ON A GENERAL WEIGHT OF TREES

R. Kemp
Johann Wolfgang Goethe-Universität
Fachbereich Informatik (20)
D-6000 Frankfurt a. M.

Abstract. We define a general weight of the nodes of a given tree $T$; it depends on the structure of the subtrees of a node, on the number of interior and exterior nodes of these subtrees and on three weight functions defined on the degrees of the nodes appearing in $T$. Choosing particular weight functions, the weight of the root of the tree is equal to its internal path length, to its external path length, to its internal degree path length, to its external degree path length, to its number of nodes of some degree $r$, etc.

For a simply generated family of rooted planar trees $\mathcal{T}$ (e.g. all trees defined by restrictions on the set of allowed node degrees), we shall derive a general approach to the computation of the average weight of a tree $T \in \mathcal{T}$ with $n$ nodes and $m$ leaves for arbitrary weight functions, on the assumption that all these trees are equally likely. This general result implies exact and asymptotic formulas for the average weight of a tree $T \in \mathcal{T}$ with $n$ nodes for arbitrary weight functions satisfying particular conditions. Furthermore, this approach enables us to derive explicit and asymptotic expressions for the different types of average path lengths of a tree $T \in \mathcal{T}$ with $n$ nodes and of all ordered trees with $n$ nodes and $m$ leaves.

I. Introduction

Let $T=(N,L,r)$ be an unlabelled rooted planar tree with the set of nodes $N$, the set of leaves $L$ and the root $r \in N$. Throughout this paper we shall constantly use the convention that the one node tree has no interior nodes and exactly one leaf. The weight $\xi_{f,g,h}(x)$ of a node $x \in N$ is recursively defined by

$$\xi_{f,g,h}(x) := \begin{cases} \xi(x) & \text{IF } x \in L \text{ THEN } 0 \\ \text{ELSE} \sum_{1 \leq i \leq d} f(x_i) + \sum_{1 \leq i \leq d} \frac{|L_i|}{\sum_{1 \leq i \leq d} |N_i \setminus L_i|} + g(d), \end{cases}$$

where $x \in N$ has the $d$ subtrees $T_i=(N_i,L_i,x_i)$, $1 \leq i \leq d$, and $f,g,h: N \to \mathbb{R}$ are given mappings, the so-called weight functions. A tree $T$ has weight $w$ if $\xi_{f,g,h}(r) = w$.

For example, consider the tree $T=(N,L,r)$ drawn in Figure 1. We obtain
the following weights of the nodes for given weight functions $f, g$ and $h$:

$$
\begin{align*}
\xi_{f,g,h}(v) &= 0 \text{ for } v \in \{v_4, v_6, v_7, v_9, v_{10}, v_{11}, v_{12}\}, \\
\xi(v_6) &= \xi(v_3) = f(1) + g(1) + h(1), \\
\xi(v_2) &= f(2) + 2g(2) + 2h(2), \\
\xi(v_1) &= f(1) + f(3) + g(1) + 4g(3) + h(1) + 3h(3), \\
\xi(v_2) &= 2f(2) + 6g(2) + 5h(2), \\
\xi(r) &= 2f(1) + 2f(2) + 2f(3) + 2g(1) + 6g(2) + 16g(3) + 2h(1) + 5h(2) + 10h(3).
\end{align*}
$$

Choosing particular weight functions $f, g$ and $h$, the weight $w$ of a tree corresponds to important parameters which are directly related to the analysis of algorithms. Here are some typical examples:

- Let $f(\lambda) = 0$ and $g(\lambda) = -h(\lambda) = 1$ for $\lambda \in \mathbb{N}$.

  Obviously, the weight of a tree $T = (N, L, r)$ is equal to the internal path length of $T$ which is defined to be the sum - taken over all interior nodes $N \sim L$ - of the lengths of the paths from the root to each node. This quantity is related to the execution time of particular algorithms, e.g. binary search in an ordered table ([7; p. 410]), binary tree searching ([7; p. 427]) or digital searching ([7; p. 495]).

- Let $f(\lambda) = g(\lambda) = 0$ and $h(\lambda) = 1$ for $\lambda \in \mathbb{N}$.

  In this case, the weight of a tree $T = (N, L, r)$ is equal to the external path length of $T$ which is defined to be the sum - taken over all leaves $L$ - of the lengths of the paths from the root to each node. This parameter plays a part in problems dealing with optimum patterns for merging on a tape ([7; p. 306]) or with the number of comparisons and the time required by particular sorting methods ([7; pp. 194, 367, 410, 427, 495]).

- Let $f(\lambda) = g(\lambda) = 0$ and $h(\lambda) = \lambda$ for $\lambda \in \mathbb{N}$.

  Evidently, the weight of a tree $T = (N, L, r)$ is equal to the external degree path length of $T$ which is defined by the sum - taken over all