Advances in Applied Mathematics 78 (2016) 76–93

Contents lists available at ScienceDirect

Advances in Applied Mathematics

www.elsevier.com/locate/yaama

The shape of random tanglegrams

Matjaž Konvalinka [a,1], Stephan Wagner [b,2]

[a] *Department of Mathematics, University of Ljubljana, and Institute for Mathematics, Physics and Mechanics, Jadranska 19, Ljubljana, Slovenia*
[b] *Department of Mathematical Sciences, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa*

ARTICLE INFO

Article history:
Received 18 December 2015
Received in revised form 31 March 2016
Accepted 4 April 2016
Available online 14 April 2016

MSC:
60C05

ABSTRACT

A tanglegram consists of two binary rooted trees with the same number of leaves and a perfect matching between the leaves of the trees. We show that the two halves of a random tanglegram essentially look like two independently chosen random plane binary trees. This fact is used to derive a number of results on the shape of random tanglegrams, including theorems on the number of cherries and generally occurrences of subtrees, the root branches, the number of automorphisms, and the height. For each of these, we obtain limiting probabilities or distributions. Finally, we investigate the number of matched cherries, for which the limiting distribution is identified as well.

1. Introduction

Tanglegrams are, intuitively, graphs obtained by taking two binary rooted trees with the same number of leaves (which is the *size* of a tanglegram) and matching each leaf

E-mail addresses: matjaz.konvalinka@fmf.uni-lj.si (M. Konvalinka), swagner@sun.ac.za (S. Wagner).

[1] Partially supported by Research Program Z1-5434 of the Slovenian Research Agency.
[2] This material is based upon work supported financially by the National Research Foundation of South Africa under grant number 96236.

http://dx.doi.org/10.1016/j.aam.2016.04.001
0196-8858/

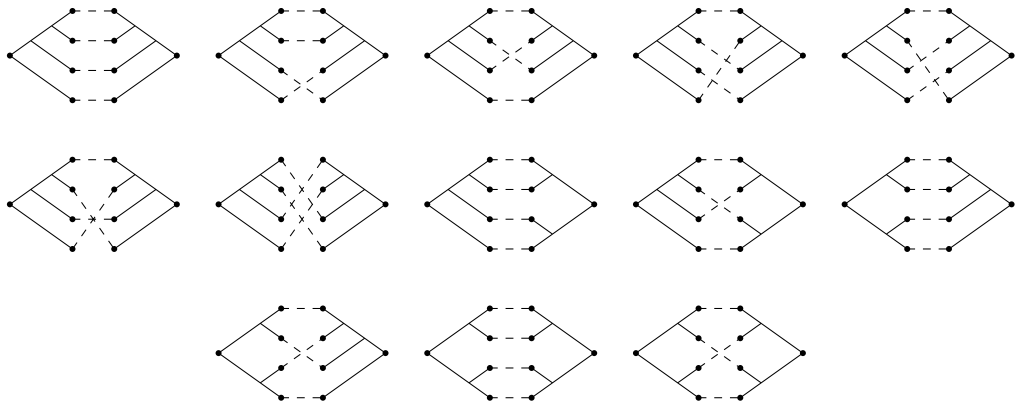

Fig. 1. The 13 tanglegrams of size 4.

from the tree on the left with a unique leaf from the tree on the right. Furthermore, we consider two tanglegrams to be the same if we can get one from the other by an isomorphism that fixes the roots. For example, Fig. 1 shows all 13 tanglegrams of size 4.

Let us make this definition more precise. A *plane binary tree* has one distinguished vertex assumed to be a common ancestor of all other vertices, and each vertex either has two children (left and right) or no children. A vertex with no children is a *leaf*, and a vertex with two children is an *internal vertex*. It is well known that the number of plane binary trees with n leaves is the Catalan number $\frac{1}{n}\binom{2n-2}{n-1}$, henceforth denoted by C_n. The sequence starts with

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, 208012,$$

see [10, A000108] and [12, §6] for more information.

Two plane binary trees with labeled leaves are said to be *equivalent* if there is an isomorphism from one to the other as graphs mapping the root of one to the root of the other. Let $\mathcal{B}_n$ be the set of inequivalent plane binary trees with $n \geq 1$ leaves. In the following, we will refer to the elements of $\mathcal{B}_n$ merely as *binary trees* for simplicity. The sets $\mathcal{B}_n$ are enumerated by the Wedderburn–Etherington numbers, whose sequence starts

$$1, 1, 1, 2, 3, 6, 11, 23, 46, 98, 207, 451, 983,$$

see [10, A001190] for more information.

For each plane binary tree T, denote by $A(T)$ its *automorphism group*, which can be interpreted as a subgroup of the permutation group of the set of leaves, i.e. as a subgroup of $\mathfrak{S}_n$. Given a permutation $v \in \mathfrak{S}_n$ along with two trees $T, S \in \mathcal{B}_n$, each with leaves labeled $1, \ldots, n$, we construct an *ordered binary rooted tanglegram* (or *tanglegram* for short) (T, v, S) of size n with T as the left tree, S as the right tree, by identifying leaf i in T with leaf $v(i)$ in S. Furthermore, (T, v, S) and (T', v', S') are considered to represent