



ELSEVIER

Contents lists available at ScienceDirect

Journal of Combinatorial Theory,
Series A

www.elsevier.com/locate/jcta



On the enumeration of tanglegrams and tangled chains [☆]



Sara C. Billey ^a, Matjaž Konvalinka ^b, Frederick A. Matsen IV ^c

^a Department of Mathematics, University of Washington, Seattle, WA 98195, USA

^b Department of Mathematics, University of Ljubljana, and Institute of Mathematics, Physics and Mechanics, Ljubljana, Slovenia

^c Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

ARTICLE INFO

Article history:

Received 3 November 2015

Available online xxxx

Keywords:

Tanglegrams

Enumeration

Binary trees

Binary partitions

ABSTRACT

Tanglegrams are a class of graphs arising in computer science and in biological research on cospeciation and coevolution. They are formed by identifying the leaves of two rooted binary trees. We give an explicit formula to count the number of distinct binary rooted tanglegrams with n matched leaves, along with a simple asymptotic formula and an algorithm for choosing a tanglegram uniformly at random. The enumeration formula is then extended to count the number of tangled chains of binary trees of any length. This includes a new formula for the number of binary trees with n leaves. We also give a conjecture for the expected number of cherries in a large randomly chosen binary tree and an extension of this conjecture to other types of trees.

© 2016 Elsevier Inc. All rights reserved.

[☆] The first author was partially supported by the National Science Foundation grant DMS-1101017. The second author was supported by Research Program Z1-5434 and Research Project BI-US/14-15-026 of the Slovenian Research Agency. The third author was supported by National Science Foundation grants DMS-1223057 and CISE-1564137.

E-mail address: matjaz.konvalinka@fmf.uni-lj.si (M. Konvalinka).

URLs: <http://www.math.washington.edu/~billey/> (S.C. Billey), <http://www.fmf.uni-lj.si/~konvalinka/> (M. Konvalinka), <http://matsen.fredhutch.org/> (F.A. Matsen).

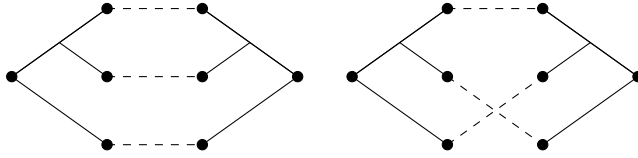


Fig. 1. The tanglegrams of size 3.

1. Introduction

Tanglegrams are graphs obtained by taking two binary rooted trees with the same number of leaves and matching each leaf from the tree on the left with a unique leaf from the tree on the right. This construction is used in the study of cospeciation and coevolution in biology. For example, the tree on the left may represent the phylogeny of a host, such as gopher, while the tree on the right may represent a parasite, such as louse [14], [22, page 71]. One important problem is to reconstruct the historical associations between the phylogenies of host and parasite under a model of parasites switching hosts, which is an instance of the more general problem of *cophylogeny estimation*. See [22–24] for applications in biology. Diaconis and Holmes have previously demonstrated how one can encode a phylogenetic tree as a series of binary matchings [7], which is a distinct use of matchings from that discussed here.

In computer science, the Tanglegram Layout Problem (TL) is to find a drawing of a tanglegram in the plane with the left and right trees both given as planar embeddings with the smallest number of crossings among (straight) edges matching the leaves of the left tree and the right tree [2]. These authors point out that tanglegrams occur in the analysis of software projects and clustering problems.

In this paper, we give the exact enumeration of tanglegrams with n matched pairs of vertices, along with a simple asymptotic formula and an algorithm for choosing a tanglegram uniformly at random. We refer to the number of pairs of matched vertices in a tanglegram as its *size*. Furthermore, two tanglegrams are considered to be equivalent if one is obtained from the other by replacing the tree on the left or the tree on the right by isomorphic trees. For example, in Fig. 1, the two non-equivalent tanglegrams of size 3 are shown.

We state our main results here postponing some definitions until Section 2. The following is our main theorem.

Theorem 1. *The number of tanglegrams of size n is*

$$t_n = \sum_{\lambda} \frac{\prod_{i=2}^{\ell(\lambda)} (2(\lambda_i + \dots + \lambda_{\ell(\lambda)}) - 1)^2}{z_{\lambda}},$$

where the sum is over binary partitions of n and z_{λ} is defined by Equation (1).

Download English Version:

<https://daneshyari.com/en/article/4655017>

Download Persian Version:

<https://daneshyari.com/article/4655017>

[Daneshyari.com](https://daneshyari.com)