CROATICA CHEMICA ACTA CCACAA **75** (4) 881–889 (2002)

ISSN-0011-1643
CCA-2838 *Original Scientific Paper*

Morgan Trees and Dyck Paths

Tomislav Došlić

Department of Informatics and Mathematics, Faculty of Agriculture, University of Zagreb, Svetošimunska c. 25, 10000 Zagreb, Croatia (E-mail: doslic@faust.irb.hr)

Received July 15, 2002; accepted September 9, 2002

A simple bijection is established between Morgan trees and Dyck paths. As a consequence, exact enumerative results for Morgan trees on given number of vertices are obtained in terms of Catalan numbers. The results are further refined by enumerating all Morgan trees with prescribed number of internal vertices and by computing the average number of internal vertices in a Morgan tree.

Key words: Morgan trees, Dyck paths, Catalan numbers, Narayana numbers.

INTRODUCTION

The main purpose of this paper is to present a simple bijection between the set of all Morgan trees on a given number of vertices n and the set of all Dyck paths on $2(n-2)$ steps. Dyck paths are just one of many combinatorial families enumerated by Catalan numbers, probably the best-researched and the best-understood of them all. Hence, by constructing the said bijection, many known enumerative results for Dyck paths can be directly applied to Morgan trees. It is our hope that the simplicity of this correspondence will prompt the researchers in mathematical chemistry to pay a closer attention to the area of lattice-paths combinatorics, whose many exact results might have a chemical significance.

MATHEMATICAL PRELIMINARIES

In this section we give only the basic definitions and results necessary for our immediate goal. For a wider combinatorial background we refer the reader to the monograph of Stanley.[1]

A Dyck path on $2n$ steps is a lattice path in the coordinate plane (x,y) from $(0,0)$ to $(2n,0)$ with steps $(1,1)$ (*Up*) and $(1,–1)$ (*Down*), never falling below the x-axis. The set of all Dyck paths of length $2n$ we denote by $\mathscr{D}(n)$. A typical Dyck path of length 14 is shown in Figure 1.

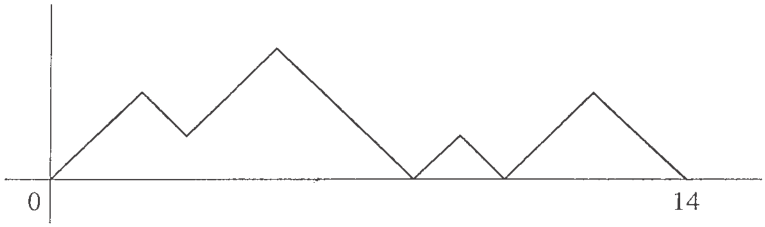

Figure 1. A Dyck path from $\mathscr{D}(7)$.

A peak of a Dyck path is a place where an *Up* step is immediately followed by a *Down* step. The set of all Dyck paths of length $2n$ with exactly k peaks we denote by $\mathscr{D}_k(n)$. Obviously, $\mathscr{D}_k(n) \bigcap \mathscr{D}_l(n) = 0$ for $k \neq l$ and $\mathscr{D}(n) = \bigcup_{1 \leq k \leq n} \mathscr{D}_k(n)$.

In a similar way, motivated by the obvious resemblance of Dyck paths to mountain landscapes, we define valleys as the places where a *Down* step is followed by an *Up* step, and ascents and descents as consecutive sequences of *Up* and *Down* steps, respectively.

There are many ways to prove that there are $\frac{1}{n+1}\binom{2n}{n}$ different Dyck paths on $2n$ steps. Probably the simplest way is to first count all paths from $(0,0)$ to $(2n,0)$ with the steps *Up* and *Down*, dropping the condition that the paths must remain above the x-axis. There are exactly $\binom{2n}{n}$ such paths, since there are $\binom{2n}{n}$ ways to chose n *Up* steps from the total of $2n$ steps. Then a simple reflection principle is applied that reduces this number by the factor of $\frac{1}{n+1}$.

For a given $n \geq 0$, the quantity $C_n = \frac{1}{n+1}\binom{2n}{n}$ is called the n-th Catalan number, and the sequence $(C_n)_{n\geq 0}$ is called Catalan sequence. The first few members are 1, 1, 2, 5, 14, 42, 132, 429, 1030... It turns out that this sequence enumerates many apparently unrelated combinatorial families. Some seventy of them are listed in Ex. 6.19 of Stanley's book,[1] and there are many more. The best known of these families are triangulations of a convex $(n + 2)$-gon, binary trees with n vertices, plane trees with n+1 vertices, n

non-intersecting chords connecting $2n$ points on the circumference of a circle and sequences of n nondecreasing integers a_i such that $a_i \le i$.

Literature on Catalan numbers is vast. A good starting point is the excellent monograph of Stanley,[1] and a reader proficient in Croatian will certainly find the recent monograph of Veljan[2] very useful. We refer the reader to these references for proofs of the following properties of Catalan numbers.

The Catalan numbers satisfy the following convolutional recurrence:

$$C_{n+1} = \sum_{k=0}^{n} C_k C_{n-k}, \quad n \ge 0,$$

with the initial condition $C_0 = 1$. They also satisfy the short recurrence, $C_{n+1} = \frac{2(2n+1)}{n+2} C_n$ for $n \ge 0$, with the same initial condition. Starting from the convolutional recurrence, it is easy to obtain the generating function for the Catalan sequence,

$$C(x) = \sum_{n \ge 0} C_n x^n = \frac{1 - \sqrt{1-4x}}{2x}.$$

For a given sequence $(a_n)_{n\ge 0}$ we define its generating function as a formal power series $a(x) = \sum_{n\ge 0} a_n x^n$. Similarly, for a two-indexed sequence $(b_{n,k})_{n,k\ge 0}$, we define its bivariate generating function by $b(x,t) = \sum_{n,k\ge 0} b_{n,k} x^n t^k$. Generating functions are very useful in establishing asymptotic properties of sequences, in finding averages, *etc*. For a more detailed treatment of this matter, we refer the reader to the classical monograph of Wilf.[3]

The set of $C_3 = 5$ Dyck paths of length 6 is shown in Figure 2. It is obvious that there is only one Dyck path in $\mathscr{D}_1(n)$, three Dyck paths in $\mathscr{D}_2(n)$, and one Dyck path in $\mathscr{D}_3(n)$. There are also other ways of partitioning the set $\mathscr{D}(n)$, for example with respect to the number of path's returns to the x -axis, but the partition with respect to the number of peaks will be the most important for our purposes. For more results on enumeration of Dyck paths with respect to various parameters, we refer the reader to the recent article of Deutsch.[4]

Figure 2. All Dyck paths from $\mathscr{D}(3)$.

The Narayana numbers $N(n,k)$ are defined for integers $n, k \geq 1$ by

$$N(n,k) = \frac{1}{n}\binom{n}{k}\binom{n}{k-1} = \frac{1}{k}\binom{n}{k-1}\binom{n-1}{k-1},$$

with the initial value $N(0,0) = 1$ and the boundary values $N(n,0) = 0$, $N(n,1) = 1$ for $n \geq 1$.

It is easy to see, by a direct computation, that the Narayana numbers decompose the Catalan numbers, *i.e.* that $\sum_{k \geq 0} N(n,k) = C_n$, for all $n \geq 0$. The more interesting is the fact that this decomposition is the one defined by the number of peaks in Dyck paths. In other words, the Narayana numbers $N(n,k)$ enumerate Dyck paths on $2n$ steps with exactly k peaks.

Proposition 1

$$\left|\mathscr{D}_k(n)\right| = N(n,k) .$$

For a combinatorial proof of this result, we refer the reader to the article.[5]

Some of the most interesting properties of Narayana numbers are summarized in the following proposition. Since all claims can be checked using the explicit formulae for $N(n,k)$, we omit the proof.

Proposition 2

For integers $n,k \geq 0$, we have

(a) $N(n, k) = N(n, n+1-k)$ (symmetry);

(b) $\binom{k+1}{2} N(n+1, k+1) = \binom{n+1}{2} N(n,k)$ (absorption law);

(c) $\binom{n}{k-1} N(n, k+1) = \binom{n}{k+1} N(n,k)$;

(d) $\binom{n-k+2}{2} N(n+1, k) = \binom{n+1}{2} N(n,k)$;

(e) $(n+1)N(n,k) = (n-1)[N(n-1,k-1) + N(n-1,k)] + 2\binom{n-1}{k-1}^2$;

(f) $N(n+1, k+1) = \binom{n}{k}^2 - \binom{n}{k-1}\binom{n}{k+1}$;

(g) The sequence $(N(n,k))_{k=1}^{n}$ is log-concave in k. As a consequence, it has one maximum for n odd, and two equal neighboring maximums for n even;

(h) The bivariate generating function $N(x,y)=\sum_{n,k\geq 0} N(n,k)x^n y^k$ satisfies the functional equation

$$xN^2(x,y)+(xy+x-1)N(x,y)+xy=0\ ;$$

(i) $$\sum_{k\geq 0}(-1)^k N(2l,k)=0\ ,\ \sum_{k\geq 0}(-1)^k N(2l+1,k)=(-1)^l C_l.$$

The Narayana numbers form a triangle whose row-sums equal the Catalan numbers. The first few rows of this triangle are shown in Table I.

TABLE I

The beginning rows of the Narayana triangle

$n \backslash k$	0	1	2	3	4	5	6	7	$\Sigma = C_n$
0	1								1
1	0	1							1
2	0	1	1						2
3	0	1	3	1					5
4	0	1	6	6	1				14
5	0	1	10	20	10	1			42
6	0	1	15	50	50	15	1		132
7	0	1	21	105	175	105	21	1	429

MORGAN TREES

Let us consider labeled trees on n vertices. The classical result of Cayley states that there are n^{n-2} such trees. A refined concept, known as a physical tree, has been introduced by Knop *et al.*[6] A physical tree is obtained by assigning labels to the vertices of a tree consecutively, and each vertex to be labeled must have an already labeled neighbour. Hence, each vertex, except the vertex labeled 1, has exactly one neighbour with a lower label. It means that the adjacency matrix $\boldsymbol{A} = \boldsymbol{A}(\mathrm{T})$ of a physical tree T must contain exactly one non-zero element in each column of its upper triangle. A direct consequence of this property is that the total number of different physical trees on n vertices is $(n-1)!$. (There are $(i-1)$ ways to chose the non-zero element in the i-th column of the upper triangle of $\boldsymbol{A}$. Since the choices are independent, the total number of ways equals $\prod_{i=2}^{n}(i-1)=(n-1)!$.)

By imposing further restriction on the labeling process, we arrive at the concept of a Morgan tree. We present here the labeling algorithm in a modified form, according to the reference.[7] It consists of four steps:

1. Assign label 1 to a vertex of the lowest degree (*i.e.* to an endpoint of the tree), and label 2 to its only neighbour.
2. If the vertex 2 is of degree p, assign the labels 3, ..., $p + 2$ to the vertices adjacent to vertex 2.
3. Consider the next vertex possessing the lowest label and which still has non-labeled neighbours. Label its q non-labeled neighbours using q consecutive numbers, starting with the smallest non-used label.
4. Repeat the step 3 while there still are non-labeled vertices.

It is obvious that so defined procedure must terminate after a finite number of steps, so it is indeed an algorithm.

It is worthwhile to note that, due to the step 1 in our naming algorithm, the element $A_{1,2}$ is the only non-zero element in the first row of the upper triangle of the adjacency matrix $\boldsymbol{A}$ of a Morgan tree. Also, the element $A_{2,3}$ is always non-zero. Let us establish another property of $\boldsymbol{A}$ that will prove crucial for our later results.

Lemma 3

Let $\boldsymbol{A}$ be the adjacency matrix of a Morgan tree on n vertices, and let $A_{i,j} = 1$, for some $1 \leq i < j \leq n$. Then $A_{k,l} = 0$, for all $k < i$, $l \geq j$.

Proof

The claim of the lemma is a simple consequence of the monotonicity property of the labeling algorithm: For any two vertices k, i, with $k < i$, all neighbours of k must be labeled before we assign any label to the neighbours of i.

Now we can state our main results.

Theorem 4

The number of all Morgan trees on n vertices is C_{n-2}. The number of all Morgan trees on n vertices with exactly k internal vertices is $N(n-2,k)$.

Proof

Let $MT(n)$ be the set of all Morgan trees on n vertices. For a given tree T $\in MT(n)$, let us consider its adjacency matrix $\boldsymbol{A}$(T). From Lemma 3 we know that this matrix has a characteristic pattern of non-zero elements, best visualized in the form of a staircase. To this matrix, we assign a lattice path D(T) on $2n - 2$ steps as follows. Start the path D(T) by an *Up* step, followed by a *Down* step. Now consider the second row of the matrix $\boldsymbol{A}$(T). For each non-zero element of this row (there must be at least one, $A_{2,3}$), add an *Up* step to the already constructed part of D(T). After reaching the last non-zero element of the second row, say, $A_{2,j}$, add a *Down* step. Now add an *Up* step for each non-zero element of the third row; if there are none, add another *Down* step. Repeat this procedure for all rows of $\boldsymbol{A}$(T). Because of Lemma 3, the path D(T) is indeed a Dyck path on $2(n - 1)$ steps, and it is obvious from the construction that the correspondence is bijective. An example is shown in Figure 3.

$$\begin{bmatrix} 0 & 1 & & & & & & \\ & 0 & 1 & 1 & & & & \\ & & 0 & 0 & 1 & 1 & 1 & \\ & & & 0 & 0 & 0 & 0 & \\ & & & & 0 & 0 & 0 & \\ & & & & & 0 & 0 & 1 \\ & & & & & & 0 & 0 \\ & & & & & & & 0 \end{bmatrix} \longleftrightarrow$$

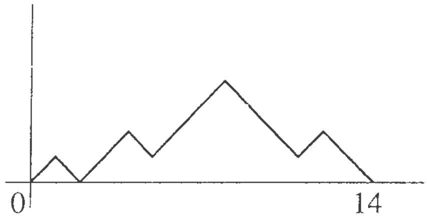

Figure 3. With the proof of Theorem 4.

So, we have reduced our task of enumerating all Morgan trees on n vertices to the enumeration of all Dyck paths on $2(n - 1)$ steps that begin with the (*Up*, *Down*) pair of steps. Obviously, there are exactly C_{n-2} such paths, and this proves the first claim of the theorem.

To prove the second claim, it suffices to note that the peaks of Dyck paths correspond bijectively to the internal vertices in Morgan trees, and *vice versa*.

In Figure 4 we show all five different Morgan trees on five vertices, together with the corresponding Dyck paths.

The second claim of Theorem 4 can be also stated in terms of leaves of Morgan trees.

Corollary 5

The number of Morgan trees on n vertices with exactly k leaves is equal to $N(n - 2, k - 1)$.

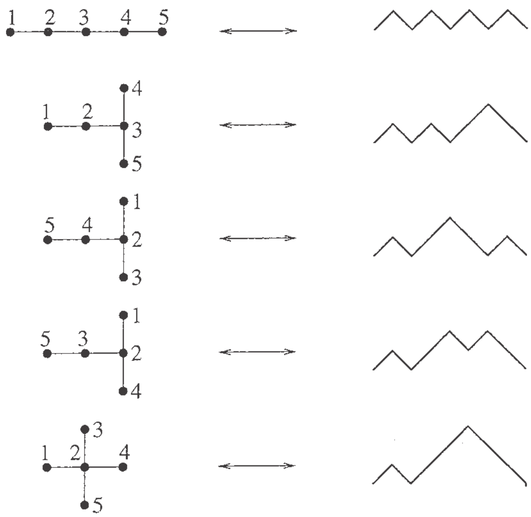

Figure 4. Morgan trees on 5 vertices and the corresponding Dyck paths.

Proof

A Morgan tree on n vertices with exactly k leaves has $n - k$ internal vertices. According to Theorem 4, there are $N(n - 2, n - k)$ such Morgan trees. Applying the symmetry property of Narayana numbers, we get

$$N(n-2, n-k) = N(n-2, n-2+1-n+k) = N(n-2, k-1)\,.$$

Our final result concerns the expected number of internal vertices in a Morgan tree on n vertices, assuming that all such trees are equiprobable.

Corollary 6

The expected number of internal vertices in a Morgan tree on n vertices equals $\frac{n-1}{2}$.

Proof

Let us consider the set of all Dyck paths of length $2n$, $\mathscr{D}(n)$, and denote by σ_n the number of all peaks in $\mathscr{D}(n)$. The generating function of the sequence σ_n is then obtained[3,4] by taking a partial derivative of the bivariate generating function for the Narayana numbers with respect to t, and substituting the value $t = 1$. In other words,

$$\sum_{n\geq 0} \sigma_n x^n = \frac{\partial N}{\partial t}(x,t)\Big|_{t=1}\ ,$$

where $N(x,t)$ is the bivariate generating function for the Narayana numbers. A simple calculation then yields

$$\sum_{n\geq 0} \sigma_n x^n = \frac{1}{2}\left(\frac{1}{\sqrt{1-4x}} - 1\right).$$

Expanding the right hand side *via* binomial theorem, we obtain $\sigma_n = \frac{1}{2}\binom{2n}{n}$, for $n \geq 0$. So, the total number of peaks in $D(n)$ is $\frac{1}{2}\binom{2n}{n}$, and the average (or expected) number of peaks is obtained by dividing this quantity by the total number of paths, C_n. Hence, in a given Dyck path on $2n$ steps there are, on average, $\frac{n+1}{2}$ peaks. Our claim now follows by replacing n by $n-2$.

REFERENCES

1. R. P. Stanley, *Enumerative Combinatorics*, Vol. 2, Cambridge University Press, Cambridge, 1999.
2. D. Veljan, *Combinatorial and Discrete Mathematics* (in Croatian), Algoritam, Zagreb, 2001.
3. H. S. Wilf, *generatingfunctionology*, 2nd ed., Academic Press, New York, 1994.
4. E. Deutsch, *Discrete Math.* **204** (1999) 167–202.
5. D. Veljan, *Math. Commun.* **6** (2001) 217–232.
6. J. V. Knop, W. R. Müller, K. Szymanski, H. W. Kroto, and N. Trinajstić, *J. Comput. Chem.* **8** (1987) 549–554.
7. I. Lukovits and I. Gutman, *Croat. Chem. Acta* **75** (2002) 563–576.

SAŽETAK

Morganova stabla i Dyckovi putovi

Tomislav Došlić

Uspostavljena je jednostavna bijekcija između Morganovih stabala i Dyckovih putova. Kao posljedica, egzaktno su prebrojena Morganova stabla sa zadanim brojem vrhova s pomoću Catalanovih brojeva. Rezultati su dalje utočnjeni prebrojavanjem Morganovih stabala sa zadanim brojem unutarnjih vrhova i računanjem očekivanog broja unutarnjih vrhova u Morganovu stablu.