IX International Seminar on
Analysis of Algorithms
San Miniato, Pisa, June 22-28, 2003

The internal path length of proper generating trees

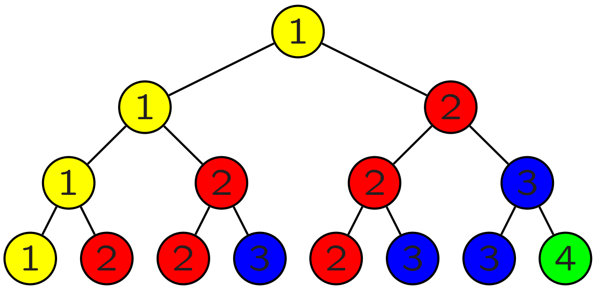
Donatella Merlini

Università di Firenze
Dipartimento di Sistemi e Informatica
via Lombroso 6/17, 50134
merlini@dsi.unifi.it

Proper generating trees

DEFINITION A generating tree is a rooted labeled tree with the property that if v_1 and v_2 are any two nodes with the same label then, for each label l, v_1 and v_2 have exactly the same number of children with label l. To define a generating tree it therefore suffices to specify: the label of the root and a set of rules explaining how to derive from the label of a parent node the labels of all of its children.

$$\begin{cases} root: & (1) \\ rule: & (k) \quad \rightarrow (k)(k+1) \end{cases}$$



DEFINITION An infinite matrix $\{d_{n,k}\}_{n,k\in\mathbf{N}}$ is *associated to a generating tree* with root (c) (AGT matrix for short) if $d_{n,k}$ is the number of nodes at level n with label $k+c$.

Under suitable conditions, this matrix corresponds to a proper Riordan Array, and vice versa (Theorem 2). In the previous example the AGT matrix is the Pascal triangle:

n/k	0	1	2	3	4
0	1				
1	1	1			
2	1	2	1		
3	1	3	3	1	
4	1	4	6	4	1

Proper Riordan Arrays

A Proper Riordan Array (pRA, for short) $\{d_{n,k}\}_{n,k\in\mathbb{N}}$ is defined by a couple of formal power series $(d(t), h(t))$, $d(0) \neq 0, h(0) \neq 0$, such that

$$d_{n,k} = [t^n]d(t)(th(t))^k$$

In the Pascal triangle, we have

$$d(t) = h(t) = \frac{1}{1-t}, \quad d_{n,k} = [t^{n-k}]\frac{1}{(1-t)^{k+1}} = \binom{n}{k}$$

Other properties are:

$$d_{n+1,k+1} = a_0 d_{n,k} + a_1 d_{n,k+1} + a_2 d_{n,k+2} + \ldots = \sum_{j=0}^{\infty} a_j d_{n,k+j}$$

$$d_{n+1,0} = z_0 d_{n,0} + z_1 d_{n,1} + z_2 d_{n,2} + \ldots = \sum_{j=0}^{\infty} z_j d_{n,j}$$

The sequences $A = \{a_0, a_1, a_2, \ldots\}$ and $Z = \{z_0, z_1, z_2, \ldots\}$ are called the $A-$sequence and the $Z-$sequence of the pRA and we have:

$$h(t) = A(th(t)), \quad A(y) = [h(t) \,|y = th(t)]$$

$$d(t) = \frac{d_0}{1 - tZ(th(t))}, \quad Z(y) = \left[\left.\frac{d(t) - d_0}{td(t)}\right| y = th(t)\right]$$

pRA have many other properties ....

In [Spr94] it is proved a result concerning the computation of combinatorial sums:

Theorem 1 *Let* $D = (d(t), h(t))$ *be a Riordan Array and* $f(t)$ *the generating function for the sequence* $\{f_k\}_{k \in \mathbf{N}}$*. Then:*

$$\sum_{k=0}^{n} d_{n,k} f_k = [t^n] d(t) f(th(t)).$$

The connection between pRA and generating trees is studied in [MV00]:

Theorem 2 *Let* $c \in \mathbf{N}$, $a_j, z_j \in \mathbf{N}$, $\forall j \geq 0$, $a_0 \neq 0$ *and* $k \geq c$ *and let*

$$\begin{cases} root: & (c) \\ rule: & (k) \quad \rightarrow (c)^{z_{k-c}} (c+1)^{a_{k-c}} (c+2)^{a_{k-c-1}} \cdots (k+1)^{a_0} \end{cases}$$

be a generating tree specification. Then, the AGT matrix is a pRA D *defined by the triple* (d_0, A, Z)*, such that*

$$d_0 = 1, \quad A = (a_0, a_1, a_2, \ldots), \quad Z = (z_0, z_1, z_2, \ldots).$$

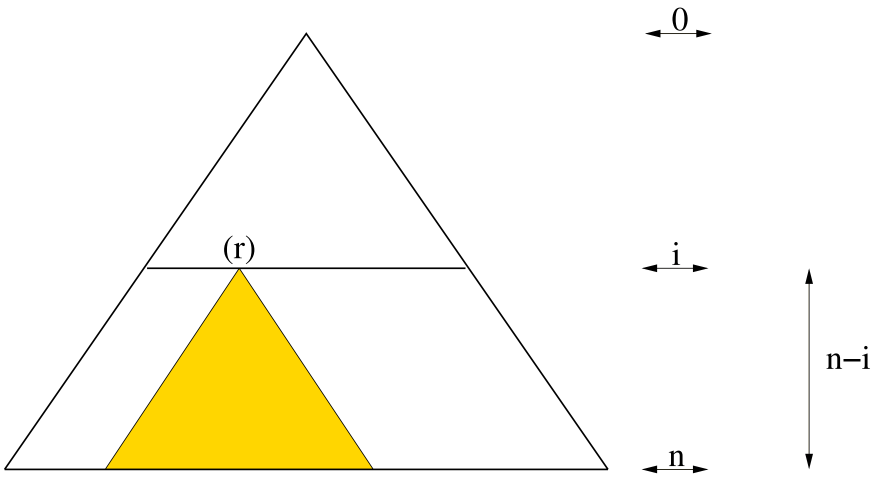
On the contrary, if D *is a pRA defined by the triple* (d_0, A, Z) *with* $d_0 = 1$ *and* $a_j, z_j \in \mathbf{N}$, $\forall j \geq 0$*, then* D *is the AGT matrix associated to the generating tree specification.*

The internal path length

Let P_n the internal path length up to level n in the tree, being 0 the level of the root and $P(t) = \sum_{n\geq 0} P_n t^n$. We have: $P_n = \sum_{i=0}^{n} P_{i,n}$ where $P_{i,n}$ is the sum of the labels at level i counted with their multiplicity.

If $f_j(t)$ is the generating function counting the number of nodes at a given level in the generating tree having root labeled (j) we have:

$$P_{i,n} = \sum_{r\geq 0} d_{i,r}(r+c)[t^{n-i}]f_{r+c}(t)$$



where $d_{i,r}$ counts the number of nodes at level i with label $(r+c)$, being (c) the label of the root.

The computation of $P(t)$

....after many computations it is possible to find formulas for the generating functions $f_j(t)$ and $F(t,w) = \sum_{r\geq 0} f_{r+c}(t)w^r$. By using Theorems 1 and 2 and the theory of Riordan arrays the generating function

$$P(t) = \sum_{n\geq 0}\left(\sum_{i=0}^{n}\sum_{r\geq 0} d_{i,r}(r+c)[t^{n-i}]f_{r+c}(t)\right)t^n$$

can be finally computed and the following Theorems 3 and 4 are proved!!!

The renewal case: $d(t) = h(t)$

In this case we have $d(t) = A(td(t))$ and

$$A(t) = 1 + a_1 t + a_2 t^2 + a_3 t^3 \dots$$

Theorem 3 *Let $P(t) = \sum_{n \geq 0} P_n t^n$ be the generating function counting the total internal path length of any generating tree corresponding to the specification rule:*

$$\begin{cases} root: & (c) \\ rule: & (k) \quad \to (c)^{a_{k-c+1}}(c+1)^{a_{k-c}}(c+2)^{a_{k-c-1}} \cdots (k)^{a_1}(k+1) \end{cases}$$

Then

$$P(t) = \frac{t^2 d(t)(1-td(t))d''(t) - 2t^2(1-c+(c-2)td(t))d'(t)^2}{2(d(t)+td'(t))(1-td(t))^3} +$$

$$+ \frac{4td(t)(c+(1-c)td(t))d'(t) + 2(1-c)td(t)^3 + 2cd(t)^2}{2(d(t)+td'(t))(1-td(t))^3}$$

For the Pascal triangle we have:

$P(t) = \frac{1-t}{(1-2t)^3} =$

$1 + 5\,t + 18\,t^2 + 56\,t^3 + 160\,t^4 + 432\,t^5 + O\left(t^6\right).$

The general case

In this case we have $h(t) = A(th(t))$ and

$$A(t) = 1 + a_1 t + a_2 t^2 + a_3 t^3 \dots$$

Theorem 4 *Let $P(t) = \sum_{n\geq 0} P_n t^n$ be the generating function counting the total internal path length of any generating tree corresponding to the specification rule:*

$$\begin{cases} root: & (c) \\ rule: & (k) \quad \to (c)^{z_{k-c}}(c+1)^{a_{k-c}}(c+2)^{a_{k-c-1}} \cdots (k+1)^{a_0} \end{cases}$$

$$P(t) = \frac{N_P(t)}{D_P(t)}$$

where

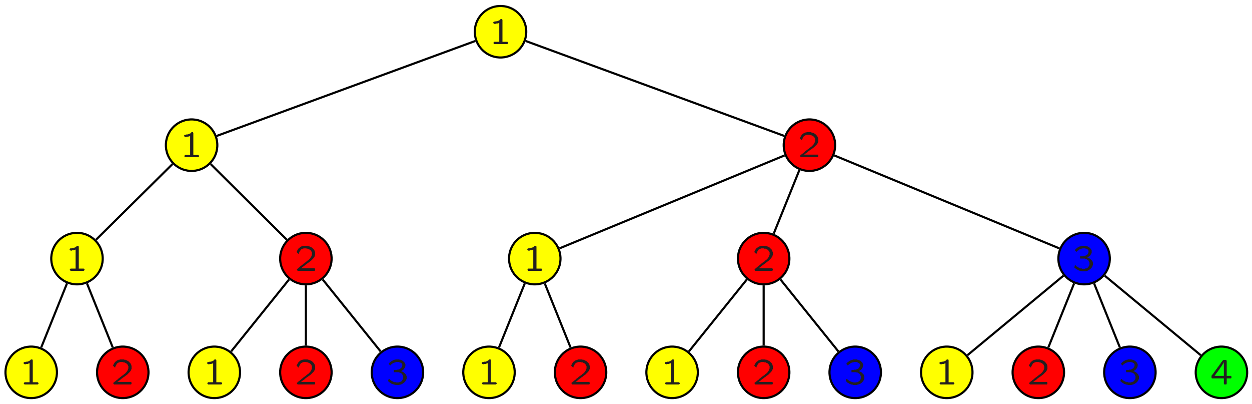
$$\begin{aligned} N_P(t) &= \tfrac{1}{2}t^2 d(t)h(t)d''(t)(1-th(t))^2 + \\ &+ \tfrac{1}{2}t^3 d(t)^2 h(t)h''(t)(1-th(t)) + \\ &+ t^3 d(t)^2 h'(t)^2 (c + (1-c)th(t)) - t^2 h(t)d'(t)^2 (1-th(t))^2 + \\ &+ ctd(t)h(t)d'(t)(1-th(t))^2 + d(t)^2 h(t)(c + (1-c)th(t)) + \\ &+ td(t)h'(t)((ct - 2ct^2 h(t) + ct^3 h(t)^2)d'(t) + \\ &+ (c + 2th(t) - ct^2 h(t)^2)d(t)), \end{aligned}$$

$$D_P(t) = d(t)(h(t) + th'(t))(1-th(t))^3.$$

The Catalan example

The specification rule is:

$$\begin{cases} root: & (1) \\ rule: & (k) \quad \rightarrow (1)\cdots(k)(k+1) \end{cases}$$



and the corresponding associated matrix is:

n/k	0	1	2	3	4
0	1				
1	1	1			
2	2	2	1		
3	5	5	3	1	
4	14	14	9	4	1

which corresponds to

$$A(t) = Z(t) = \frac{1}{1-t} \quad \text{or} \quad d(t) = h(t) = \frac{1-\sqrt{1-4t}}{2t}.$$

In this case, Theorem 3 gives:

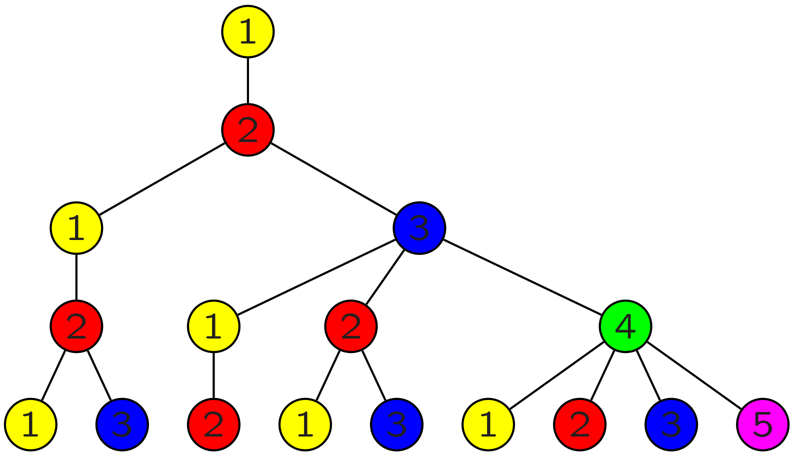
$$P(t) = \frac{1-\sqrt{1-4t}}{2t(1-4t)} = 1+5\,t+22\,t^2+93\,t^3+386\,t^4+O\left(t^5\right).$$

This is sequence A000346 in *http://www.research.att.com/~njas/sequences* and can also be found in [MSV02] in connection with "the tennis ball problem".

The Motzkin example

The specification rule is:

$$\begin{cases} root: & (1) \\ rule: & (k) \quad \rightarrow (1)\ldots(k-1)(k+1) \end{cases}$$



and the corresponding associated matrix is:

n/k	0	1	2	3	4
0	1				
1	0	1			
2	1	0	1		
3	1	2	0	1	
4	3	2	3	0	1

which corresponds to $d(t) = h(t) = \frac{1+t-\sqrt{1-2t-3t^2}}{2t(1+t)}$ or

$$A(t) = \frac{1-t+t^2}{1-t}, \qquad Z(t) = \frac{t}{1-t}$$

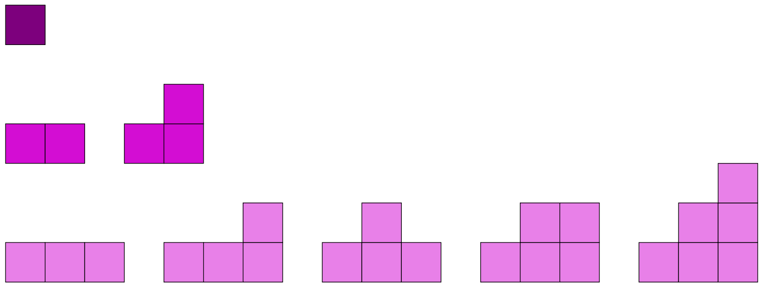
In this case, Theorem 3 gives:

$$P(t) = \frac{2(1+t-\sqrt{1-2t-3t^2})(2t^2+t-1-\sqrt{1-2t-3t^2})}{t(3t-1)(1+t+\sqrt{1-2t-3t^2})^3} =$$

$$= 1 + 3\,t + 10\,t^2 + 31\,t^3 + 96\,t^4 + O\left(t^5\right).$$

Conclusions and Open Problems

- The total internal path length corresponds to the total area of the histograms obtained by juxtaposing n columns having height equal to the labels found in all the paths of length n in the generating tree. For the Catalan example, the histograms with $n \leq 3$ are:



- The internal path length of a generating tree is also related to the area of lattice paths according to the model of [BM02].

- Other interpretations of the internal path length can be found in [MSV01] and [MSV02]

- The problem of determining the generating function counting the total internal path length of any generating tree remains an OPEN PROBLEM.

- What about the asymptotic behaviour of the generating function $P(t)$ in Theorems 3 and 4?

Some Bibliography

- [BM02] C. Banderier and D. Merlini. Lattice paths with an infinite set of jumps. *Proceedings of the 14-th International Conference on Formal Power Series and Algebraic Combinatorics, Melbourne, Australia*, 2002.

- [MSV0] D. Merlini and M. C. Verri. Generating trees and proper Riordan Arrays. *Discrete Mathematics*, 218: 167–183, 2000.

- [MSV01] D. Merlini, R. Sprugnoli, and M. C. Verri. Waiting patterns for a printer. To appear in Discrete Applied Mathematics.

- [MSV02] D. Merlini, R. Sprugnoli, and M. C. Verri. The tennis ball problem. *Journal of Combinatorial Theory, A*, 307–344, 2002.

- [Spr94] R. Sprugnoli. Riordan arrays and combinatorial sums. *Discrete Mathematics*, 132:267–290, 1994.