell msmr$".

- *$deg(T) = deg(T')$ and for some $1 \leq i \leq k$, $T_j =_B T'_j$ for all $j = 1,2,\ldots,i-1$ and $T_i \prec_B T'_i$.*

Our generation algorithm, which is given in the Section 4, produces the sequences corresponding to T_n^Δ trees in A-order. For a given tree $T \in T_n^\Delta$, the *generation algorithm* generates all the successor trees of T in T_n^Δ; the position of tree T in T_n^Δ is called *rank*, the *rank function* determines the rank of T; the inverse operation of ranking is *unranking*. These functions can be also employed in any random generation of T_n^Δ trees, for example.

3 The encoding schema

The main point in generating trees is to choose a suitable encoding to represent them, and generate their corresponding codewords. Regarding the properties of T_n^Δ, we present our new encoding. For any tree $T \in T_n^\Delta$, the encoding over 4 letters $\{s,\ell,m,r\}$ is defined as follows. The root of T is labeled by s, and for any internal node, if it has only one child, that child is labeled by s, otherwise the leftmost child is labeled by ℓ, and the rightmost child is labeled by r, and the children between the leftmost and the rightmost children (if exist) are all labeled by m. Nodes are labeled in the same way for any internal node in each level recursively, and by a pre-order traversal of T, the codeword will be obtained (one can say the labels ℓ and r operate as left and right parenthesis to define the fingerprint of a subtree in the codewords, and s is used when the left and right bounds are the same). This labeling is illustrated in Figure 3. Using this encoding, the 4-letters alphabet codeword corresponding to the first and last T_n^Δ trees in A-order are respectively "$s\ell m^{\Delta-2}r\ell m^{\Delta-2}r\ldots\ell m^{(n \bmod \Delta)-2}r$" and "$s^n$" which are shown in the Figure 4-a and Figure 4-b. Now, we prove the validity of this encoding for T_n^Δ trees (one-to-one correspondence).

Definition 3 *Suppose that $\{s,\ell,m,r\}^*$ is the set of all sequences with alphabet of s,m,ℓ,r and let A be a proper subset of $\{s,\ell,m,r\}^*$, then we call the set A a* CodeSet$^\Delta$ *iff A satisfies the following properties:*

1. $\varepsilon \in A$ *(ε is a string of length 0),*
2. $\forall x \in A : sx \in A$*,*
3. $\forall\, x_1, x_2, \ldots, x_i \in A$*, and* $2 \leq i \leq \Delta$*:* $\ell x_1 m x_2 m x_3 \ldots m x_{i-1} r x_i \in A$.

Now we show that a **valid codeword** is obtained by the concatenation of the character s and each element of CodeSet$^\Delta$.

Theorem 1 *Let A be the "CodeSet$^\Delta$" and $\delta \in A$. Let C be a codeword obtained by the concatenation of character s and δ (denoted by $s\delta$). There is a one-to-one correspondence between C and a T^Δ tree.*

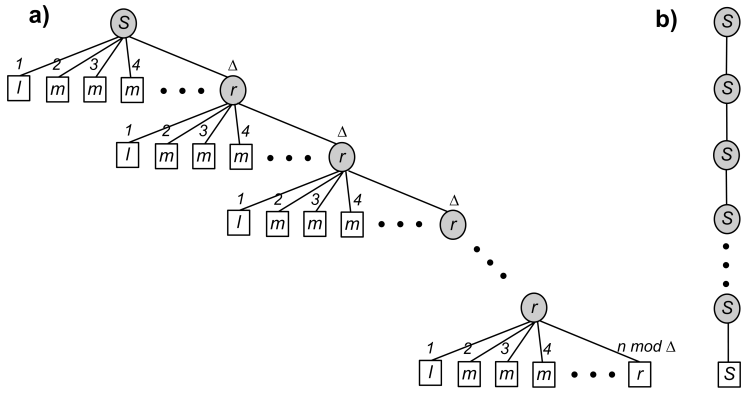

Figure 4: a) The first T_n^Δ tree in A-order. b) The last T_n^Δ tree in A-order.

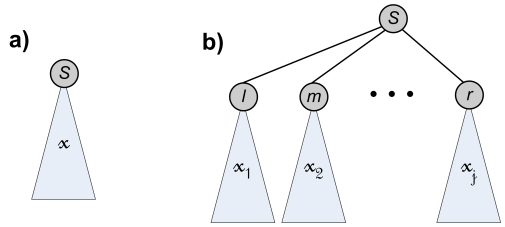

Figure 5: T^Δ trees encoded by $C = sx$ and $C = s\ell x_1 m x_2 \dots m x_{j-1} r x_j$.

Proof. It can be proved by induction on the length of C. Initially for a codeword of length equal to 1, the proof is trivial. Assume that any codeword obtained in the above manner with length less than n encodes a unique T^Δ tree. For a given codeword with length n, because of that concatenation of s and δ, we have:

1. $C = sx$, such that $x \in A$, or
2. $C = s\ell x_1 m x_2 \dots m x_{j-1} r x_j$, such that $x_i \in A$, $\forall 1 \leq i \leq j \leq \Delta$.

For the first case by induction hypothesis, x is a valid codeword of a T^Δ tree T; therefore, sx is another codeword corresponding to another T^Δ tree by adding a new root to the top of T. This tree is unique and shown in Figure 5-a. For the second case, by induction hypothesis and that concatenation of s and δ, each sx_i for $1 \leq i \leq j$ is a valid codeword for a T^Δ tree; therefore, with replacement of 's with ℓ in sx_1' and 's with m in sx_i for $2 \leq i \leq j-1$' and finally 's with r in sx_j', we can produce $\ell x_1, m x_2, \dots m x_{j-1}, r x_j$ codewords. Now they all are subtrees of a T^Δ tree whose codeword is $C = s\ell x_1 m x_2 \dots m x_{j-1} r x_j$ (add a new root and connect it to each one of them). This tree is unique and shown in Figure 5-b. ■

For a T_n^Δ tree, this encoding needs only 4 alphabet letters and has length n. This encoding is simple and powerful, so it can be used for many other applications besides the generation algorithm. In the next section, we use it to generate T^Δ trees in A-order.

4 The generation algorithm

In this section, we present an algorithm that generates the successor sequence of a given codeword of a T_n^Δ tree in A-order. For generating the successor of a given codeword C corresponding to a T_n^Δ tree T,

the codeword C is scanned from right to left. Scanning the codeword C from right to left, corresponds to a reverse pre-order traversal of T. First we describe how this algorithm works directly on T, then we present the pseudocode of the algorithm. For generating the successor of a given T_n^Δ tree T we traverse the tree in reverse pre-order as follows.

1. Let v be the last node of T in pre-order traversal.
2. If v doesn't have any brothers, then
   - repeat $\{v =$ parent of $v.\}$
     until v has at least one brother or v be the root of tree T.
   - If $v =$ root, then the tree is the last tree in A-order and there is no successor.
3. If v has at least one brother (obviously it has to be a left brother), delete one node from the subtree of v and insert this node into its left brother's subtree, then rebuild both subtrees (each one as a first tree with corresponding nodes in A-order).

The pseudo code of this algorithm for codewords corresponding to T_n^Δ trees is presented in Figure 6. In this algorithm, the codeword is stored in a global array C so it is used both as the input and the output codewords (this will be useful to generate all the codewords in constant average time); therefore, when the algorithms stops, C is updated to the successor codeword. In this algorithm, n shows the size of the codeword (the number of nodes of the tree corresponded to C), $STsize$ is a variable contains the size of the subtree rooted by node corresponded to $C[i]$ and $SNum$ holds the number of consecutive visited s characters. This algorithm also calls two functions $updateChildren(i,ChNum)$ presented in Figure 7, and $updateBrothers(i,ChNum)$ presented in Figure 8.

The procedure $updateChildren(i,ChNum)$ regenerates the codeword corresponding to the children of an updated node and the procedure $updateBrothers(i,ChNum)$ regenerates the codeword corresponding to the brothers of a node with regard to the maximum degree Δ for each node. In these algorithms, C is the global array of characters storing the codeword, i is the position of the current node in the array C. In $updateChildren(i,ChNum)$, $ChNum$ is the number of children of $C[i]$ to regenerate the corresponding codeword and in $updateBrothers(i,ChNum)$, $ChNum$ is the number of its brothers, instead. Each node has at most Δ children, so $NChild$ is a global array which $NChild[i]$ stores the number of children of the parent of the current node (current node is the node corresponding to $C[i]$) so it wont exceed Δ, *i.e.*, $NChild[i]$ holds the number of left brothers of the node corresponding to $C[i]$ including itself.

Theorem 2 *The algorithm Next presented in Figure 6 has a worst case time complexity of $O(n)$ and an average time complexity of $O(1)$.*

Proof. In the algorithm given in Figure 6, we scan the sequence from right to left (corresponding to the reverse pre-order traversal of the corresponding tree as discussed before). Inside or after every loop of the algorithm, the variable *Current* (which keeps track of the node we currently process), decreases equivalently to the number of iterations, and it can not be decreased more than n times; therefore, the worst case time complexity of the algorithm is $O(n)$. For computing the average time, it should be noted that during the scanning process, every time we visit the characters s, m or ℓ, the algorithm terminates, so we define $S_i^{n,\Delta}$ as the number of codewords of T_n^Δ trees whose last character s, m or ℓ has distance i from the end, and $S^{n,\Delta}$ as the total number of T_n^Δ trees. Obviously we have:

$$S^{n,\Delta} = \sum_{i=1}^{n} S_i^{n,\Delta}. \tag{1}$$

We define H_n as the average time of generating all codewords of T_n^Δ trees,

```
Function AOrder-Next(n : integer);
var i,Current,STsize,SNum: integer; finished,RDeleted: boolean;
begin
    Current := n; STSize := 0; RDeleted := false; finished := false;
    while ( (C[Current] = 's') & (Current ≥ 1) ) do
        STSize++; Current--;
    if (Current = 0) then return (no successor);
    while (not finished) do begin
        STSize++;
        switch C[Current] of
            case 'r':
                i := Current − 1; SNum := 0;
                while (C[i] = 's') do
                    SNum := SNum + 1; i−−;
                if (C[i] = 'r') then begin
                    updateBrothers ( Current , STSize);
                    Current := i; STSize := SNum;
                end;
                if ( (C[i] = 'm') or (C[i] = 'ℓ') ) then begin
                    if (STSize = 1) then RDeleted := true;
                    if (STSize > 1) then begin
                        STSize−−; updateBrothers(Current + 1,STSize);
                        Current := i; STSize := SNum + 1;
                    end;
                end;
            case 'm':
                if (RDeleted = true) then C[Current] := 'r';
                updateChildren( Current + 1,STSize − 1); finished:= true;
            case 'ℓ':
                if (RDeleted = true) then C[Current] := 's';
                updateChildren( Current + 1,STSize − 1); finished := true;
    end;
end;
```

Figure 6: Algorithm for generating the successor codeword for T_n^{Δ} trees in A-order.

$$H_n \leq (k/S^{n,\Delta}) \sum_{i=1}^{n} iS_i^{n,\Delta},$$
$$\leq (k/S^{n,\Delta}) \sum_{j=1}^{n} \sum_{i=j}^{n} S_i^{n,\Delta}.$$

Where k is a constant value. On the other hand, consider that for $S_j^{n+1,\Delta}$ we have two cases, in the first case, the last character s, m or ℓ is a leaf and in the second one, it is not. Therefore, $S_j^{n+1,\Delta}$ is greater than or equal to just the first case, and in that case by removing the node corresponding to the 'last character s, m or ℓ of the codeword', the remaining tree will have a corresponding codeword belongs to exactly one of $S_k^{n,\Delta}$ cases, for $j \leq k \leq n$. By substituting k and i we have:

$$S_j^{n+1,\Delta} \geq \sum_{i=j}^{n} S_i^{n,\Delta}.$$

Therefore, for H_n we have:

```
procedure updateChildren( i,ChNum: integer);
begin
    while (ChNum > 0) do begin
        if ChNum = 1 then begin
            C[i] := 's'; NChild[i] := 1; i++; ChNum--;
        end;
        if ChNum > 1 then begin
            C[i] := 'ℓ'; NChild[i] := 1; i++; ChNum--;
            while ( (NChild[i] < (Δ-1)) & (ChNum > 1) ) do begin
                C[i] := 'm'; NChild[i] := NChild[i-1]+1; i++; ChNum--;
            end;
            C[i] := 'r'; NChild[i] := NChild[i-1]+1; i++; ChNum--;
        end
    end;
end;
```

Figure 7: Algorithm for updating the children.

```
Procedure updateBrothers( i,ChNum: integer);
begin
    if ChNum = 1 then begin
        C[i] := 'r'; NChild[i] := NChild[i-1]; ChNum--;
    end;
    if ChNum > 1 then begin
        C[i] := 'm'; ChNum--; i++;
        while ( (NChild[i] < (Δ-1) ) & (ChNum > 1) ) do begin
            C[i] := 'm'; NChild[i] := NChild[i-1]+1; i++; ChNum--;
        end;
        C[i] := 'r'; NChild[i] := NChild[i-1]+1;
        i++; ChNum--; updateChildren(i,ChNum);
    end;
end;
```

Figure 8: Algorithm for updating the neighbors.

$$H_n \leq (k/S^{n,\Delta}) \sum_{j=1}^{n} S_j^{n+1,\Delta},$$

then by using Equation (1),

$$H_n \leq kS^{n+1,\Delta}/S^{n,\Delta}.$$

Finally from [39] we know that the total number of ordered trees is growing same as Catalan number, while T_n^{Δ} is a subset of ordered trees can not grow faster than that, this guarantees that for large enough values of n, $S^{n+1,\Delta}/S^{n,\Delta} = O(1)$. Therefore, $H_n \leq kO(1) = O(1)$. ■

It should be mentioned that this constant average time complexity is without considering the input or the output time.

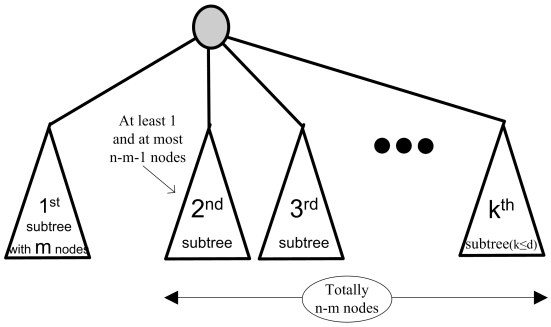

Figure 9: T_n^{Δ} tree whose first subtree has exactly m nodes and its root has maximum degree of d.

5 Ranking and Unranking algorithms

By designing a generation algorithm in a specific order, the ranking of algorithm is desired. In this section, ranking and unranking algorithms for these trees in A-order will be given. Ranking and unranking algorithms usually use a precomputed table of the number of a subclass of given trees with some specified properties to achieve efficient time complexities; these precomputations will be done only once and stored in a table for further use. Let $S^{n,\Delta}$ be the number of T_n^{Δ} trees, $S_{m,d}^{n,\Delta}$ be the number of T_n^{Δ} trees whose first subtree has **exactly** m nodes and its root has maximum degree of d, and $D_{m,d}^{n,\Delta}$ be the number of T_n^{Δ} trees whose first subtree has **at most** m nodes and its root has maximum degree of d.

Theorem 3

- $D_{m,d}^{n,\Delta} = \sum_{i=1}^{m} S_{i,d}^{n,\Delta}$,
- $S^{n,\Delta} = \sum_{i=1}^{n-1} S_{i,\Delta}^{n,\Delta}$.

Proof. The proof is trivial. ■

Theorem 4

$$S_{m,d}^{n,\Delta} = S_{m,1}^{m+1,\Delta} \times \sum_{i=1}^{n-m-1} (S_{i,d-1}^{n-m,\Delta}).$$

Proof. Let T be a T_n^{Δ} tree whose first subtree has exactly m nodes and its root has maximum degree of d; by the definition and as shown in the Figure 9, the number of the possible cases for the first subtree is $S_{m,1}^{m+1,\Delta}$ and the number of cases for the other parts of the tree is $\sum_{i=1}^{n-m-1}(S_{i,d-1}^{n-m,\Delta})$ (by removing the first subtree from T). Therefore, we have:

$$S_{m,d}^{n,\Delta} = S_{m,1}^{m+1,\Delta} \times \sum_{i=1}^{n-m-1} (S_{i,d-1}^{n-m,\Delta}).$$

■

Now, let T be a T_n^{Δ} tree whose subtrees are defined by $T_1, T_2, \ldots, T_k$ and for $1 \leq i \leq k \leq \Delta : |T_i| = n_i$ and $\sum_{i=1}^{k} n_i = n-1$. One way to compute the rank of tree T is to enumerate the number of trees generated

before T. Let *Rank(T,n)* be the rank of T. The number of T^{Δ} trees whose first subtree is smaller than T_1 is equal to:

$$\sum_{i=1}^{n_1-1} S_{i,\Delta}^{n,\Delta} + (Rank(T_1,n_1)-1) \times \sum_{i=1}^{n-n_1} S_{i,\Delta-1}^{n-n_1,\Delta},$$

and the number of T^{Δ} trees whose first subtree is equal to T_1 but the second subtree is smaller than T_2 is equal to:

$$\sum_{i=1}^{n_2-1} S_{i,\Delta-1}^{n-n_1,\Delta} + (Rank(T_2,n_2)-1) \times \sum_{i=1}^{n-n_1-n_2} S_{i,\Delta-2}^{n-n_1-n_2,\Delta}.$$

Similarly, the number of T^{Δ} trees whose first $(j-1)$ subtrees are equal to $T_1, T_2, \ldots, T_{j-1}$ and the j^{th} subtree is smaller than T_j is equal to:

$$\sum_{i=1}^{n_j-1} S_{i,\Delta-j+1}^{(n-\sum_{\ell=1}^{j-1} n_\ell),\Delta} + (Rank(T_j,n_j)-1) \times \sum_{i=1}^{n-\sum_{\ell=1}^{j} n_\ell} S_{i,\Delta-j}^{n-\sum_{\ell=1}^{j} n_\ell,\Delta}.$$

Therefore, regarding enumerations explained above, for given tree $T \in T_n^{\Delta}$ whose subtrees are defined by $T_1, T_2, \ldots, T_k$, we can write:

$$\begin{aligned} Rank(T,1) &= 1, \\ Rank(T,n) &= 1 + \sum_{j=1}^{k} \Big(\sum_{i=1}^{n_j-1} S_{i,\Delta-j+1}^{(n-\sum_{\ell=1}^{j-1} n_\ell),\Delta} + (Rank(T_j,n_j)-1) \sum_{i=1}^{n-\sum_{\ell=1}^{j} n_\ell} S_{i,\Delta-j}^{(n-\sum_{\ell=1}^{j} n_\ell),\Delta} \Big). \end{aligned}$$

Hence, from Theorem 3, by using $D_{m,d}^{n,\Delta} = \sum_{i=1}^{m} S_{i,d}^{n,\Delta}$, we have:

$$\begin{aligned} Rank(T,1) &= 1, \\ Rank(T,n) &= 1 + \sum_{j=1}^{k} \Big(D_{(n_j-1),(\Delta-j+1)}^{(n-\sum_{\ell=1}^{j-1} n_\ell),\Delta} + (Rank(T_j,n_j)-1) D_{(n-\sum_{\ell=1}^{j} n_\ell),(\Delta-j)}^{(n-\sum_{\ell=1}^{j} n_\ell),\Delta} \Big). \end{aligned}$$

To achieve the most efficient time for ranking and unranking algorithms, we need to precompute $D_{m,d}^{n,\Delta}$ and store it for further use. Assuming Δ is constant, to store $D_{m,d}^{n,\Delta}$ values, a 3-dimensional table denoted by $D[n,m,d]$ is enough, this table will have a size of $O(n \times n \times \Delta) = O(n^2)$ and can be computed using Theorems 3 and 4 with time complexity of $O(n \times n \times \Delta) = O(n^2)$.

To compute the rank of a codeword stored in array C, we also need an auxiliary array $N[i]$ which keeps the number of nodes in the subtree whose root is labeled by $C[i]$ and corresponds to n_i in the above formula. This array can be computed by a pre-order traversal or a level first search (DFS) algorithm just once before we call the ranking algorithm.

The pseudo code for ranking algorithm is given in Figure 10. In this algorithm, *Beg* is the variable that shows the positions of the first character in the array C whose rank is being computed (*Beg* is initially set to 1), and Fin is the variable that returns the position of the last character of C.

Now the time complexity of this algorithm is discussed. Obviously computing the array $N[i]$ takes $O(n)$. Hence we discuss the complexity of ranking algorithm which was given in Figure 10.

Theorem 5 *The ranking algorithm has the time complexity of* $O(n)$.

```
Function Rank( Beg : integer; var Fin: integer) ;
Var R, Point, PointFin, j, Nodes, n: integer;
begin
    n := N[Beg];
    if (n = 1) then begin
        Fin := Beg; return(1) end;
    else begin
        Point := Beg + 1; R := 0; Nodes := 0; j := 1;
        while ( Nodes < n ) do begin
            R := R + D[n − Nodes, N[Point] − 1, Δ − j + 1] +
                (Rank(Point, PointFin) − 1) ×
                D[(n − Nodes − N[Point]), (n − Nodes − N[Point]), Δ − j];
            Nodes := Nodes + N[Point]; j:=j+1;
            Point := PointFin + 1;
        end;
        Fin := Point − 1;
        return( R + 1);
    end;
end
```

Figure 10: Ranking algorithm for T_n^Δ trees.

Proof. Let T be a T_n^Δ tree whose subtrees are defined by $T_1, T_2, \ldots, T_k$ and for $1 \le i \le k \le \Delta$: $|T_i| = n_i$ and $\sum_{i=1}^{k} n_i = n-1$, and let $T(n)$ be the time complexity of the ranking algorithm, then we can write:

$$T(n) = T(n_1) + T(n_2) + \ldots + T(n_k) + \alpha k,$$

where α is a constant and αk is the time complexity of the non-recursive parts of the algorithm. By using a simple induction, we prove that if β is a value greater than α then $T(n) \le \beta n$. We have $T(1) \le \beta$. We assume $T(m) \le \beta(m-1)$ for each $m < n$; therefore,

$$\begin{aligned} &T(n) \le \beta(n_1 - 1) + \beta(n_2 - 1) + \ldots + \beta(n_k - 1) + \alpha k, \\ &T(n) \le \beta(n_1 + \ldots + n_k - k) + \alpha k, \\ &T(n) \le \beta n - \beta k + \alpha k \le \beta n, \end{aligned}$$

So the induction is complete and $T(n) \le \beta n = O(n)$. ■

If a and b are integer numbers, let $(a \; div \; b)$ and $(a \; mod \; b)$ denote *integer division* and *remainder* of the division of a and b, respectively ($a = (a \; div \; b) \times b + (a \; mod \; b)$). Before giving the description of the unranking algorithm, we have to define two new operators, namely $(a \; div^+ \; b)$ and $(a \; mod^+ \; b)$ as follows.

- If $b \mid a$, then $(a \; div^+ \; b) = (a \; div \; b) - 1$, and $(a \; mod^+ \; b) = b$.
- If $b \nmid a$, then $(a \; div^+ \; b) = (a \; div \; b)$, and $(a \; mod^+ \; b) = (a \; mod \; b)$.

For unranking algorithm, we also need the values of $S^{n,\Delta}$, these values can be stored in an array of size n, denoted by $S[n]$ (we assume Δ is constant). The unranking algorithm is the reverse manner of ranking algorithm, this algorithm is given in Figure 11. In this algorithm, the rank R is the main input, *Beg* is

```
Function UnRank ( R,Beg,n: integer; Root: char);
var Point, i, t, ChildNum: integer;
begin
    if ( (n = 0) or (R = 0) ) then return(Beg − 1)
    else begin
        if (n = 1) then begin
            C[Beg] := Root; return(Beg);
        end;
        else begin
            C[Beg] := Root; Point := Beg + 1;
            Root := 'ℓ'; ChildNum := 0;
            while (n > 0) do begin
                ChildNum + +;
                find the smallest i that D[n,i,Δ − ChildNum + 1] ≥ R;
                R := R − D[n,i − 1,Δ − ChildNum + 1];
                if (n − i)= 1 then
                    if (ChildNum = 1) then  Root := 's';
                    else  Root := 'r';
                t := S[n];
                Point := UnRank( (div+(R,t)) + 1,Point,i,Root ) +1;
                R := mod+(R,t);
                n := n − i; Root := 'm';
            end;
            return(Point − 1);
        end;
    end;
end
```

Figure 11: Unranking algorithm for T_n^Δ trees.

the variable to show the position of the first character in the global array C and initially is set to 1. The generated codeword will be stored in global array C. The variable n is the number of nodes and *Root* stores the character corresponding to the node we consider for the unranking procedure. For the next character, we have two possibilities. If the root is r or s, then the next character, if exists, will be ℓ or s (based on the number of root's children). If the root is m or ℓ, we have again two possible cases: if all the nodes of the current tree are not produced, then the next character is m, otherwise, the next character will be r.

Theorem 6 *The time complexity of the unranking algorithm is* $O(n\log n)$.

Proof. Let T be a T_n^Δ tree whose subtrees are defined by T_1, T_2, ..., T_k and for $1 \le i \le k \le \Delta$: $|T_i| = n_i$ and $\sum_{i=1}^{k} n_i = n-1$, and let $T(n)$ be the time complexity of the unranking algorithm. With regards to the unranking algorithm, the time complexity of finding j such that $D[n, j, \Delta - ChildNum + 1] \ge R$ for each T_i of T is $O(\log n_i)$; therefore,

$$T(n) = O(\log n_1 + \log n_2 + \ldots + \log n_k) + T(n_1) + T(n_2) + \ldots + T(n_k).$$

We want to prove that $T(n) = O(n\log n)$. In order to obtain an upper bound for $T(n)$ we do as

follows. First we prove this assumption for $k = 2$ then we generalize it. For $k = 2$ we have $T(n) = O(\log(n_1) + \log(n_2)) + T(n_1) + T(n_2)$. Let $n_1 = x$ then we can write the above formula as

$$T(n) = T(x) + T(n-x) + O(\log(x) + \log(n-x)) = T(x) + T(n-x) + C' \log(n).$$

For proving that $T(n) = O(n\log(n))$, we use an induction on n. We assume $T(m) \leq Cm\log(m)$ for all $m \leq n$, thus in $T(n)$ we can substitute

$$T(n) \leq C \times x\log(x) + C \times (n-x)\log(n-x) + C' \log(n).$$

Let $f(x) = C \times x\log(x) + C \times (n-x)\log(n-x)$, now the maximum value of $f(x)$ with respect to x and by considering n as a constant value can be obtained by evaluating the derivation of $f(x)$ which is $f'(x) = C \times \log(x) - C \times \log(n-x)$. Thus if $f'(x) = 0$ we get $x = (n-1)/2$ and by computing $f(1)$, $f(n-2)$ and $f((n-1)/2)$ we have:

$$f(1) = f(n-2) = C \times (n-2)\log(n-2),$$
$$f((n-1)/2) = 2C \times ((n-1)/2) \times \log((n-1)/2) \ < C \times (n-2)\log(n-2),$$

so the maximum value of $f(x)$ is equal to $C \times (n-2)\log(n-2)$; therefore,

$$T(n) \leq C \times (n-2)\log(n-2) + C' \times \log(n).$$

It is enough to assume $C = C'$, then $T(n) \leq C \times (n-2)\log(n) + C \times \log(n) \ \leq C \times n\log(n)$.

Now, for generalizing the above proof and proving $T(n) = O(n\log n)$, we should find the maximum of the function $f(n_1, n_2, \ldots, n_k) = \prod_{i=1}^{k} n_i$. By the Lagrange method we prove that the maximum value of $f(n_1, n_2, \ldots, n_k)$ is equal to $(\frac{n}{k})^k$. Then $\frac{\delta f}{\delta k} = (\frac{n}{k})^k(\ln(\frac{n}{k}) - 1) = 0$, and

$$\ln(\tfrac{n}{k}) - 1 = 0,$$
$$\tfrac{n}{k} = e \Rightarrow k = \tfrac{n}{e},$$

so the maximum value of $f(n_1, n_2, \ldots, n_k)$ is equal to $e^{\frac{n}{e}}$. We know that:

$$T(n) = O(\log n_1 + \log n_2 + \ldots + \log n_k) + T(n_1) + T(n_2) + \ldots + T(n_k),$$
$$T(n) = O(\log(\textstyle\prod_{i=1}^{k} n_i) + \sum_{i=1}^{k} T(n_i),$$
$$T(n) < O(\log(n^{\frac{n}{e}})) + \textstyle\sum_{i=1}^{k} T(n_i),$$
$$T(n) < O(\tfrac{n}{e}\log e) = O(n) + \textstyle\sum_{i=1}^{k} T(n_i).$$

Finally, by using induction, we assume that for any $m < n$ we have $T(m) < \beta m \log m$, therefore:

$$T(n) = O(n) + \textstyle\sum_{i=1}^{k} T(n_i),$$
$$T(n) < O(n) + \textstyle\sum_{i=1}^{k} \beta O(n_i \log n_i),$$
$$T(n) < O(n) + \beta \log(\textstyle\prod_{i=1}^{k}(n_i^{n_i})),$$
$$T(n) < O(n) + O(\log(n^n)),$$
$$T(n) = O(n\log n).$$

Hence, the proof is complete. ■

6 Conclusion

In this paper, we studied the problem of generation, ranking and unranking of ordered trees of size n and maximum degree Δ which are a generalization of chemical trees; we presented an efficient algorithm for the generation of these trees in A-order with an encoding over 4 letters and size n. Also two efficient ranking and unranking algorithms were designed for this encoding. The generation algorithm has $O(n)$ time complexity in worst case and $O(1)$ in average case. The ranking and unranking algorithms have $O(n)$ and $O(n\log n)$ time complexity, respectively. The presented ranking and unranking algorithms use a precomputed table of size $O(n^2)$ (assuming Δ is constant). All presented algorithms at this paper are also implemented. For the future works, generating this class of trees in B-order and minimal change ordering, and finding some explicit relations for counting them, are major unresolved problems.

References

[1] A. Ahmadi-Adl, A. Nowzari-Dalini & H. Ahrabian (2011): *Ranking and unranking algorithms for loopless generation of t-ary trees*. Logic Journal of IGPL 19(1), pp. 33–43, doi:10.1093/jigpal/jzp097.

[2] H. Ahrabian & A. Nowzari-Dalini (1998): *On the generation of binary trees from (0–1) codes*. International journal of computer mathematics 69(3-4), pp. 243–251, doi:10.1080/00207169808804721.

[3] H. Ahrabian & A. Nowzari-Dalini (1999): *On the generation of binary trees in A-order*. International journal of computer mathematics 71(3), pp. 351–357, doi:10.1080/00207169908804813.

[4] H. Ahrabian & A. Nowzari-Dalini (2005): *Parallel generation of binary trees in A-order*. Parallel Computing 31(8), pp. 948–955, doi:10.1016/j.parco.2005.06.002.

[5] R. Aringhieri, P. Hansen & F. Malucelli (2003): *Chemical trees enumeration algorithms*. Quarterly Journal of the Belgian, French and Italian Operations Research Societies 1(1), pp. 67–83, doi:10.1007/s10288-002-0008-9.

[6] T. S. Balaban, P. A. Filip & O. Ivanciuc (1992): *Computer generation of acyclic graphs based on local vertex invariants and topological indices. Derived canonical labelling and coding of trees and alkanes*. Journal of Mathematical Chemistry 11(1), pp. 79–105, doi:10.1007/BF01164196.

[7] T. Beyer & S. M. Hedetniemi (1980): *Constant time generation of rooted trees*. SIAM Journal on Computing 9(4), pp. 706–712, doi:10.1137/0209055.

[8] G. Caporossi, I. Gutman & P. Hansen (1999): *Variable Neighborhood Search for Extremal Graphs: IV: Chemical Trees with Extremal Connectivity Index*. Computers & Chemistry 23(5), pp. 469–477, doi:10.1016/S0097-8485(99)00031-5.

[9] A. A. Dobrynin & I. Gutman (1999): *The Average Wiener Index of Trees and Chemical Trees*. Journal of Chemical Information and Computer Sciences 39(4), pp. 679–683, doi:10.1021/ci980158r.

[10] M. C. Er (1992): *Efficient generation of k-ary trees in natural order*. The Computer Journal 35(3), pp. 306–308, doi:10.1093/comjnl/35.3.306.

[11] H. Fujiwara, J. Wang, L. Zhao, H. Nagamochi & T. Akutsu (2008): *Enumerating Treelike Chemical Graphs with Given Path Frequency*. Journal of Chemical Information and Modeling 48(7), pp. 1345–1357, doi:10.1021/ci700385a.

[12] I. Gutman (1987): *Graphs and graph polynomials of interest in chemistry*, pp. 177–187. Springer Berlin Heidelberg, doi:10.1007/3-540-17218-1.

[13] I. Gutman, P. Hansen & H. Mélot (2005): *Variable Neighborhood Search for Extremal Graphs. 10. Comparison of Irregularity Indices for Chemical Trees*. Journal of Chemical Information and Modeling 45(2), pp. 222–230, doi:10.1021/ci0342775.

[14] I. Gutman & O. E. Polansky (1986): *Mathematical concepts in organic chemistry*. Springer Berlin Heidelberg, doi:10.1007/978-3-642-70982-1.

[15] P. Hansen, B. Jaumard, C. Lebatteux & M. Zheng (1994): *Coding Chemical Trees with the Centered N-tuple Code*. Journal of Chemical Information and Computer Sciences 34(4), pp. 782–790, doi:10.1021/ci00020a010.

[16] J. B. Hendrickson & C. A. Parks (1991): *Generation and enumeration of carbon skeletons*. Journal of Chemical Information and Computer Sciences 31(1), pp. 101–107, doi:10.1021/ci00001a018.

[17] S. Heubach, N. Y. Li & T. Mansour (2008): *Staircase tilings and k-Catalan structures*. Discrete Mathematics 308(24), pp. 5954–5964, doi:10.1016/j.disc.2007.11.012.

[18] J. F Korsh (2005): *Generating t-ary trees in linked representation*. The Computer Journal 48(4), pp. 488–497, doi:10.1093/comjnl/bxh110.

[19] J. F. Korsh & P. LaFollette (1999): *Loopless generation of Gray codes for k-ary trees*. Information processing letters 70(1), pp. 7–11, doi:10.1016/S0020-0190(99)00035-6.

[20] M. Lepovic & I. Gutman (1998): *A collective property of trees and chemical trees*. Journal of chemical information and computer sciences 38(5), pp. 823–826, doi:10.1021/ci980004b.

[21] L. Li (1986): *Ranking and unranking of AVL-trees*. SIAM Journal on Computing 15(4), pp. 1025–1035, doi:10.1137/0215073.

[22] J. M. Lucas, D. R. Vanbaronaigien & F. Ruskey (1993): *On rotations and the generation of binary trees*. Journal of Algorithms 15(3), pp. 343–366, doi:10.1006/jagm.1993.1045.

[23] S. Nakano & T. Uno (2005): *Constant time generation of trees with specified diameter*. In: 30th International Workshop on Graph-Theoretic Concepts in Computer Science, Springer, pp. 33–45, doi:10.1007/978-3-540-30559-03.

[24] J. M. Pallo (1987): *Generating trees with n nodes and m leaves*. International journal of computer mathematics 21(2), pp. 133–144, doi:10.1080/00207168708803562.

[25] J. M. Pallo (1990): *A simple algorithm for generating neuronal dendritic trees*. Computer methods and programs in biomedicine 33(3), pp. 165–169, doi:10.1016/0169-2607(90)90038-B.

[26] J. M. Pallo & R. Racca (1985): *A note on generating binary trees in A-order and B-order*. International Journal of Computer Mathematics 18(1), pp. 27–39, doi:10.1080/00207168508803477.

[27] F. Ruskey (1978): *Generating t-ary trees lexicographically*. SIAM Journal on Computing 7(4), pp. 424–439, doi:10.1137/0207034.

[28] E. Seyedi-Tabari, H. Ahrabian & A. Nowzari-Dalini (2010): *A new algorithm for generation of different types of RNA*. International Journal of Computer Mathematics 87(6), pp. 1197–1207, doi:10.1080/00207160802140049.

[29] M. Shimizu, H. Nagamochi & T. Akutsu (2011): *Enumerating tree-like chemical graphs with given upper and lower bounds on path frequencies*. BMC Bioinformatics 12(14), pp. 1–9, doi:10.1186/1471-2105-12-S14-S3.

[30] V. Vajnovszki & J. M. Pallo (1994): *Generating binary trees in A-order from codewords defined on a four-letter alphabet*. Journal of Information and Optimization Sciences 15(3), pp. 345–357, doi:10.1080/02522667.1994.10699193.

[31] V. Vajnovszki & J. M. Pallo (1997): *Ranking and unranking k-ary trees with a 4 k-4 letter alphabet*. Journal of Information and Optimization Sciences 18(2), pp. 271–279, doi:10.1080/02522667.1997.10699333.

[32] H. S. Wilf & N. A. Yoshimura (1989): *Ranking rooted trees, and a graceful application*. Annals of the New York Academy of Sciences 576(1), pp. 633–640, doi:10.1111/j.1749-6632.1989.tb16444.x.

[33] P. Willett, J. M. Barnard & G. M. Downs (1998): *Chemical similarity searching*. Journal of chemical information and computer sciences 38(6), pp. 983–996, doi:10.1021/ci9800211.

[34] R. A. Wright, B. Richmond, A. Odlyzko & B. D. McKay (1986): *Constant time generation of free trees*. SIAM Journal on Computing 15(2), pp. 540–548, doi:10.1137/0215039.

[35] R. Wu, J. Chang & C. Chang (2011): *Ranking and unranking of non-regular trees with a prescribed branching sequence*. Mathematical and Computer Modelling 53(5), pp. 1331–1335, doi:10.1016/j.mcm.2010.12.019.

[36] R. Wu, J. Chang & Y. Wang (2006): *A linear time algorithm for binary tree sequences transformation using left-arm and right-arm rotations*. Theoretical Computer Science 355(3), pp. 303–314, doi:10.1016/j.tcs.2006.01.022.

[37] L. Xiang, K. Ushijima & C. Tang (2001): *On generating k-ary trees in computer representation*. Information processing letters 77(5), pp. 231–238, doi:10.1016/S0020-0190(00)00155-1.

[38] K. Yamanaka, Y. Otachi & S. Nakano (2009): *Efficient enumeration of ordered trees with k leaves*. In: WALCOM: Algorithms and Computation, Springer, pp. 141–150, doi:10.1016/j.tcs.2011.01.017.

[39] S. Zaks (1980): *Lexicographic generation of ordered trees*. Theoretical Computer Science 10(1), pp. 63–82, doi:10.1016/0304-3975(80)90073-0.

[40] B. Zhuang & H. Nagamochi (2010): *Constant Time Generation of Trees with Degree Bounds*. In: 9th International Symposium on Operations Research and Its Applications, pp. 183–194, doi:10.1.1.385.6436.