Trees, Tanglegrams, and Tangled Chains

Sara Billey
University of Washington

Based on joint work with:
Matjaž Konvalinka and Frederick (Erick) Matsen IV
arXiv:1507.04976

WPI, September 12, 2015

Outline

Background

Formulas for Trees, Tanglegrams and Tangled Chains

Algorithms for random generation

Open Problems

Rooted Binary Trees

- B_n= set of rooted inequivalent binary trees with n leaves
- $|B_n| \longrightarrow 1, 1, 1, 2, 3, 6, 11, 23, 46, 98, \ldots$

Rooted Binary Trees

- B_n= set of rooted inequivalent binary trees with n leaves
- $|B_n| \longrightarrow 1, 1, 1, 2, 3, 6, 11, 23, 46, 98, \ldots$

Examples.

- (1), (2), (3) represent the unique rooted binary trees for $n = 1, 2, 3$ respectively.

Rooted Binary Trees

- B_n= set of rooted inequivalent binary trees with n leaves
- $|B_n| \longrightarrow 1, 1, 1, 2, 3, 6, 11, 23, 46, 98, \ldots$

Examples.

- (1), (2), (3) represent the unique rooted binary trees for $n = 1, 2, 3$ respectively.

- $B_4 = \{((1)(3)),\ ((2)(2))\}$,
- $B_5 = \{((1)((1)(3))),\ ((1)((2)(2))),\ ((2)(3))\}$,
- $((1)(((1)((1)((1)(3))))(((2)(2))(((1)(3))((2)(3))))))$ is in B_{20}. $|B_{20} = 293, 547|$

Catalan objects

- C_n= set of plane rooted binary trees with n leaves
- $|C_n| \longrightarrow 1, 1, 2, 5, 14, 42, \ldots$

Example.

- ((1)(2)) and ((2)(1)) are distinct as plane trees.

Automorphism Groups of Rooted Binary Trees

- Let $T \in B_n$ rooted binary tree with n leaves.
- $A(T)$ is the automorphism group of T given a canonical labeling of its leaves.

Example. $T = ((1)((2)(2)))$ generated by 3 involutions

$$[1,3,2,4,5], \ [1,2,3,5,4], \ [1,4,5,2,3]$$

$$|| \qquad\qquad || \qquad\qquad ||$$

$$(2\ 3) \qquad\qquad (4\ 5) \qquad\qquad (2\ 4)(3\ 5)$$

$|A(T)| = 2^3 = 8.$

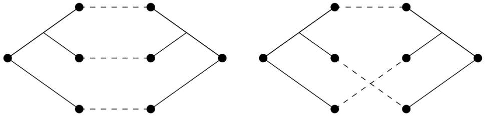
Tanglegrams

Defn. An *(ordered binary rooted) tanglegram* of size n is a triple (T, w, S) where $S, T \in B_n$ and $w \in S_n$.

Two tanglegrams (T, w, S) and (T', w', S') are equivalent provided $T = T'$, $S = S'$ and $w' \in A(T)wA(S)$.

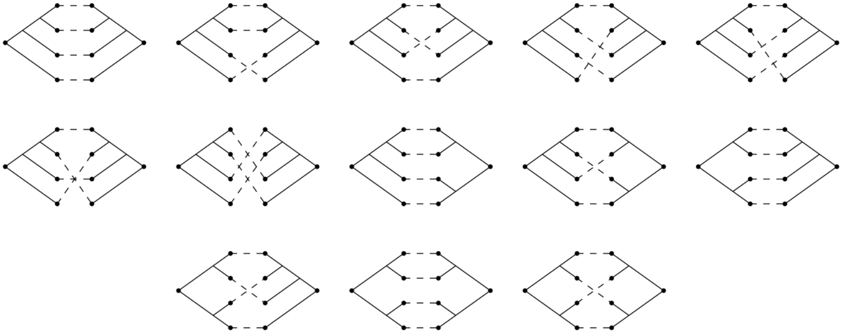
- T_n= set of inequivalent tanglegrams with n leaves
- $t_n = |T_n| \longrightarrow 1, 1, 2, 13, 114, 1509, 25595, 535753, \ldots$

Example. $n = 3$, $t_3 = 2$



Tanglegrams

Case $n = 4,\ t_4 = 13$:



Enumeration of Tanglegrams

Questions.(Matsen) How many tanglegrams are in T_n?
How does t_n grow asymptotically?

First formula.:

$$t_n = \sum_{S \in B_n} \sum_{T \in B_n} \sum_{w \in S_n} \frac{1}{|A(T)wA(S)|}$$

.

This formula allowed us to get data up to $n = 10$. Sequence wasn't in OEIS.

Motivation to study tanglegrams

Cophylogeny Estimation Problem in Biology.: Reconstruct the history of genetic changes in a host vs parasite or other linked groups of organisms.

Tanglegram Layout Problem in CS.: Find a drawing of a tanglegram in the plane with planar embeddings of the left and right trees and a minimal number of crossing (straight) edges in the matching. Eades-Wormald (1994) showed this is NP-hard.

Tanglegrams appear in analysis of software development in CS.

Main Enumeration Theorem

Thm 1. The number of tanglegrams of size n is

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

summed over *binary partitions* of n.

Defn. A *binary partition* $\lambda = (\lambda_1 \geq \lambda_1 \geq \ldots)$ has each part $\lambda_k = 2^j$ for some $j \in \mathbb{N}$.

Defn. $z_\lambda = 1^{m_0} 2^{m_1} 4^{m_2} \cdots (2^j)^{m_j} m_0! m_1! m_2! \cdots m_j!$
for $\lambda = 1^{m_0} 2^{m_1} 4^{m_2} 8^{m_3} \cdots$.

The numbers z_λ are famous!

Defn. More generally, $z_\lambda = 1^{m_1} 2^{m_2} 3^{m_3} \cdots j^{m_j} m_1! m_2! m_2! \cdots m_j!$ for $\lambda = 1^{m_1} 2^{m_2} 3^{m_3} \cdots$.

Facts.:

1. The number of permutations in S_n of cycle type λ is $\dfrac{n!}{z_\lambda}$.
2. If $v \in S_n$ has cycle type λ, then z_λ is the size of the stabilizer of v under the conjugation of S_n on itself.
3. For fixed $u, v \in S_n$ of cycle type λ,

$$z_\lambda = \#\{w \in S_n \mid wvw^{-1} = u|\}.$$

4. The symmetric function $h_n(X) = \sum_\lambda \dfrac{p_\lambda(X)}{z_\lambda}$.

Main Enumeration Theorem

Thm. The number of tanglegrams of size n is

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

summed over *binary partitions* of n and z_λ.

Example. The 4 binary partitions of $n = 4$ are

λ :	(4)	(22)	(211)	(1111)
z_λ :	4	$2^2 2!$	$1^2 2^1 2!$	$1^4 4!$

$$t_4 = \frac{1}{4} + \frac{3^2}{8} + \frac{3^2 \cdot 1^2}{4} + \frac{5^2 \cdot 3^2 \cdot 1^2}{24} = 13$$

Corollaries

Cor 1. $t_n = \frac{c_{n-1}^2 n!}{4^{n-1}} \sum_\mu \frac{n(n-1)\cdots(n-|\mu|+1)}{z_\mu \cdot \prod_{i=1}^{\ell(\mu)} \prod_{j=1}^{\mu_i - 1} (2n-2(\mu_1+\cdots+\mu_{i-1})-2j-1)^2}$,
summed is over binary partitions μ with all parts equal to a positive power of 2 and $|\mu| \leq n$.

Cor 2.: As n gets large, $\frac{t_n}{n!} \sim \frac{e^{\frac{1}{8}} 4^{n-1}}{\pi n^3}$.

Cor 3.: There is an efficient recurrence relation for t_n based on stripping off all copies of the largest part of λ.
We can compute t_{4000} exactly.

Second Enumeration Theorem

Thm 2. The number of binary trees in B_n is

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

summed over *binary partitions* of n.

Second Enumeration Theorem

Thm 2. The number of binary trees in B_n is

$$b_n = \sum_{\lambda} \frac{\prod_{i=2}^{\ell(\lambda)} \left(2(\lambda_i + \cdots + \lambda_{\ell(\lambda)}) - 1\right)}{z_\lambda},$$

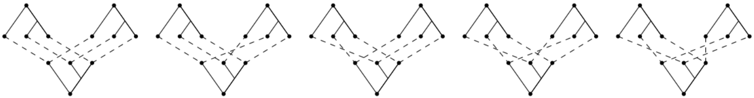
summed over *binary partitions* of n.

Question. What if the exponent k is bigger than 2?

$$t(k,n) = \sum_{\lambda} \frac{\prod_{i=2}^{\ell(\lambda)} \left(2(\lambda_i + \cdots + \lambda_{\ell(\lambda)}) - 1\right)^k}{z_\lambda}.$$

Tangled Chains

Defn. A *tangled chain* of size n and length k is an ordered sequence of binary trees with complete matchings between the leaves of neighboring trees in the sequence.



Thm 3. The number of tangled chains of size n and length k is

$$t(k,n) = \sum_{\lambda} \frac{\prod_{i=2}^{\ell(\lambda)} \Big(2(\lambda_i + \cdots + \lambda_{\ell(\lambda)}) - 1\Big)^k}{z_\lambda}.$$

Outline of Proof of Theorem 1

$$t_n = \sum_{S \in B_n} \sum_{T \in B_n} \sum_{w \in S_n} \frac{1}{|A(T)wA(S)|}$$

Outline of Proof of Theorem 1

$$t_n = \sum_{S \in B_n} \sum_{T \in B_n} \sum_{w \in S_n} \frac{1}{|A(T)wA(S)|}$$

For S, T fixed

$$|A(T)wA(S)| = \frac{|A(T)| \cdot |A(S)|}{|A(T) \cap wA(S)w^{-1}|}$$

Outline of Proof of Theorem 1

$$t_n = \sum_{S \in B_n} \sum_{T \in B_n} \sum_{w \in S_n} \frac{1}{|A(T)wA(S)|}$$

For S, T fixed

$$|A(T)wA(S)| = \frac{|A(T)| \cdot |A(S)|}{|A(T) \cap wA(S)w^{-1}|}$$

$$\sum_{w \in S_n} |A(T) \cap wA(S)w^{-1}| = \sum_{w \in S_n} \sum_{u \in A(T)} \sum_{v \in A(S)} \chi[u = wvw^{-1}]$$

Outline of Proof of Theorem 1

$$t_n = \sum_{S \in B_n} \sum_{T \in B_n} \sum_{w \in S_n} \frac{1}{|A(T)wA(S)|}$$

For S, T fixed

$$|A(T)wA(S)| = \frac{|A(T)| \cdot |A(S)|}{|A(T) \cap wA(S)w^{-1}|}$$

$$\sum_{w \in S_n} |A(T) \cap wA(S)w^{-1}| = \sum_{w \in S_n} \sum_{u \in A(T)} \sum_{v \in A(S)} \chi[u = wvw^{-1}]$$

$$= \sum_{u \in A(T)} \sum_{v \in A(S)} \sum_{w \in S_n} \chi[u = wvw^{-1}]$$

Outline of Proof of Theorem 1

$$t_n = \sum_{S \in B_n} \sum_{T \in B_n} \sum_{w \in S_n} \frac{1}{|A(T)wA(S)|}$$

For S, T fixed

$$|A(T)wA(S)| = \frac{|A(T)| \cdot |A(S)|}{|A(T) \cap wA(S)w^{-1}|}$$

$$\sum_{w \in S_n} |A(T) \cap wA(S)w^{-1}| = \sum_{w \in S_n} \sum_{u \in A(T)} \sum_{v \in A(S)} \chi[u = wvw^{-1}]$$

$$= \sum_{u \in A(T)} \sum_{v \in A(S)} \sum_{w \in S_n} \chi[u = wvw^{-1}]$$

$$= \sum_{\lambda \vdash n} |A(T)_\lambda| \cdot |A(S)_\lambda| \cdot z_\lambda$$

where $A(T)_\lambda = \{w \in A(T) \mid type(w) = \lambda\}$. Only binary partitions occur!

Outline of Proof of Main Theorem

$$t_n = \sum_{S \in B_n} \sum_{T \in B_n} \sum_{w \in S_n} \frac{1}{|A(T)wA(S)|}$$

$$= \sum_{S \in B_n} \sum_{T \in B_n} \sum_{\lambda} \frac{|A(T)_\lambda| \cdot |A(S)_\lambda| \cdot z_\lambda}{|A(T)| \cdot |A(S)|}$$

Outline of Proof of Main Theorem

$$t_n = \sum_{S \in B_n} \sum_{T \in B_n} \sum_{w \in S_n} \frac{1}{|A(T)wA(S)|}$$

$$= \sum_{S \in B_n} \sum_{T \in B_n} \sum_{\lambda} \frac{|A(T)_\lambda| \cdot |A(S)_\lambda| \cdot z_\lambda}{|A(T)| \cdot |A(S)|}$$

$$= \sum_{\lambda} z_\lambda \left(\sum_{T \in B_n} \frac{|A(T)_\lambda|}{|A(T)|} \right)^2$$

Outline of Proof of Main Theorem

$$t_n = \sum_{S \in B_n} \sum_{T \in B_n} \sum_{w \in S_n} \frac{1}{|A(T)wA(S)|}$$

$$= \sum_{S \in B_n} \sum_{T \in B_n} \sum_{\lambda} \frac{|A(T)_\lambda| \cdot |A(S)_\lambda| \cdot z_\lambda}{|A(T)| \cdot |A(S)|}$$

$$= \sum_{\lambda} z_\lambda \left(\sum_{T \in B_n} \frac{|A(T)_\lambda|}{|A(T)|} \right)^2$$

To show:

$$\sum_{T \in B_n} \frac{|A(T)_\lambda|}{|A(T)|} = \frac{\prod_{i=2}^{\ell(\lambda)} \left(2(\lambda_i + \cdots + \lambda_{\ell(\lambda)}) - 1\right)}{z_\lambda} =: q_\lambda$$

via the recurrence

$$2q_\lambda = q_{\lambda/2} + \sum_{\lambda^1 \cup \lambda^2 = \lambda} q_{\lambda^1} q_{\lambda^2}$$

Conclusion: $t_n = \sum z_\lambda q_\lambda^2$.

Random Generation of Tanglegrams

Input: n

Step 1: Pick a binary partition $\lambda \vdash n$ with prob $z_\lambda q_\lambda^2 / t_n$.

Step 2: Choose T and $u \in A(T)_\lambda$ uniformly by subdividing $\lambda = \lambda^1 \cup \lambda^2$ according to the recurrence for q_λ. Similarly, choose S and $v \in A(T)_\lambda$ uniformly by subdividing.

Step 3: Among the z_λ permutations w such that $u = wvw^{-1}$, pick one uniformly.

Output: (T, w, S).

Random Generation of a Permutation in $A(T)$

Input: Binary tree $T \in B_n$ with left and right subtrees T_1 and T_2.

If $n = 1$, set $w = (1) \in A(T)$, unique choice.

Otherwise, recursively find $w_1 \in A(T_1)$ and $w_2 \in A(T_2)$ at random.

- If $T_1 \neq T_2$, set $w = w_1 w_2$.
- If $T_1 = T_2$, choose either $w = w_1 w_2$ or $w = \pi w_1 w_2$ with equal probability.

Here $\pi = (1\ k)(2\ (k+1))(3\ (k+3)) \cdots (k\ n)$ where $k = n/2$ flips the labels on the leaves of the two subtrees.

Output: Permutation $w \in A(T)$.

Random Generation of Tanglegrams:Step 2

Step 2: Choose T and $u \in A(T)_\lambda$ uniformly by subdividing $\lambda = \lambda^1 \cup \lambda^2$ according to the recurrence for q_λ.

Input: $\lambda \vdash n$.

- If $n = 1$, output $T = \bullet$, $u = (1) \in A(T)$, unique choice.
- Otherwise, pick a subdivision of λ from two types

$$\{(\lambda/2, \lambda/2)\} \bigcup \{(\lambda^1, \lambda^2) : \lambda^1 \cup \lambda^2 = \lambda\}$$

with probability proportional to

$$q_{\lambda/2} + \sum q_{\lambda^1} q_{\lambda^2} = 2q_\lambda.$$

Random Generation of Tanglegrams:Step 2

Step 2: Choose T and $u \in A(T)_\lambda$ uniformly by subdividing $\lambda = \lambda^1 \cup \lambda^2$ according to the recurrence for q_λ.

- Type 1: $(\lambda/2, \lambda/2)$. Use the algorithm recursively to compute $T_1 \in B_{n/2}$ and a permutations $u_2 \in A(T_1)_{\lambda/2}$. Uniformly at random, generate anther permutation $u_1 \in A(T_1)$. Set

$$T = (T_1, T_1),\ u = \pi u_1 \pi u_1^{-1} \pi u_2.$$

- Type 2: (λ^1, λ^2). Use the algorithm recursively to compute trees T_1, T_2 and permutations $u_1 \in A(T_1)_{\lambda^1}$ $u_2 \in A(T_2)_{\lambda^2}$. Switch if necessary so $T_1 \leq T_2$. Set

$$T = (T_1, T_2),\ u = u_1 u_2.$$

Output: (T, u).

Random Generation of Tanglegrams:Step 2

Example If $\lambda = (6,4)$, then $|\lambda| = 10$, $\lambda/2 = (3,2)$ and $\pi = (1\ 6)(2\ 7)(3\ 8)(4\ 9)(5\ 10)$. If

$$w_1 = (1\ 4)(2\ 5)(3) \text{ and } w_2 = (6\ 9\ 7)(8\ 10)$$

then

$$w = \pi w_1 \pi w_1^{-1} \pi w_2 = (6\ 1\ 9\ 5\ 7\ 4)(8\ 2\ 10\ 3),$$

all in cycle notation.

Review: Random Generation of Tanglegrams

Input: n

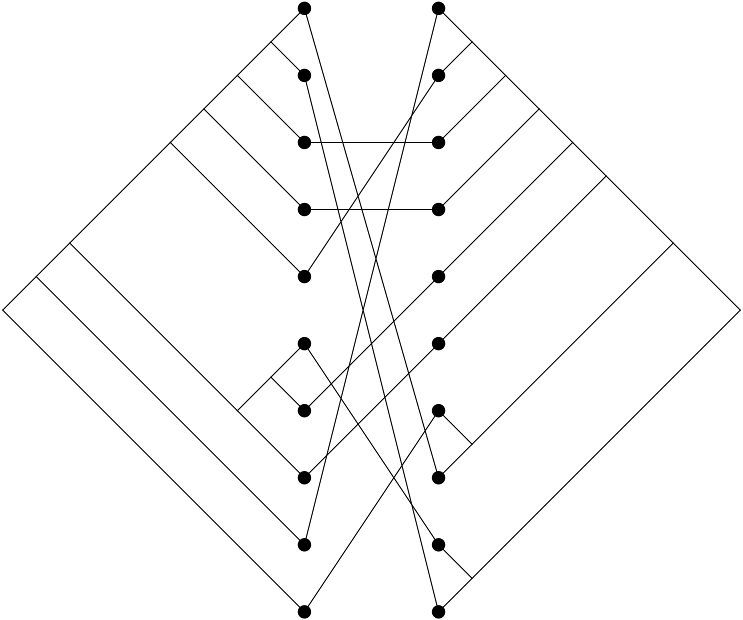
Step 1: Pick a binary partition $\lambda \vdash n$ with prob $z_\lambda q_\lambda^2 / t_n$.

Step 2: Choose T and $u \in A(T)_\lambda$ uniformly by subdividing $\lambda = \lambda^1 \cup \lambda^2$ according to the recurrence for q_λ. Similarly, choose S and $v \in A(T)_\lambda$ uniformly by subdividing.

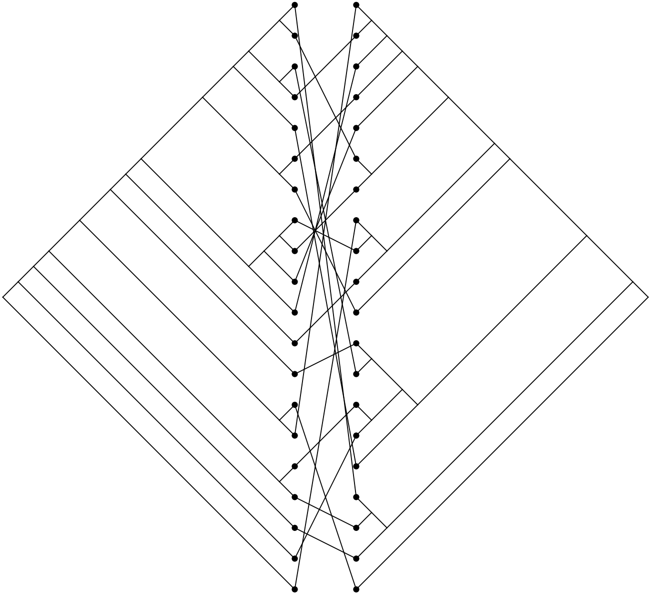
Step 3: Among the z_λ permutations w such that $u = wvw^{-1}$, pick one uniformly.

Output: (T, w, S).

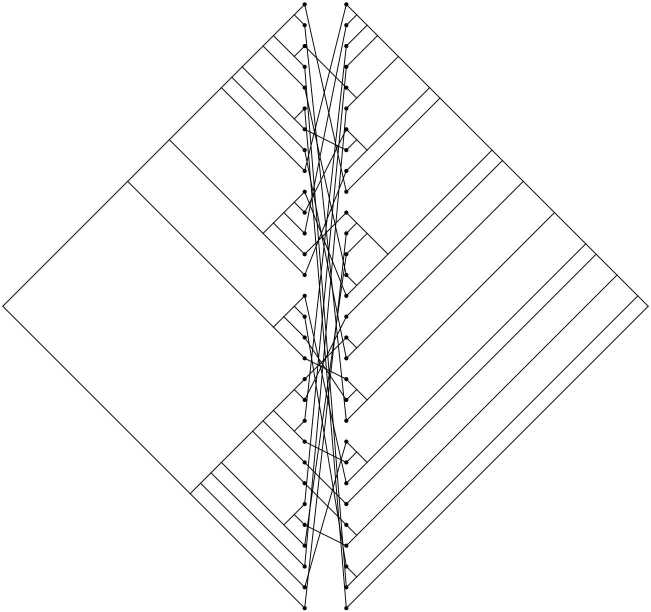
Random Tanglegrams: n=10



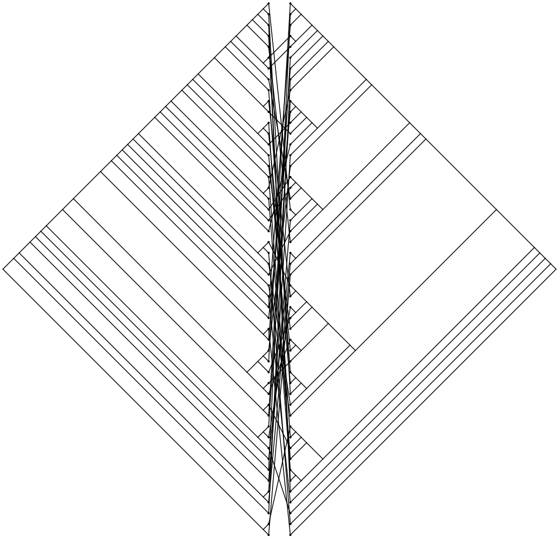
Random Tanglegrams: n=20



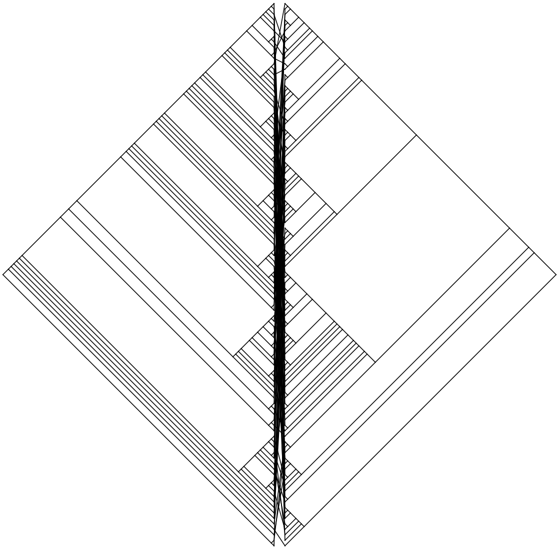
Random Tanglegrams: n=30



Random Tanglegrams: n=50



Random Tanglegrams: n=100



Positivity and symmetric functions
go hand in hand with enumeration.

This is a story that began with an enumeration question and via work of Gessel now connects to symmetric functions, plethysm of Schur functions, and Kronecker coefficients.

Open Problems

1. Is there a closed form or functional equation for $T(x) = \sum t_n x^n$ like there is for binary trees $B(x)$?

$$B(x) = x + \frac{1}{2}\left(B(x)^2 + B(x^2)\right)$$

2. Is there an efficient algorithm for depth first search on tanglegrams?

3. Can one describe the lex minimal permutations in the double cosets $A(T)\backslash S_n/A(S)$ for $S, T \in B_n$?