Enumeration of non-crossing trees

Marc Noy*

Abstract

We consider several enumeration problems concerning labelled trees whose vertices lie in a circle and whose edges are rectilinear and do not cross.

Sommaire

On considère quelques problèmes d'énumération d'arbres étiquetés dont les sommets se trouvent sur un cercle et les arêtes sont des segments rectilignes qui ne se croissent pas.

1 Introduction

Take n points in a circle and consider trees whose vertices are the given points and whose edges are rectilinear and do not cross. Such trees will be called hereafter *non-crossing trees* (*nc-trees* for short). The enumeration of such trees presents several interesting problems. The first fact is that the number of nc-trees on n points is a generalized Catalan number (see [5] for a survey on these numbers). This result could be in principle deduced from formulas obtained by Domb and Barret [3], where the more general problem of enumerating non-crossing graphs in a circle is studied. A simple combinatorial proof appears in [4], where the problem is related to configurations of chords in the circle and to permutations.

Theorem 1.1 ([4]) *The number of nc-trees on n vertices is equal to the number of ternary trees with $n-1$ internal vertices, that is,*

$$t_n = \tau_{n-1} = \frac{1}{2n-1}\binom{3n-3}{n-1}.$$

*Dept. Matemàtica Aplicada II, Universitat Politècnica de Catalunya, Pau Gargallo 5, 08028 Barcelona, Spain. noy@ma2.upc.es

The goal of this paper is to further the enumerative study of non-crossing trees. In Section 2 we determine the number of nc-trees in which the degree of a node is specified, in Section 3 we compute the number of unicyclic non-crossing subgraphs (that is, connected non-crossing subgraphs in a circle with a single cycle), in Section 4 we find close and asymptotic formulas for the number of bipartite nc-trees and the number of nc-forests, while in Section 5 we determine the number of unlabelled nc-trees. Section 6 contains some partial results on the enumeration of nc-trees by the number of leaves and in Section 7 we consider random nc-trees. We conclude with some remarks and open problems.

2 Enumeration according to the degree of a node

Let $t(n,d)$ be the number of trees on n vertices in which a given vertex (say number 1) has degree d. We show how to derive a closed formula for these numbers.

Lemma 2.1 *The numbers $t(n,d)$ can be written in terms of the numbers t_n as follows:*

$$t(n,d) = \sum_{\substack{i_1+\cdots+i_{2d}=n+d-1 \\ i_1,\ldots,i_{2d}\geq 1}} t_{i_1} t_{i_2} \cdots t_{i_{2d}}.$$

Let us introduce the generating functions $T = \sum t_n z^n$ and $T_d = \sum t(n,d) z^n$. Then the above Lemma implies that

$$T_d = \frac{1}{z^{d-1}} T^{2d}.$$

Since $T/z = S$ satisifes $S - 1 = zS^3$, we can apply Lagrange's inversion (see [2]) to get a closed formula.

Theorem 2.2 *For every $d \geq 1$ and $n > d$*

$$t(n,d) = \frac{2d}{3n-d-3}\binom{3n-d-3}{n-d-1}.$$

We will also need a simple consequence of Theorem 2.2.

Corollary 2.3 *The number of nc-trees on n vertices containing the edge $(1,n)$ is equal to*

$$t(n,1) = \frac{2}{3n-4}\binom{3n-4}{n-2}.$$

It is interesting to compare these results with the corresponding formulas for arbitrary labelled trees. If we let $c(n,d)$ be the number of labelled trees on n vertices in which a given node has degree d then it is known [6] that $c(n,d) = \binom{n-2}{d-1}(n-1)^{n-d-1}$ (and of course the number of labelled trees is n^{n-2}). Compare now the ratios $t(n,d)/t_n \sim 4d/3^{d+1}$ and $c(n,d)/n^{n-2} \sim (e(d-1)!)^{-1}$ for fixed d as n goes to infinity.

3 Unicyclic non-crossing graphs

A non-crossing unicyclic graph (unicyclic nc-graph for short) on n vertices on a cricle is a non-crossing connected graph with a unique cycle or, equivalently, a connected non-crossing graph of size n. Unicyclic graphs are very much like trees so the similarity of the following formula with the corresponding one for nc-trees is not surprising.

Theorem 3.1 *The number u_n of unicyclic nc-graphs on n points is given by*

$$u_n = \binom{3n-3}{n-3}.$$

The proof of Theorem 3.1 requires a technical Lemma analogous to Corollary 2.3

Lemma 3.2 *Let w_n be the number of unicyclic nc-graphs containing the edge $(1,n)$. Then for $n \geq 3$ the following equations hold:*

$$\begin{aligned} u_n &= t(2,1)w_n + \cdots + t(n,1)w_2; \\ w_n &= t_n - t(n,1) + 2(u_1 t_{n-1} + \cdots + u_{n-1}t_1). \end{aligned}$$

From Lemma 3.2 one obtains functional equations involving the generating functions $U = \sum u_n z^n$ and $W = \sum w_n z^n$:

$$\begin{aligned} z^2U &= (T_1 - z)W \\ W &= T - T_1 + 2UT, \end{aligned}$$

where T and $T_1 = T^2 + z$ as before (for technical reasons it is convenient to set $t(1,1) = 1$ in this case). And from these equations one gets

$$U = z\frac{S(S-1)^2}{3-2S} = z^3\frac{S^7}{3-2S},$$

where $T = zS$ and $S - 1 = zS^3$. One could in principle use Lagrange's inversion once more to find the coefficients of U but the ensuing expressions become too clumsy. Instead we first prove that $2dU/dz = z^2d^2S/dz^2$ and deduce from it the following simple relation between the number of nc-trees and the number of nc-unicyclic graphs

$$nu_n = \binom{n-1}{2}t_n.$$

The simplicity of this relation between u_n and t_n suggests the possibility of a simple combinatorial proof. This would be certainly desirable but we have not succeeded in finding such a proof.

4 The number of bipartite nc-trees and the number of nc-forests

Consider r black points and s white points distributed in a circle, where $n = r + s$, and let us ask about the number of bipartite nc-trees on these two sets of vertices. The answer will of course depend on the particular distribution of the colours, so we consider two extreme cases: first when the black points are consecutive and then when $r = s$ and the two colours alternate.

Theorem 4.1 *The number of (r,s) bipartite nc-trees when the colors are completely separated is equal to*

$$\binom{n-2}{r-1} = \binom{n-2}{s-1}.$$

When $r = s$ and the colors alternate the number a_n of bipartite nc-trees on n vertices is, asymptotically,

$$a_n \sim K n^{-3/2} \Big(\frac{135 + 78\sqrt{3}}{16}\Big)^n$$

To proof the last statement, one shows that the a_n satisfy the recurrence

$$a_n = \sum_{\substack{i,j,k \geq 1 \\ i+j+k=n+1 \\ i+j \text{ even}}} a_i a_j a_k,$$

which tanslates into a simple functional equation for the generating function $A(z) = \sum a_{n-1} z^n$, namely

$$A(z) - 1 = z\, A(z) \frac{A(z)^2 + A(-z)^2}{2}.$$

After some manipulation one arrives at two equations for $A^+(z)$ and $A^-(z)$, the even and odd parts of $A(z)$. Putting $A^+(z) = f(z^2)$ and $A^-(z) = zg(z^2)$ they are

$$x f(x)^3 (2f(x) - 1)^2 = f(x) - 1 \quad \text{and}$$
$$(1 + x^3 g(x)^3)(1 + 4x g(x)^2) = g(x).$$

Finally we apply a general result of Bender on asymptotics from functional equations (Theorem 5 in [1]).

A similar technique can be used for enumerating non-crossing forests. Let f_n be the number of nc-forests on n points on a circle.

Theorem 4.2 *The generating function $F = \sum f_n z^n$ satisfies $zF^3 + (z^2 - z)F^2 + (2z - 1)F + 1 = 0$. The numbers f_n are, asymptotically, $f_n \sim K n^{-3/2} \alpha^n$, where K is a constant and $\alpha \approx 8.22$ is the largest positive root of $4\alpha^3 - 32\alpha^2 - 8\alpha + 5 = 0$.*

5 Unlabelled nc-trees

When we speak of an unlabelled structure we mean the orbit of a set of labelled structures under the action of some group of symmetries. In the case of non-crossing trees the natural symmetries (automorphisms) to consider are given by the action of the dihedral group. Hence we say two nc-trees on n labelled vertices are equivalent if one is obtained by a rotation and/or a reflection from the other. One can count exactly the number of inequivalent nc-trees using an idea of Moon and Moser [7] in a similar problem concerning triangulations of a convex polygon.

Theorem 5.1 *The number t_n^* of inequivalent nc-tress on n vertices is*

$$t_n^* = \begin{cases} \dfrac{t_n}{2n} + \dfrac{3}{4}t(\dfrac{n}{2}+1,1) & \textit{for even } n; \\ \dfrac{t_n}{2n} + \dfrac{1}{2}t_{\frac{n+1}{2}} & \textit{for odd } n. \end{cases}$$

From the previous theorem it follows that $t_n^* \sim t_n/2n$. This means that most non-crossing trees are rigid, that is, have no symmetry besides the identity.

Since it will be needed later, we consider the particular case of non-crossing Hamiltonian paths. They are easy to count, as well as the number of classes under the action of the dihedral group.

Theorem 5.2 *The number of non-crossing Hamiltonian paths on n points in convex position is given by*

$$p_n = n2^{n-3}$$

and the number of them which are inequivalent is

$$p_n^* = \begin{cases} 2^{n-4} + 2^{(n-4)/2} & \textit{for even } n; \\ 2^{n-4} + 2^{(n-5)/2} & \textit{for odd } n. \end{cases}$$

6 The number of leaves in an nc-tree

Let $r_{n,k}$ be the number of nc-trees on n vertices having exactly k leaves. This is the main enumerative problem for which we have not found a simple answer. In the case of labelled trees the answer is $n!/k!S(n-2, n-k)$, where the $S(n,k)$ are Stirling numbers of the second kind [6], and we also expected a simple closed formula for nc-trees. We have found however a recursive formula that, though rather involved, allows the actual computation of the numbers $r_{n,k}$. After some calculations we were led to the following conjecture:

$$\text{(C)} \qquad r_{n,k} = n\binom{n-2}{k-1} q_k(n) 2^n \qquad \text{for } k \geq 3,$$

where $q_k(n)$ is a polynomial in n of degree $k-3$ with rational coefficients. The conjecture agrees with our data for $3 \leq k \leq 15$ and $2 \leq n \leq 16$. If this were actually correct then the problem would be to determine the polynomials q_k. Note that if α_k is the leading coefficient of q_k then (C) would imply the following simple asymptotic estimate

$$r_{n,k} \sim \frac{\alpha_k}{(k-1)!} 2^n n^{2k-3}$$

for fixed k.

For small values of k we can compute $r_{n,k}$ exactly. As usual, a k-star is a tree isomorphic to $K_{1,k}$. Also, let a (k,j)-star be the disjoint union of a k-star and a j-star plus an edge connecting the corresponding vertices of degree k and j. Also recall that two graphs are homeomorphic if one can be obtained from the other by a sequence of subdivision of edges or deletion of vertices of degree two.

Theorem 6.1 *The number of nc-trees on n vertices homeomorphic to a k-star is equal to*

$$n 2^{n-k-1} \binom{n-2}{k-1} \qquad \text{for } k > 2.$$

The number of nc-trees on n vertices homeomorphic to a (k,j)-star is equal to

$$\begin{cases} n 2^{n-k-j-2}(k+1)(j+1)\binom{n-2}{k+j} & \text{if } k \neq j; \\ \frac{1}{2} n 2^{n-k-j-2}(k+1)(j+1)\binom{n-2}{k+j} & \text{if } k = j. \end{cases}$$

Every tree with only 3 leaves is homeomorphic to a 3-star, hence $r(n,3) = n2^{n-4}\binom{n-2}{2}$. In the same vein one gets $r(n,4) = n2^{n-5}\binom{n-2}{3} + n2^{n-7}\binom{n-2}{4}$. But already $r(n,5)$ cannot be computed using Lemma 6.1 An interesting question would be to compute more generally the number of nc-trees on n vertices homeomorphic to a given general unlabelled tree.

7 Random nc-trees

There are many results on random labelled trees which stem directly from enumerative formulas (see [6]). Theorem 2.2 gives the distribution of nc-trees according to the degrees of the nodes. If $d(x)$ denotes the degree of node x in a random nc-tree, then

$$P(d(x) = d) = t(n,d)/t_n = 2\binom{3n-3}{n-1}^{-1} d \binom{3n-4-d}{2n-3}.$$

We have computed the basic statistics of this distribution.

Theorem 7.1 *The mean and the variance of the distribution $d(x)$ are given by $E(d(x)) = 2(1-1/n)$ and $\sigma^2(d(x)) = (1-1/n)(1-2/n)(3n-2)/(2n+1)$.*

The mean is obviously the same as for unrestricted labelled trees: it is just the sum of the degrees divided by n. The variance is however about 3/2 times higher, since for labelled trees it is equal to $(1-1/n)(1-2/n)$. If we now consider (as in [6]) the number $X = X(n,k)$ of nodes of degree k in a random nc-tree, it is easy to compute the mean $E(X) = n\,t(n,k)/t_n$ but not the second moment. The results mentioned at the end of Section 3 imply that the mean number of leaves in a random nc-tree is $\sim 4n/9 = .444n$ while for random labelled trees it is $\sim n/e = 0.367n$. Hence a typical random nc-tree has more leaves than a random labelled tree.

Another interesting aspect is the generation of random nc-trees, which can be useful for checking conjectures about the distribution of a given variable and for looking at properties of typical nc-trees. For general labelled trees, the Prüfer correspondence between trees and sequences of integers can be used to prove many enumerative formulas and it also leads directly to a random generation scheme; for nc-trees it is the proof of Theorem 1.1 which is useful.

Proof of Theorem 1.1. Let T be an nc-tree on n vertices, and let i be the last index such that i is a neighbor of 1. Then T restricted to the vertex set $\{i, i+1, \ldots, n\}$ is again an nc-tree. Also, since $(1,i)$ is an edge and a tree contains no cycle, there exists a unique $j \in \{1, 2, \ldots, n-1\}$ such that T is again an nc-tree when restricted to $\{1, \ldots, j\}$ and to $\{j+1, \ldots, i\}$.

Conversely, any triple of nc-trees on the corresponding vertex sets uniquely determines an nc-tree on n vertices. □

Now we know the number of nc-trees on n vertices in which i is the last neighbor of 1, namely $t(i,1)t_{n-i+1}$. We select i according to this probability. On the vertex set $\{i, i+1, \ldots, n\}$ we only have to select another random nc-tree. Next we select j as in the proof of Theorem 1.1 with probability $t_j t_{n-j}/t(n,1)$ and further random nc-trees on the sets $\{1, \ldots, j\}$ and $\{j+1, \ldots, i\}$. Then we recurse on the three resulting subsets. A short description of a recursive function *tree(n)* generating a random nc-tree on n vertices follows (the function returns a tree as a set of edges).

$tree(1) := \{\}$
$tree(2) := \{(1,2)\}$
$tree(n) :=$
  select i randomly in $\{2, ..., n\}$ with probability $t(i,1)t_{n-i+1}/t_n$
  select j randomly in $\{1, ..., i-1\}$ with probability $t_j t_{n-j}/t(n,1)$
  return $(1,i) \cup tree(j) \cup (j + tree(i-j)) \cup (i - 1 + tree(n-i+1))$

8 Conclusions and open problems

We have introduced the family of non-crossing trees which, although much smaller than that of labelled trees, is also a rich source of enumeration problems. The classical recursive techniques for the enumeration of labelled trees do not apply in this case and we have developed new techniques, which have proved useful in the solution of several particular problems.

Some other problems concerning nc-trees, like the enumeration of nc-trees with a given partition, determination of the average number of spanning trees in a non-crossing graph (any graph whose points lie in a circle and whose edges are rectilinear and do not cross) and investigation of analogues to the matrix- tree theorem for nc-trees will be the object of future research.

The main open problem left in this paper is to find a reasonably simple expression for $r_{n,k}$, the number of nc-trees on n vertices having exactly k leaves.

Acknowledgement

The author whishes to thank A. García and J. Tejel for suggesting the proof of Theorem 4.2, J.L. Andrés Yebra for useful comments and the referees for calling his attention to reference [4].

References

[1] E.A. Bender, Asymptotic methods in enumeration, SIAM Review 16 (1974), 485-515.

[2] L. Comtet, *Advanced Combinatorics*, Reidel, 1974.

[3] C. Domb and A.J. Barrett, Enumeration of Ladder Graphs, Discrete Math. 9 (1974), 341-358.

[4] S. Dulucq and J.G. Penaud, *Chordes, arbres et permutations*, Discrete Math. 117 (1993), 89-105.

[5] P. Hilton and J. Pedersen, Catalan numbers, Their Generalization and Their Uses, The Math. Intelligencer 13 (1991), 64-75.

[6] J. Moon, *Counting labelled trees*, Canadian Math. Congress, 1970.

[7] J. Moon and L. Moser, Triangular dissections of n-gons, Canad. Math. Bull. 6 (1963), 175-178.