Patterns in Trees

Thomas Klausner
TU Wien
Joint work with Michael Drmota

INRIA Rocquencourt, December 9, 2002

Supported by the Austrian Science Fund, Project S8302-MAT.

Outline

- Present assumptions and basic techniques.
- Define patterns.
- Find representations as combinatorial objects.
- Convert to corresponding generating functions.
- Compute asymptotics.

1. Basics

Labelled universe

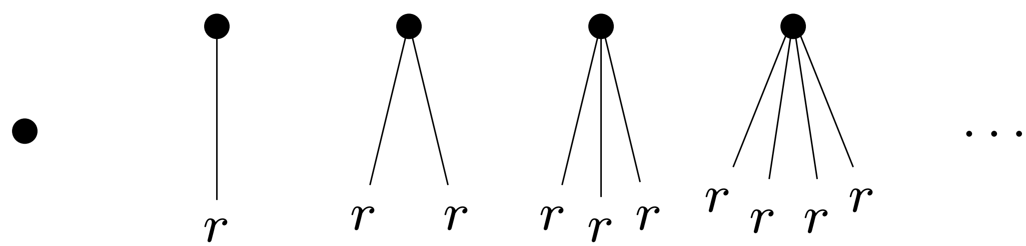
Counting of combinatorical structures by exponential generating functions

$$p(z) = \sum p_n \frac{z^n}{n!}$$

First model: All trees of size n have equal probability.

Rooted Trees

Construction (rooted):



Tree function:

$$p(z) \;=\; z\left(1 + p(z) + \frac{p(z)^2}{2!} + \frac{p(z)^3}{3!} + \dots\right) = ze^{p(z)}$$

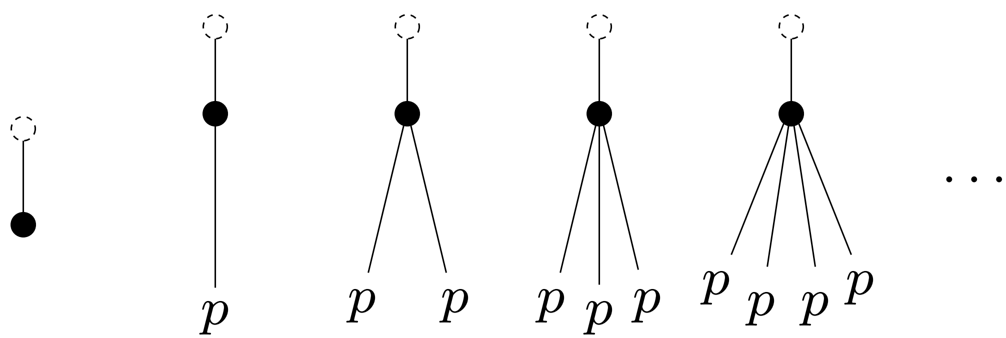
Explicit number of rooted trees of size n via e.g. Lagrange inversion formula

$$p_n = n^{n-1}$$

Planted Rooted Trees

Sometimes helpful: planted rooted trees.

Node degree does not change during construction.



Results in same function $p(z)$.

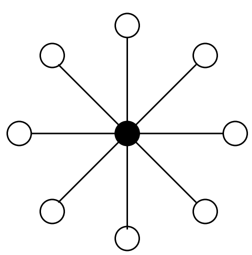
First order asymptotics

$$p(z) \; = \; 1 - \sqrt{2}\sqrt{1 - ez} + O(z)$$

Bivariate Generating Functions

Mark special properties with a second variable u

E.g.: Nodes with particular degree $k+1$



$$p \;=\; ze^p - \frac{p^k}{k!} + u\frac{p^k}{k!} = ze^p + (u-1)\frac{p^k}{k!}$$

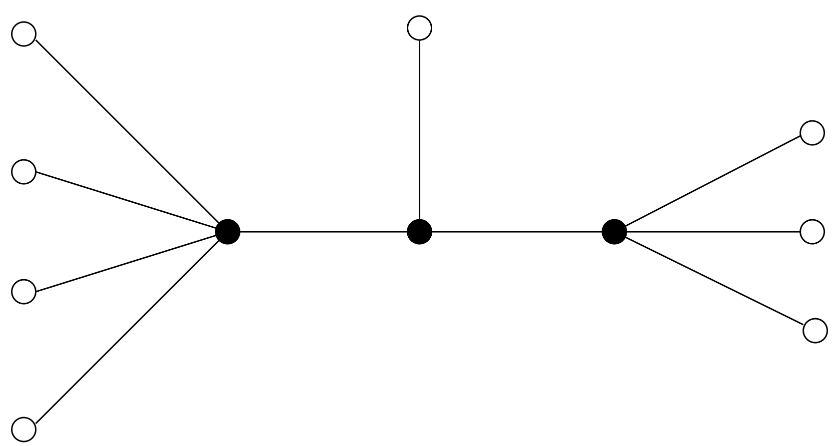
Possible to find limit distribution and compute parameters. In this case [Drmota and Gittenberger 1997]: asymptotically Gaussian with mean $\mu_k n$ and variance $\sigma_k^2 n$, where $m_k = \frac{1}{ek!}$ und $\sigma_k = -\frac{1+(k-2)^2}{e^2(k-1)!^2} + \frac{1}{e(k-1)!}$

2. Patterns

What is a pattern?

In our case, connected sub-tree $\mathcal{M}$.

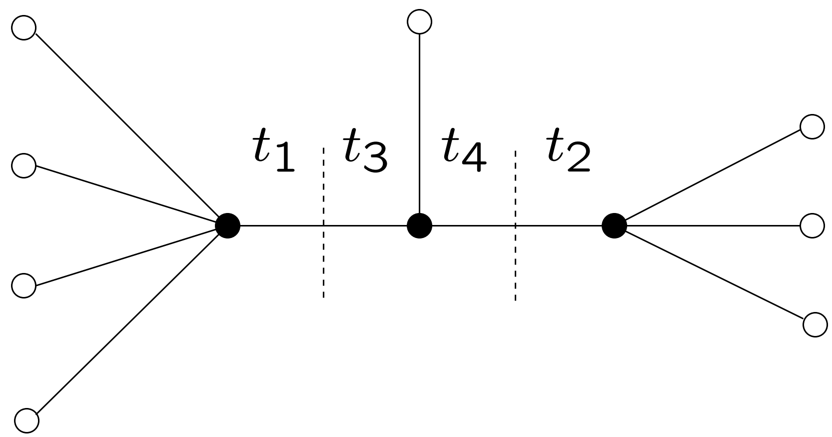
Easy example:



How to Mark a Pattern?

More difficult than with single nodes.

Split in parts:



Aim: Describe patterns by generating functions for each part
$\longrightarrow$ system of functional equations.

Proposition (Planted Rooted Trees)

Let $\mathcal{M}$ be a pattern. Then there exist L auxiliary functions $a_j(x,u)$ $(1 \le j \le L)$ with

$$p(x,u) = \sum_{j=1}^{L} a_j(x,u)$$

and polynomials $P_j(y_1,\ldots,y_L,u)$ $(1 \le j \le L-1)$ with non-negative coefficients such that

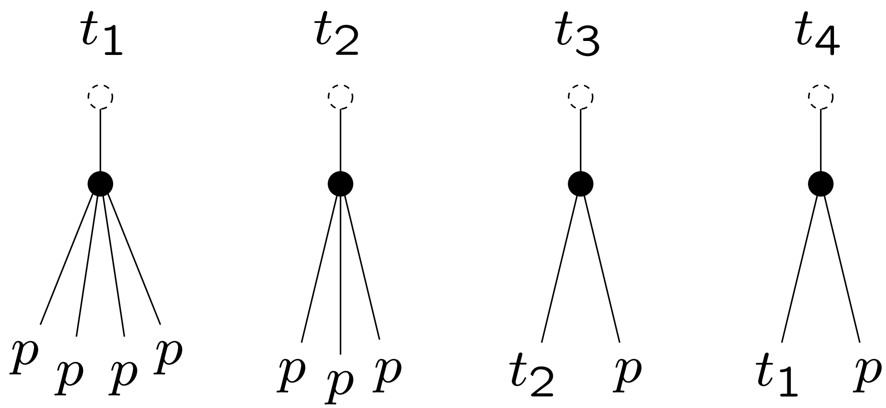
$$\begin{aligned} a_1(x,u) &= x \cdot P_1(a_1(x,u),\ldots,a_L(x,u),u) \\ &\vdots \\ a_{L-1}(x,u) &= x \cdot P_{L-1}(a_1(x,u),\ldots,a_L(x,u),u) \qquad (1) \\ a_L(x,u) &= xe^{a_1(x,u)+\cdots+a_L(x,u)} - x\sum_{j=1}^{L-1} P_j(a_1(x,u),\ldots,a_L(x,u),1). \end{aligned}$$

Notation to Describe Patterns

$\circ$ is a (planted) root node, $\times$ the Cartesian product, $\cap$ the intersection, $\cup$ the union, and $\dot{\cup}$ the disjunct union.

$\times$ binds stronger than either of $\cap$, $\cup$, and $\dot{\cup}$.

Note: no immediate one-to-one correspondance to the generating functions (relative probabilities, u).



$$
\begin{aligned}
t_1 &= \{\circ\} \times p \times p \times p \times p, \\
t_2 &= \{\circ\} \times p \times p \times p, \\
t_3 &= \{\circ\} \times t_2 \times p, \\
t_4 &= \{\circ\} \times t_1 \times p.
\end{aligned}
$$

3. Convert to Functions

- Standardise: no duplicate descriptions of the same structure
- Find coefficients
- Sprinkle with u

Standard Form

Each tree a_j represented as disjoint union of trees of the kind

$$\{\circ\} \times a_{l_1} \times \cdots \times a_{l_d}, \tag{2}$$

(d = degree of the root of a_j).

Standardising the Functions

Build intersections (symbolically):

$$\begin{aligned} x_1 &= \{\circ\} \times x_{1,1} \times \cdots \times x_{1,l} \\ x_2 &= \{\circ\} \times x_{2,1} \times \cdots \times x_{2,l} \\ x_1 \cap x_2 &= \bigcup_{\substack{\{m_1,\ldots,m_l\}=\{1,\ldots,l\} \\ \{n_1,\ldots,n_l\}=\{1,\ldots,l\}}} \{\circ\} \times (x_{1,m_1} \cap x_{2,n_1}) \times \cdots \times (x_{1,m_l} \cap x_{2,n_l}) \\ x_1 \cup x_2 &= x_1 \dot{\cup} x_2 \setminus (x_1 \cap x_2) \end{aligned}$$

Intersection Example

$$\begin{aligned} t_1 &= \{\circ\} \times p \times p \times p \times p, \\ t_2 &= \{\circ\} \times p \times p \times p, \\ t_3 &= \{\circ\} \times t_2 \times p, \\ t_4 &= \{\circ\} \times t_1 \times p. \end{aligned}$$

Only one non-empty intersection:

$$\begin{aligned} t_3 \cap t_4 &= \{\circ\} \times (t_1 \cap t_2) \times p \cup \{\circ\} \times t_1 \times t_2 = \\ &= \{\circ\} \times t_1 \times t_2. \end{aligned}$$

Example in Standardised Form

$$
\begin{aligned}
a_1 &= t_3 \setminus (t_3 \cap t_4) = \{\circ\} \times a_5 \times (a_1 \cup a_2 \cup a_3 \cup a_5 \cup a_6), \\
a_2 &= t_4 \setminus (t_3 \cap t_4) = \{\circ\} \times a_4 \times (a_1 \cup a_2 \cup a_3 \cup a_4 \cup a_6), \\
a_3 &= t_3 \cap t_4 = \{\circ\} \times a_4 \times a_5, \\
a_4 &= t_1 = \{\circ\} \times \bigcup_{i=1}^{6} a_i \times \bigcup_{i=1}^{6} a_i \times \bigcup_{i=1}^{6} a_i \times \bigcup_{i=1}^{6} a_i, \\
a_5 &= t_2 = \{\circ\} \times \bigcup_{i=1}^{6} a_i \times \bigcup_{i=1}^{6} a_i \times \bigcup_{i=1}^{6} a_i, \\
a_6 &= p \setminus (a_1 \cup \cdots \cup a_5) = \\
&= \{\circ\} \cup \{\circ\} \times \bigcup_{i=1}^{6} a_i \cup \{\circ\} \times (a_1 \cup a_2 \cup a_3 \cup a_6) \times (a_1 \cup a_2 \cup a_3 \cup a_6) \cup \\
&\quad \cup \bigcup_{n=5}^{\infty} \left(\{\circ\} \times \underbrace{\bigcup_{i=1}^{6} a_i \times \bigcup_{i=1}^{6} a_i \times \cdots \times \bigcup_{i=1}^{6} a_i}_{n} \right).
\end{aligned}
$$

Coefficients

$A_{j,l_1,\dots,l_L,k}$:= number of possible configurations of type (2)
(l_i sub-trees of type a_i, k new occurrences of $\mathcal{M}$)

Coefficients are computed by simple combinatorics (k is implicitly given by l_i).

Resulting Functions

$$P_j(y_1,\ldots,y_L,u) = \sum_{l_1,\ldots,l_L \ge 0} \frac{A_{j,l_1,\ldots,l_L,k}}{l_1! \cdots l_L!} y_1^{l_1} \cdots y_L^{l_L} u^k, \qquad 1 \le j \le L-1$$

$$P_L(y_1,\ldots,y_L) = e^{y_1+\cdots+y_L} - \sum_{j=1}^{L-1} P_j(y_1,\ldots,y_L,1).$$

$$a_j(x,u) = x \cdot P_j(a_1(x,u),\ldots,a_L(x,u),u).$$

$\rightarrow$ proposed structure of the system of functional equations (1).

How to Find $k = k(l_1, \ldots, l_L)$

New patterns occur when all necessary sub-trees are attached to a node of proper degree.

In example, three cases:

1. Node of degree three with a t_1 and t_2.

2. Node of degree four with a t_4 attached. Each t_4 produces another pattern.

3. Node of degree five with a t_3 attached. Each t_3 produces another pattern.

Coefficients in Example

$$
\begin{aligned}
P_1 &= y_5(y_1 + y_2 + y_3 + y_6) + \frac{1}{2!}y_5^2, \\
P_2 &= y_4(y_1 + y_2 + y_3 + y_6) + \frac{1}{2!}y_4^2, \\
P_3 &= uy_4y_5, \\
P_4 &= \frac{(uy_1 + y_2 + uy_3 + y_4 + y_5 + y_6)^4}{4!}, \\
P_5 &= \frac{(y_1 + uy_2 + uy_3 + y_4 + y_5 + y_6)^3}{3!}
\end{aligned}
$$

Coefficients in Example (cont'd)

$$
\begin{aligned}
a_1(x,u) = a_1 &= x a_5(a_1 + a_2 + a_3 + a_6) + x\frac{1}{2!}a_5^2, \\
a_2(x,u) = a_2 &= x a_4(a_1 + a_2 + a_3 + a_6) + x\frac{1}{2!}a_4^2, \\
a_3(x,u) = a_3 &= x u a_4 a_5, \\
a_4(x,u) = a_4 &= x\frac{(ua_1 + a_2 + ua_3 + a_4 + a_5 + a_6)^4}{4!}, \\
a_5(x,u) = a_5 &= x\frac{(a_1 + ua_2 + ua_3 + a_4 + a_5 + a_6)^3}{3!}, \\
a_6(x,u) = a_6 &= x + x\sum_{i=1}^{6} a_i + x\frac{(a_1 + a_2 + a_3 + a_6)^2}{2!} + \\
&\quad + x\sum_{n=5}^{\infty}\frac{1}{n!}\left(\sum_{i=1}^{6} a_i\right)^n .
\end{aligned}
$$

Strong Connectivity

a_6 depends on itself and all others (last term).

Each a_i depends on a_6 either directly, or through a chain to a leaf (see pattern).

Proposition (Rooted Trees)

There exists an analytic function

$$G(x, u, a_1, \ldots, a_L)$$

with non-negative Taylor coefficients such that

$$r(x, u) = G(x, u, a_1(x, u), \ldots, a_L(x, u)).$$

where the a_i were defined earlier.

Counting Patterns in Rooted Trees

For the example:

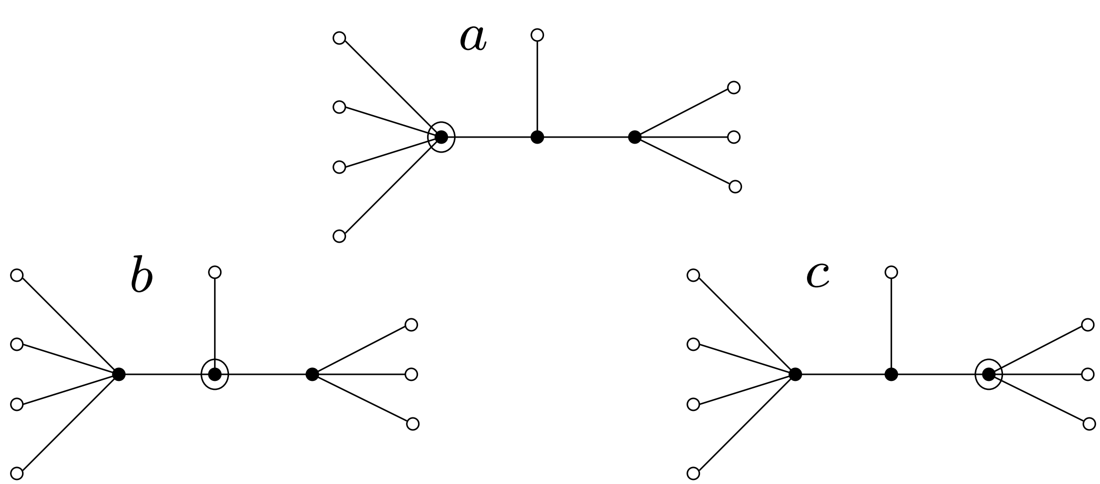
$$r_0 := xe^p - \frac{xp^5}{5!} - \frac{xp^4}{4!} - \frac{xp^3}{3!}.$$

(“uninteresting” sub-trees)

Marks in sub-trees already counted correctly, only have to distribute marks for newly appearing patterns.

Newly Appearing Patterns

$$r = r(x,u) = r_0 + x \sum_{\substack{\sum e_i = 5 \\ e_i \geq 0}} u^{e_1+e_3} \frac{\prod_{j=1}^{6} a_j^{e_j}}{\prod e_i!} + \\ + x \sum_{\substack{\sum e_i = 3 \\ e_i \geq 0}} u^{e_4 e_5} \frac{\prod_{j=1}^{6} a_j^{e_j}}{\prod e_i!} + x \sum_{\substack{\sum e_i = 4 \\ e_i \geq 0}} u^{e_2+e_3} \frac{\prod_{j=1}^{6} a_j^{e_j}}{\prod e_i!}$$



a
b
c

Unrooted Trees

$$t_{n,m} = r_{n,m}/n$$

4. Asymptotics

Use Drmota's Theorem on systems of functional equations. *

Paraphrased: Under certain conditions for the system of equations, the coefficients asymptotically follow a Gaussian distribution with mean and variance asymptotically proportional to n.

Often the difficult part:

Finding the singularity

*M. Drmota, *Systems of Functional Equations*. Random Structures and Algorithms 10, 103-124, 1997.

For Our Case

Singularity known:

$$\frac{1}{e}$$

To find: left eigenvector (for the eigenvalue one) of the derivative of the functional matrix with respect to the functions

Application of theorem then gives expectation for examples as

$$\frac{384e - 19}{12e(576e^3 + 24e^2 - 25)} = 0.0026803\ldots$$

Possible Future Extensions

- Different kind of trees or tree distributions

- Logical terms using $\vee$, $\wedge$, $\neg$, $\forall$ and $\exists$ describing more general patterns.

- Does a 0-1-law hold?