Compositions, canonical trees, acyclic digraphs and their common structural properties

Stephan Wagner

Stellenbosch University

AofA, Menorca, 28 May 2013

based on joint work with Clemens Heuberger and Daniel Krenn

Compositions

Definition

A *composition* of n is a representation of n as an ordered sum of positive integers: e.g.,

$$5 + 3 + 1 + 2 + 3 + 1$$

is a composition of 15.

Compositions

Definition

A *composition* of n is a representation of n as an ordered sum of positive integers: e.g.,

$$5 + 3 + 1 + 2 + 3 + 1$$

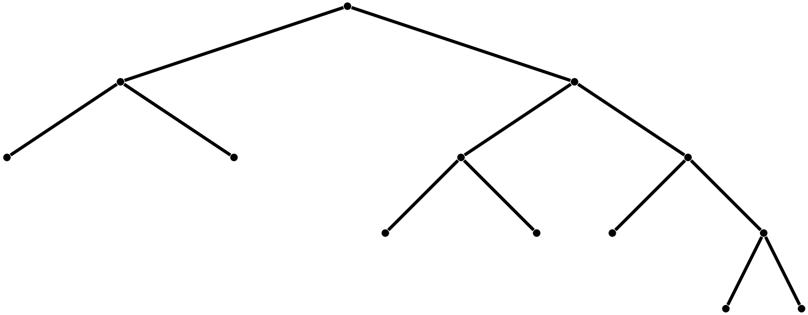
is a composition of 15.

It is well known that there are 2^{n-1} compositions of n. The length is asymptotically normally distributed with mean $\frac{n+1}{2}$ and variance $\frac{n-1}{4}$, the largest summand is typically around $\log_2 n$, ...

Canonical trees

Definition

We call a rooted plane t-ary tree *canonical* if the vertex degrees are weakly increasing from left to right, as in the following example:



Canonical trees

Canonical t-ary trees are in bijection with *canonical compact prefix-free codes*: these are t-ary codes (codes over an alphabet of size t) such that:

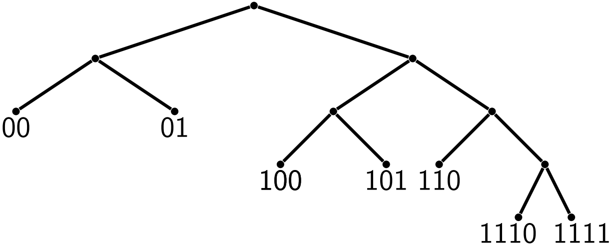
- no word in the code is a proper prefix of another word (prefix-free),
- no word can be added to the code so that it remains prefix-free (compact),
- the lexicographic ordering corresponds to a non-decreasing ordering of word-lengths (canonical).

Canonical trees

Canonical t-ary trees are in bijection with *canonical compact prefix-free codes*: these are t-ary codes (codes over an alphabet of size t) such that:

- no word in the code is a proper prefix of another word (prefix-free),
- no word can be added to the code so that it remains prefix-free (compact),
- the lexicographic ordering corresponds to a non-decreasing ordering of word-lengths (canonical).

The following picture illustrates the bijection:



Canonical trees

Canonical trees are also in bijection with *partitions* of 1 into powers of t, i.e., representations of the form

$$1 = t^{-a_1} + t^{-a_2} + \cdots + t^{-a_n}$$

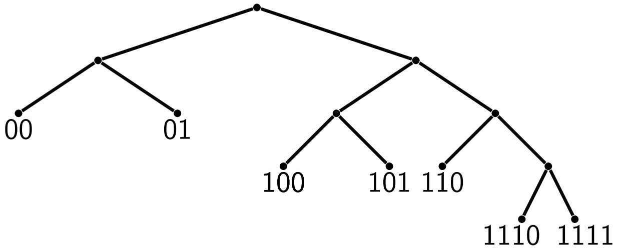
with $a_1 \leq a_2 \leq \cdots \leq a_n$.

Canonical trees

Canonical trees are also in bijection with *partitions* of 1 into powers of t, i.e., representations of the form

$$1 = t^{-a_1} + t^{-a_2} + \cdots + t^{-a_n}$$

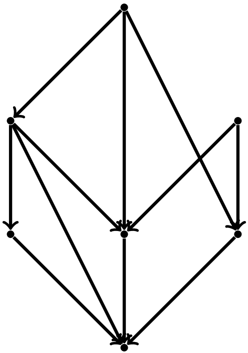
with $a_1 \leq a_2 \leq \cdots \leq a_n$.



$$1 = 2^{-2} + 2^{-2} + 2^{-3} + 2^{-3} + 2^{-3} + 2^{-4} + 2^{-4}$$

Acyclic digraphs

A digraph is called *acyclic* if it does not contain a directed cycle.



An acyclic digraph is thus (in some sense) the directed analogue of a forest.

Generating functions and asymptotics

The generating function for canonical t-ary trees with a given number of internal vertices is

$$C(x) = \frac{\sum_{j \geq 0} (-1)^j x^{[j]} \prod_{i=1}^{j} \frac{x^{[i]}}{1-x^{[i]}}}{\sum_{j \geq 0} (-1)^j \prod_{i=1}^{j} \frac{x^{[i]}}{1-x^{[i]}}},$$

where $[j] = 1 + t + \cdots + t^{j-1}$.

Generating functions and asymptotics

The generating function for canonical t-ary trees with a given number of internal vertices is

$$C(x) = \frac{\sum_{j\geq 0}(-1)^j x^{[j]} \prod_{i=1}^{j} \frac{x^{[i]}}{1-x^{[i]}}}{\sum_{j\geq 0}(-1)^j \prod_{i=1}^{j} \frac{x^{[i]}}{1-x^{[i]}}},$$

where $[j] = 1 + t + \cdots + t^{j-1}$. Singularity analysis yields an asymptotic formula:

Theorem (Boyd 1975, Komlos/Moser/Nemetz 1984, Flajolet/Prodinger 1987)

The number of canonical t-ary trees is asymptotically equal to $K_t \rho_t^n$ for some constants K_t and ρ_t. Moreover, $\rho_t \to 2$ as $t \to \infty$.

Generating functions and asymptotics

For acyclic digraphs, one needs a special type of generating function: if a_n is the number of labelled acyclic digraphs with n vertices, then

$$A(x) = \sum_{n=0}^{\infty} \frac{a_n}{n! 2^{\binom{n}{2}}} = \frac{1}{\sum_{n \geq 0} \frac{(-1)^n x^n}{n!} 2^{-\binom{n}{2}}}.$$

Generating functions and asymptotics

For acyclic digraphs, one needs a special type of generating function: if a_n is the number of labelled acyclic digraphs with n vertices, then

$$A(x) = \sum_{n=0}^{\infty} \frac{a_n}{n! 2^{\binom{n}{2}}} = \frac{1}{\sum_{n \geq 0} \frac{(-1)^n x^n}{n!} 2^{-\binom{n}{2}}}.$$

Again, singularity analysis strikes:

Theorem (Robinson 1971)

The number of labelled acyclic digraphs of order n is

$$a_n \sim K n! \cdot 2^{\binom{n}{2}} z_0^{-n},$$

where $z_0 \approx 1.488079$.

A first connection

A canonical t-ary tree is uniquely determined if we know the number a_k of internal vertices at distance k from the root for all k. Clearly, $a_0 = 1$ (unless the root is the only vertex) and

$$a_k \leq t a_{k-1}.$$

A first connection

A canonical t-ary tree is uniquely determined if we know the number a_k of internal vertices at distance k from the root for all k. Clearly, $a_0 = 1$ (unless the root is the only vertex) and

$$a_k \leq t a_{k-1}.$$

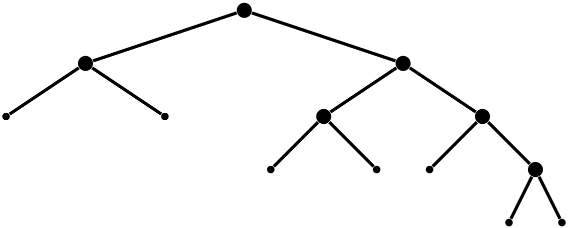
This means that canonical t-ary trees are also in bijection with *restricted compositions*: compositions that satisfy the inequality stated above.

A first connection

A canonical t-ary tree is uniquely determined if we know the number a_k of internal vertices at distance k from the root for all k. Clearly, $a_0 = 1$ (unless the root is the only vertex) and

$$a_k \leq t a_{k-1}.$$

This means that canonical t-ary trees are also in bijection with *restricted compositions*: compositions that satisfy the inequality stated above.



$$1 + 2 + 2 + 1 = 6$$

A first connection

As $t \to \infty$, the restriction

$$a_k \leq ta_{k-1}$$

disappears, and we get ordinary compositions. This also explains why the constant ρ_t in the asymptotic formula $K_t \rho_t^n$ for the number of canonical t-ary trees has to tend to 2.

A first connection

As $t \to \infty$, the restriction

$$a_k \leq t a_{k-1}$$

disappears, and we get ordinary compositions. This also explains why the constant ρ_t in the asymptotic formula $K_t \rho_t^n$ for the number of canonical t-ary trees has to tend to 2.

There is a series of papers on properties of compositions with various local restrictions by Bender, Canfield and Gao (2005 –).

Transfer matrices

Let $C_k(x)$ be the generating function for canonical t-ary trees with the property that there are exactly k vertices with maximum distance from the root. Then

$$C_k(x) = x^k \sum_{j=\lceil \frac{k}{t} \rceil}^{\infty} C_j(x) + [k = 1]x.$$

This is a linear system of equations with an *infinite* transfer matrix.

Transfer matrices

Let $C_k(x)$ be the generating function for canonical t-ary trees with the property that there are exactly k vertices with maximum distance from the root. Then

$$C_k(x) = x^k \sum_{j=\lceil \frac{k}{t} \rceil}^{\infty} C_j(x) + [k=1]x.$$

This is a linear system of equations with an *infinite* transfer matrix. For $t = 2$, the matrix looks like this:

$$\begin{bmatrix} x & x & x & x & x & \cdots \\ x^2 & x^2 & x^2 & x^2 & x^2 & \cdots \\ 0 & x^3 & x^3 & x^3 & x^3 & \cdots \\ 0 & x^4 & x^4 & x^4 & x^4 & \cdots \\ 0 & 0 & x^5 & x^5 & x^5 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Transfer matrices

A similar idea works for acyclic digraphs: every such digraph has a number of *sinks* (vertices of outdegree 0). For each vertex v, let the *level* be the length of the longest directed path starting at v (so that the level of a sink is 0). If $a_{n,k}$ is the number of labelled acyclic digraphs of order n with k vertices at the highest level, then we have the recursion

$$a_{n,k} = \binom{n}{k} \sum_{j=1}^{\infty} a_{n-k,j} 2^{k(n-k)} (1 - 2^{-j})^k$$

with $a_{k,k} = 1$ for $k \geq 1$ and $a_{n,k} = 0$ for $n < k$.

Transfer matrices

A similar idea works for acyclic digraphs: every such digraph has a number of *sinks* (vertices of outdegree 0). For each vertex v, let the *level* be the length of the longest directed path starting at v (so that the level of a sink is 0). If $a_{n,k}$ is the number of labelled acyclic digraphs of order n with k vertices at the highest level, then we have the recursion

$$a_{n,k} = \binom{n}{k} \sum_{j=1}^{\infty} a_{n-k,j} 2^{k(n-k)} (1-2^{-j})^k$$

with $a_{k,k} = 1$ for $k \geq 1$ and $a_{n,k} = 0$ for $n < k$. For the generating functions $A_k(x) = \sum_{n \geq 0} a_{n,k} 2^{-\binom{n}{2}} \frac{x^n}{n!}$, this means

$$A_k(x) = \frac{x^k}{k!} 2^{-\binom{k}{2}} \left(1 + \sum_{j=1}^{\infty} (1-2^{-j})^k A_j(x) \right).$$

Transfer matrices

A similar idea works for acyclic digraphs: every such digraph has a number of *sinks* (vertices of outdegree 0). For each vertex v, let the *level* be the length of the longest directed path starting at v (so that the level of a sink is 0). If $a_{n,k}$ is the number of labelled acyclic digraphs of order n with k vertices at the highest level, then we have the recursion

$$a_{n,k} = \binom{n}{k} \sum_{j=1}^{\infty} a_{n-k,j} 2^{k(n-k)} (1 - 2^{-j})^k$$

with $a_{k,k} = 1$ for $k \geq 1$ and $a_{n,k} = 0$ for $n < k$. For the generating functions $A_k(x) = \sum_{n \geq 0} a_{n,k} 2^{-\binom{n}{2}} \frac{x^n}{n!}$, this means

$$A_k(x) = \frac{x^k}{k!} 2^{-\binom{k}{2}} \left(1 + \sum_{j=1}^{\infty} (1 - 2^{-j})^k A_j(x) \right).$$

This can again be seen as a system of linear equations with an infinite transfer matrix.

How to proceed from there

The infinite matrices can be inverted – a priori only in the ring of formal power series, but it can be shown that one actually obtains analytic functions.

How to proceed from there

The infinite matrices can be inverted – a priori only in the ring of formal power series, but it can be shown that one actually obtains analytic functions.

We get a good heuristic idea of the shape by regarding the number of vertices at each level as an approximate *Markov chain*.

How to proceed from there

The infinite matrices can be inverted – a priori only in the ring of formal power series, but it can be shown that one actually obtains analytic functions.

We get a good heuristic idea of the shape by regarding the number of vertices at each level as an approximate *Markov chain*.

The similarities in the functional equations explain why shape parameters of canonical trees and acyclic digraphs behave in a similar way (which is quite different from other random tree models such as Galton-Watson trees or recursive trees).

Some structural results: height

Theorem (McKay 1989)

The height (longest directed path) of a random labelled (or unlabelled) acyclic digraph has a Gaussian limiting distribution with linear mean and variance.

Theorem (Heuberger/Krenn/SW 2013)

The height of a random canonical t-ary tree has a Gaussian limiting distribution with linear mean and variance.

Some structural results: width

Theorem (Krenn/SW 2013+)

The width (largest number of vertices on a single level) of a random acyclic digraph of order n is concentrated around $C_1\sqrt{\log n}$.

Theorem (Heuberger/Krenn/SW 2013)

The width of a random canonical t-ary tree of order n is concentrated around $C_2 \log n$.

Some structural results: profile

For the profile, we obtain the following behaviour for both acyclic digraphs and canonical t-ary trees:

Some structural results: profile

For the profile, we obtain the following behaviour for both acyclic digraphs and canonical t-ary trees:

- For fixed ℓ, the number of vertices at level ℓ asymptotically follows a discrete limiting distribution that depends on ℓ.

Some structural results: profile

For the profile, we obtain the following behaviour for both acyclic digraphs and canonical t-ary trees:

- For fixed ℓ, the number of vertices at level ℓ asymptotically follows a discrete limiting distribution that depends on ℓ.
- If $\ell = H - m$ for fixed m, where H denotes the height, then the number of vertices at level ℓ asymptotically follows a discrete limiting distribution that depends on m.

Some structural results: profile

For the profile, we obtain the following behaviour for both acyclic digraphs and canonical t-ary trees:

- For fixed ℓ, the number of vertices at level ℓ asymptotically follows a discrete limiting distribution that depends on ℓ.
- If $\ell = H - m$ for fixed m, where H denotes the height, then the number of vertices at level ℓ asymptotically follows a discrete limiting distribution that depends on m.
- If $1 \ll \ell \ll H$, then the number of vertices at level ℓ asymptotically follows a discrete limiting distribution that does not depend on ℓ.