PATTERNS IN RANDOM TREES

Michael Drmota*

Institute of Discrete Mathematics and Geometry

Vienna University of Technology

A 1040 Wien, Austria

michael.drmota@tuwien.ac.at

http://www.dmg.tuwien.ac.at/drmota/

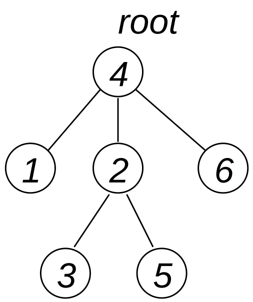
* supported by the Austrian Science Foundation FWF, National Research Network *Analytic Combinatorics and Probabilistic Number Theory*, grant S96.

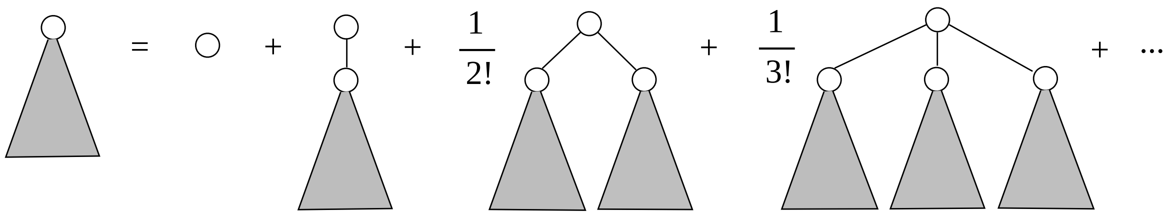
Contents

Contents

Patterns in Trees

Cayley Trees: rooted labelled trees





$$\mathcal{R} = \circ + \circ * \mathcal{R} + \frac{1}{2!} \circ * \mathcal{R} * \mathcal{R} + \frac{1}{3!} * \circ \mathcal{R} * \mathcal{R} * \mathcal{R} + \cdots$$

Patterns in Trees

Generating functions

r_n ... number of **rooted** labelled trees with n nodes

$$R(x) = \sum_{n \geq 1} r_n \frac{x^n}{n!}$$

$$\mathcal{R} = \circ + \circ * \mathcal{R} + \frac{1}{2!} \circ * \mathcal{R} * \mathcal{R} + \frac{1}{3!} \circ * \mathcal{R} * \mathcal{R} * \mathcal{R} + \cdots$$

$$R(x) = x + xR(x) + \frac{1}{2!} xR(x)^2 + \frac{1}{3!} xR(x)^3 + \cdots$$

$$\boxed{R(x) = xe^{R(x)}}$$

Patterns in Trees

Cayley's formula (derived with Lagrange inversion)

$$r_n = n!\,\frac{1}{n}\,[u^{n-1}]\,e^{un} = n^{n-1}$$

$$\boxed{r_n = n^{n-1}}$$

t_n ... number of **unrooted** labelled trees with n nodes ($= r_n/n$)

$$\boxed{t_n = n^{n-2}}$$

Patterns in Trees

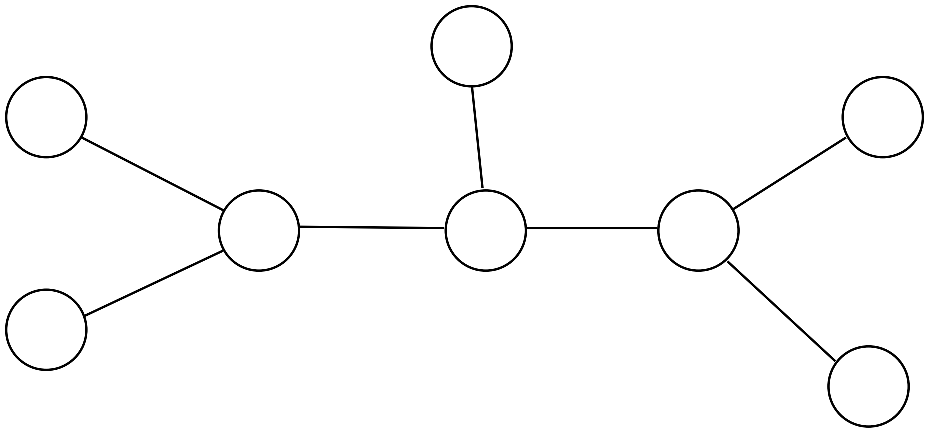
Probabilistic Model

Every unrooted labelled tree τ with n nodes is equally likely

$$\boxed{\mathbb{P}\{\tau \text{ occurs}\} = \frac{1}{n^{n-2}}}$$

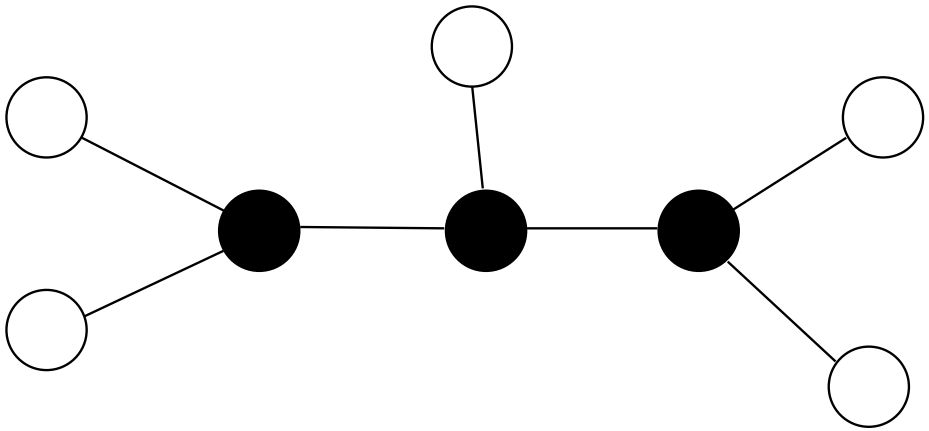
Patterns in Trees

Pattern $\mathcal{M}$



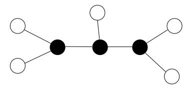
Patterns in Trees

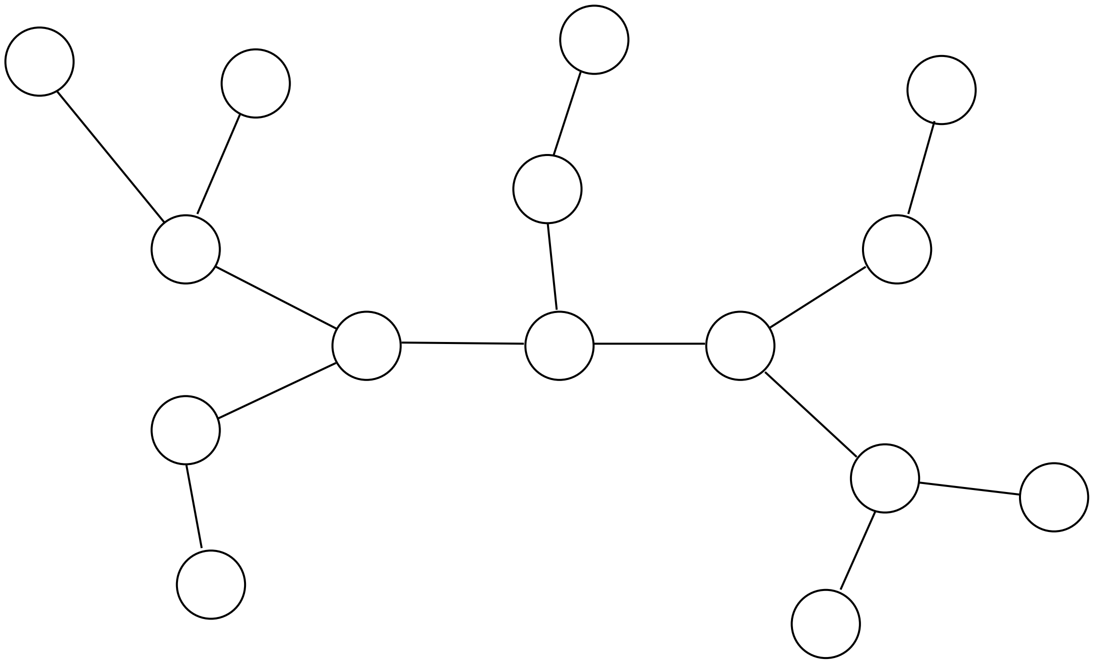
Pattern $\mathcal{M}$



Patterns in Trees

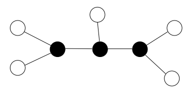
Occurence of a pattern $\mathcal{M}$

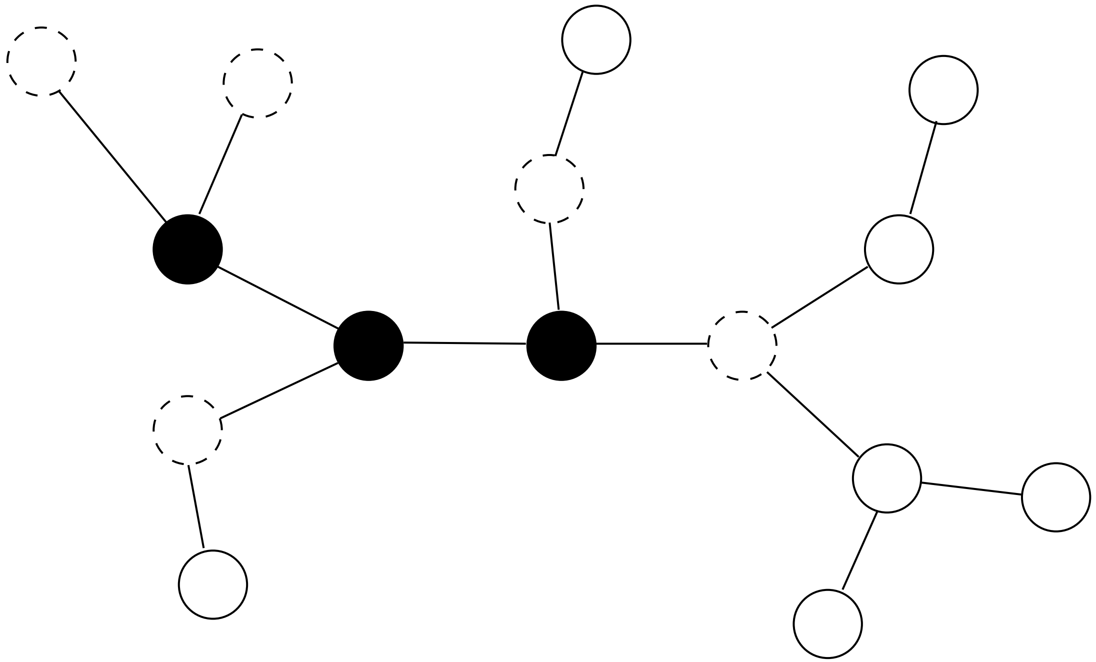




Patterns in Trees

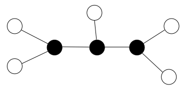
Occurence of a pattern $\mathcal{M}$

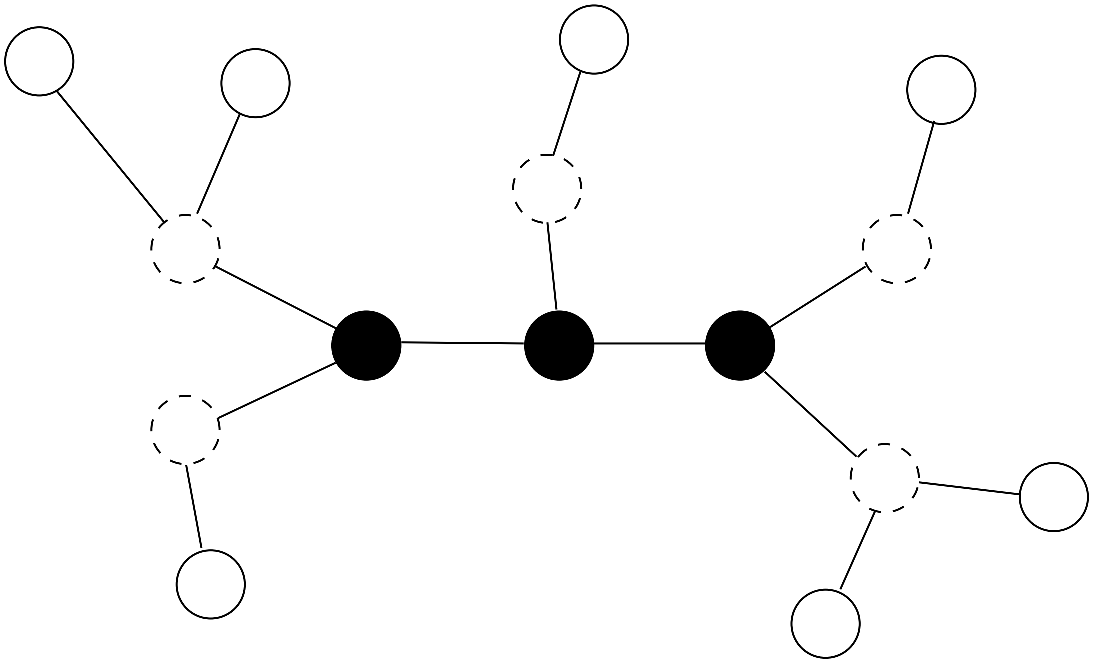




Patterns in Trees

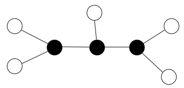
Occurence of a pattern $\mathcal{M}$

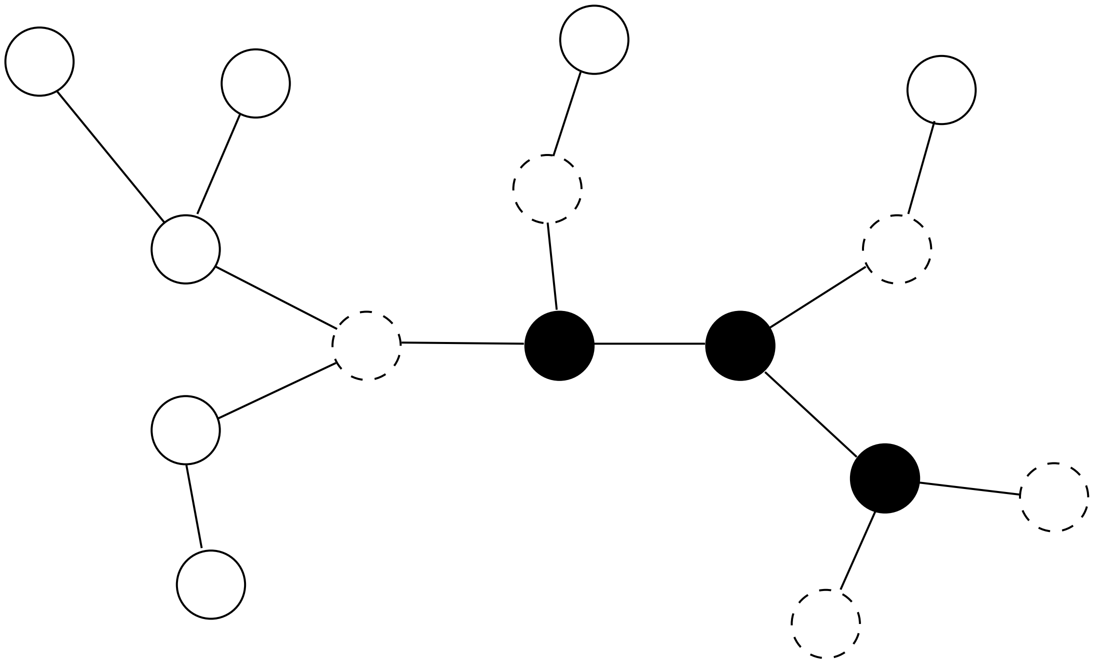




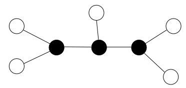
Patterns in Trees

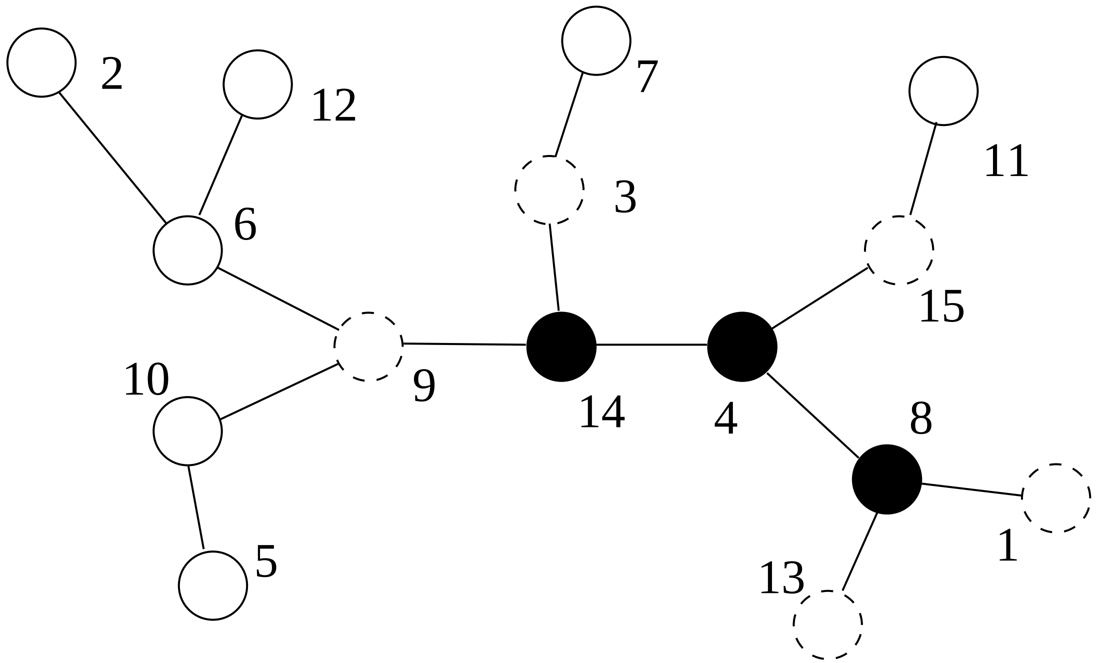
Occurence of a pattern $\mathcal{M}$





Patterns in Trees

Occurence of a pattern $\mathcal{M}$  **in a labelled tree**



Patterns in Trees

Theorem (Chyzak & D. & Klausner & Kok, CPC '08)

$\mathcal{M}$ ... be a given finite tree.

X_n ... number of occurences of of $\mathcal{M}$ in a labelled tree of size n

$\Longrightarrow$ X_n satisfies a **central limit theorem** with

$$\mathbb{E}\, X_n \sim \mu n \quad \text{and} \quad \mathbb{V}\, X_n \sim \sigma^2 n.$$

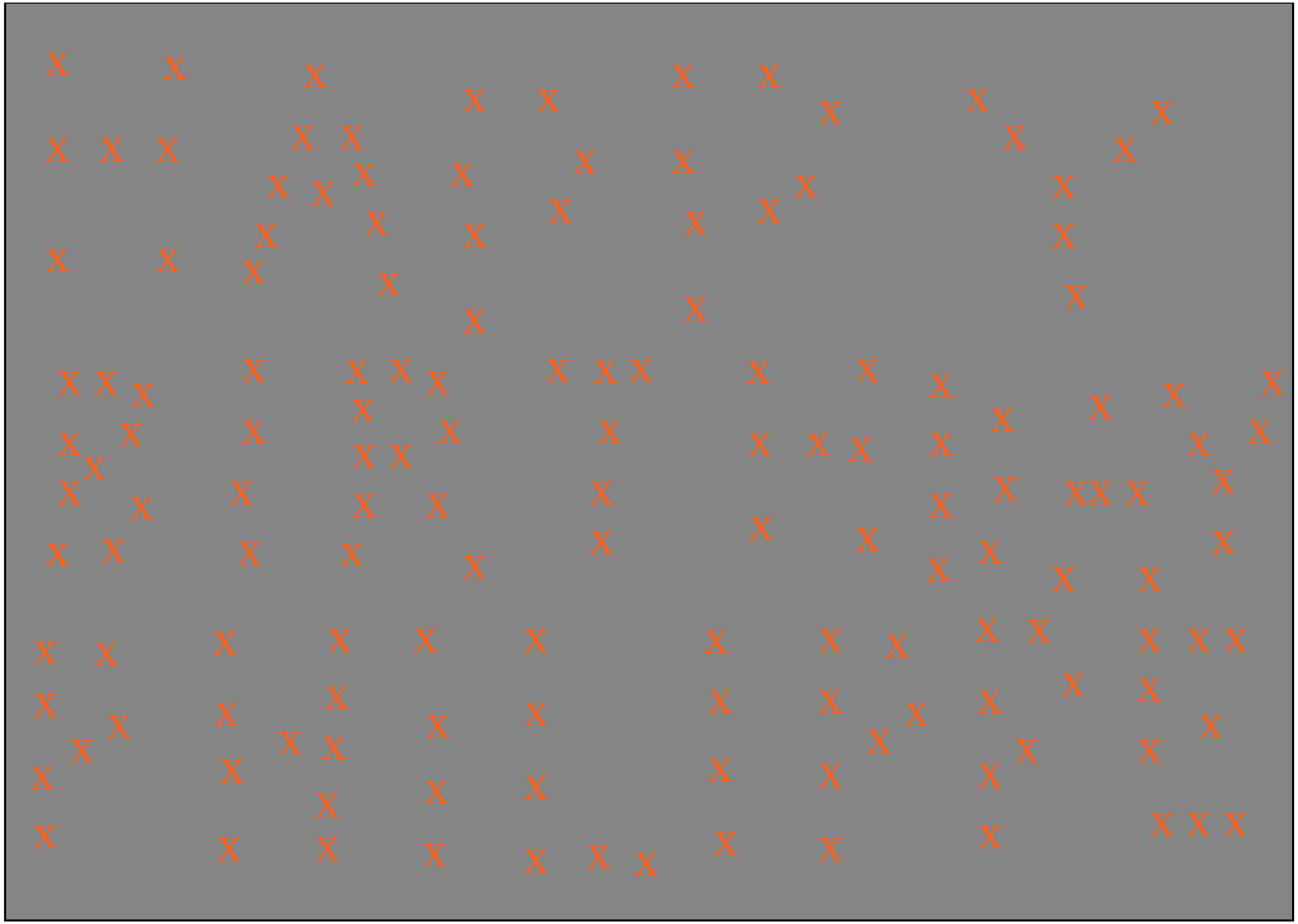
$\mu > 0$ and $\sigma^2 \geq 0$ depend on the pattern $\mathcal{M}$ and can be computed explicitly and algorithmically and can be represented as polynomials (with rational coefficients) in $1/e$.

Patterns in Trees

Sum of weakly dependent random variables

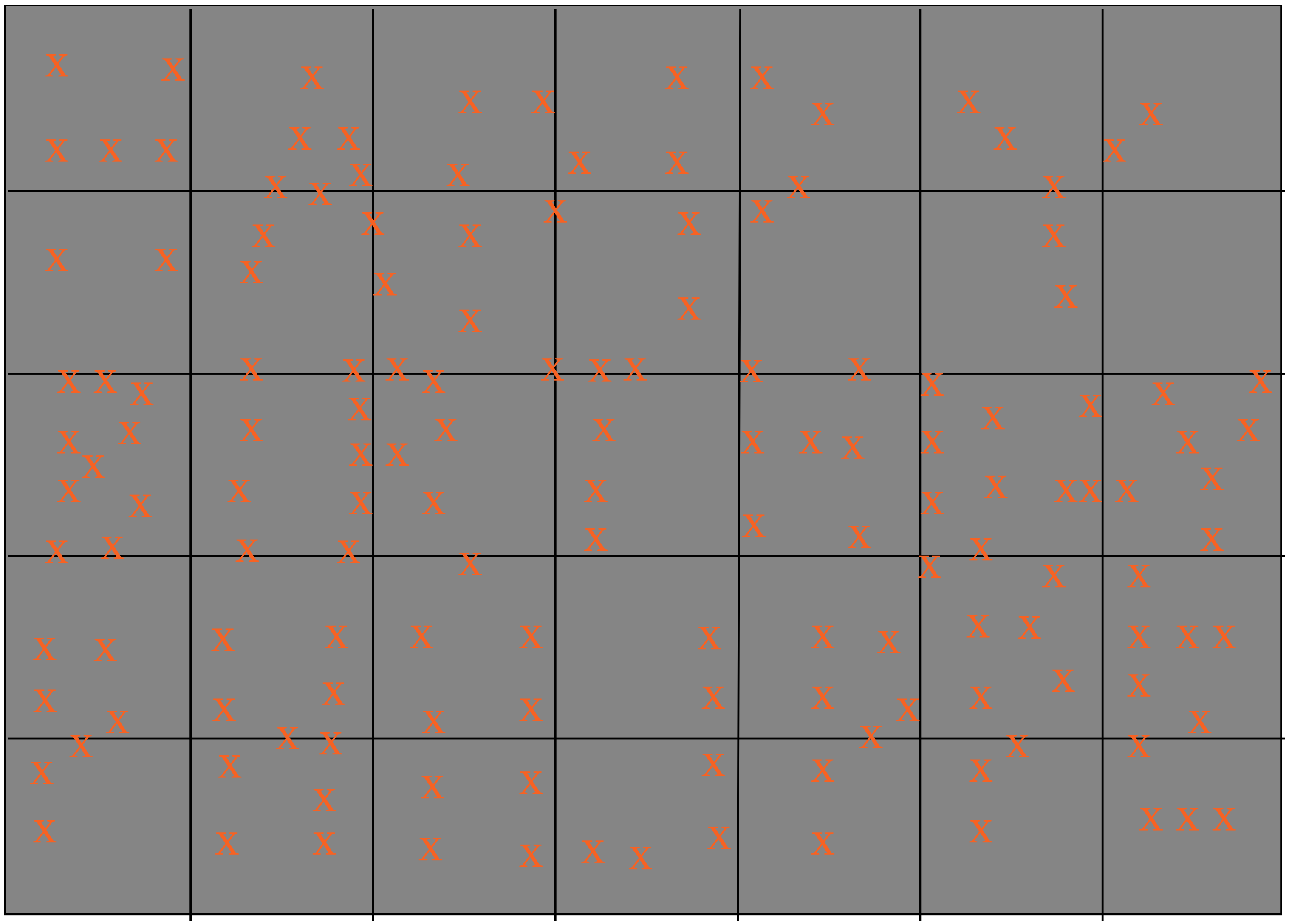
Patterns in Trees

Sum of weakly dependent random variables

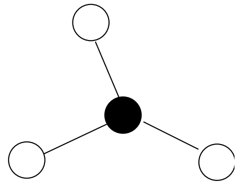


Patterns in Trees

Sum of weakly dependent random variables



Functional equations



Number of nodes of degree 3
= number of nodes of out-degree 2

$r_{n,m}$ ... number of **rooted** labelled trees with n nodes and m nodes of out-degree 2

$$R(x,u) = \sum_{n,m} r_{n,m} \frac{x^n}{n!} u^m$$

$$\mathcal{R} = \circ + \circ * \mathcal{R} + \frac{1}{2!} \bullet * \mathcal{R} * \mathcal{R} + \frac{1}{3!} \circ * \mathcal{R} * \mathcal{R} * \mathcal{R} + \cdots$$

$$R(x,u) = x + xR(x,u) + \boxed{u} \frac{1}{2!} xR(x,u)^2 + \frac{1}{3!} xR(x,u)^3 + \cdots$$

$$\boxed{R(x,u) = xu \frac{R(x,u)^2}{2!} + x\left(e^{R(x,u)} - \frac{R(x,u)^2}{2!}\right)}$$

Functional equations

Recursive structure leads to functional equation for gen. func.:

$$\boxed{A(x,u) = \Phi(x,u,A(x,u))}$$

Functional equations

Theorem (Bender, Canfield, Meir & Moon, D.)

Suppose that $\boxed{A(x,u) = \Phi(x,u,A(x,u))}$, where $\Phi(x,u,a)$ has a power series expansion at $(0,0,0)$ with non-negative coefficients and $\Phi_{aa}(x,u,a) \ne 0$.

Let $x_0 > 0$, $a_0 > 0$ (inside the region of convergence) satisfy the system of equations:

$$\boxed{a_0 = \Phi(x_0,1,a_0), \quad 1 = \Phi_a(x_0,1,a_0)}.$$

Then there exists analytic function $g(x,u), h(x,u)$, and $\rho(u)$ such that locally

$$\boxed{A(x,u) = g(x,u) - h(x,u)\sqrt{1 - \frac{x}{\rho(u)}}}.$$

Functional equations

Idea of the Proof.

Set $F(x,u,a) = \Phi(x,u,a) - a$. Then we have

$$\begin{aligned} F(x_0,1,a_0) &= 0 \\ F_a(x_0,1,a_0) &= 0 \\ F_x(x_0,1,a_0) &\neq 0 \\ F_{aa}(x_0,1,a_0) &\neq 0. \end{aligned}$$

Weierstrass preparation theorem implies that there exist analytic functions $H(x,u,a)$, $p(x,u)$, $q(x,u)$ with $H(x_0,1,a_0) \neq 0$, $p(x_0,1) = q(x_0,1) = 0$ and

$$F(x,u,a) = H(x,u,a)\Big((a-a_0)^2 + p(x,u)(a-a_0) + q(x,u)\Big).$$

Functional equations

$$F(x,u,a)=0 \quad \Longleftrightarrow \quad (a-a_0)^2+p(x,u)(a-a_0)+q(x,u)=0.$$

Consequently

$$\begin{aligned} A(x,u) &= a_0-\frac{p(x,u)}{2}\pm\sqrt{\frac{p(x,u)^2}{4}-q(x,u)} \\ &= \boxed{g(x,u)-h(x,u)\sqrt{1-\frac{x}{\rho(u)}}}, \end{aligned}$$

where we write

$$\frac{p(x,u)^2}{4}-q(x,u)=K(x,u)(x-\rho(u))$$

which is again granted by the Weierstrass preparation theorem and we set

$$g(x,u)=a_0-\frac{p(x,u)}{2} \quad \text{and} \quad h(x,u)=\sqrt{-K(x,u)\rho(u)}.$$

Functional equations

A Central Limit Theorem for Functional Equations

Suppose that $\boxed{A(x,u) = \Phi(x,u,A(x,u))}$, where $\Phi(x,u,a)$ has a power series expansion at $(0,0,0)$ with non-negative coefficients and $\Phi_{aa}(x,u,a) \neq 0$ (+ *minor* technical conditions). Set

$$\mu = \frac{x_0 \Phi_x(x_0,1,a_0)}{\Phi(x_0,1,a_0)} \quad \text{and} \quad \sigma^2 = \text{``}long\ formula\text{''}.$$

Then then random variable X_n defined by $\boxed{\mathbb{P}\{X_n = m\} = a_{n,m}/a_n}$ satisfies a **central limit theorem** with

$$\mathbb{E}\, X_n \sim \mu n \quad \text{and} \quad \mathbb{V}\, X_n \sim \sigma^2 n.$$

Remark. $\mathbb{E}\, u^{X_n} = \sum_m \mathbb{P}\{X_n = m\}\, u^m = \frac{[x^n]A(x,u)}{[x^n]A(x,1)}$

Functional equations

Idea of the Proof.

$$A(x,u) = g(x,u) - h(x,u)\sqrt{1 - \frac{x}{\rho(u)}}$$

for certain analytic function $g(x,u), h(x,u)$, and $\rho(u)$.

application of singularity analysis (Flajolet & Odlyzko)

$$\Longrightarrow \quad A_n(u) = [x^n]A(x,u) = \sum_{m\ge 0} a_{n,m}u^m \sim \frac{h(\rho(u),u)\cdot \rho(u)^{-n}\cdot n^{-3/2}}{2\sqrt{\pi}}$$

$$\Longrightarrow \quad \boxed{\mathbb{E}\, u^{X_n} = \frac{A_n(u)}{A_n(1)} \sim \frac{h(\rho(u),u)}{h(\rho(1),1)}\left(\frac{\rho(1)}{\rho(u)}\right)^n}$$

$\Longrightarrow$ **central limit theorem** by Quasi Power Theorem

Functional equations

Number of nodes of degree 3 in Cayley trees

$$R(x,u) = xe^{R(x,u)} + x(u-1)\frac{R(x,u)^2}{2}$$

$$x_0 = \frac{1}{e}, \quad r_0 = R(x_0) = 1.$$

$\Longrightarrow$ **central limit theorem** with

$$\mathbb{E}\, X_n \sim \frac{1}{2e} n \quad \text{and} \quad \mathbb{V}\, X_n \sim \left(\frac{1}{2e} - \frac{1}{2e^2}\right) n.$$

Functional equations

Systems of functional equations

Suppose, that several generating functions

$$A_1(x,u) = \sum_{n,k} a_{1;n,k} u^k x^n, \ \ldots \ , A_r(x,u) = \sum_{n,k} a_{r;n,k} u^k x^n$$

satisfy a **system of non-linear equations**

$$\boxed{A_j(x,u) = \Phi_j(x,u,A_1(x,u),\ldots,A_r(x,u))},$$

where $\Phi_j(x,u,a_1,\ldots,a_r)$ is non-linear in $a_1,\ldots,a_r$ for some j and has a power series expansion at $(0,0,0)$ with non-negative coefficients (for all j).

Let $x_0 > 0$, $\mathbf{a}_0 = (a_{0,0},\ldots,a_{r,0}) > 0$ (inside the region of convergence) satisfy the system of equations: ($\mathbf{\Phi} = (\Phi_1,\ldots,\Phi_r)$)

$$\boxed{\mathbf{a}_0 = \mathbf{\Phi}(x_0,1,\mathbf{a}_0), \quad 0 = \det(\mathbb{I} - \mathbf{\Phi}_{\mathbf{a}}(x_0,1,\mathbf{a}_0)}.$$

Functional equations

Suppose further, that the **dependency graph** of the system $\mathbf{a} = \Phi(x, u, \mathbf{a})$ is **strongly connected**.

Then there exists analytic function $g_j(x,u), h_j(x,u)$, and $\rho(u)$ (that is **independent of** j) such that locally

$$\boxed{A_j(x,u) = g_j(x,u) - h_j(x,u)\sqrt{1 - \frac{x}{\rho(u)}}}.$$

If $\boxed{A(x,u) = \sum_{n,k} a_{n,k} x^n u^k = F(x, u, A_1(x,u), \ldots, A_j(x,u))}$ (for some analytic function F satisfying certain conditions) then then random variable X_n defined by $\boxed{\mathbb{P}\{X_n = m\} = a_{n,m}/a_n}$ satisfies a **central limit theorem** with

$$\mathbb{E}\, X_n \sim \mu n \quad \text{and} \quad \mathbb{V}\, X_n \sim \sigma^2 n.$$

where μ and σ^2 can be computed.

Functional equations

Dependency graph: $G_{\mathbf{\Phi}} = (V, E)$

V ... vertex set $= \{A_1, A_2, \dots, A_r\}$
E ... (directed) edge set:

$$\begin{aligned}(A_i, A_j) \in E :&\Longleftrightarrow A_i(x,u) \text{ depends on } A_j(x,u) \\ &\Longleftrightarrow \Phi_i \text{ depends on } A_j \\ &\Longleftrightarrow \frac{\partial \Phi_i}{\partial a_j} \neq 0.\end{aligned}$$

$$\boxed{G_{\mathbf{\Phi}} \text{ is stongly connected} \Longleftrightarrow \mathbf{\Phi_a} := \left(\frac{\partial \Phi_i}{\partial A_j}\right) \text{ irreducible}}$$

Functional equations

$$\det\left(\mathbb{I} - \mathbf{\Phi}_{\mathbf{a}}(x_0, 1, \mathbf{a}_0)\right) = 0 \quad \Longleftrightarrow \quad \mathbf{\Phi}_{\mathbf{a}} \text{ has dominant eigenvalue } 1$$

Fact

$\mathbf{\Phi}_{\mathbf{a}}$ irreducible

$\Longrightarrow$ Every principle submatrix of $\mathbf{\Phi}_{\mathbf{a}}$ has smaller dominant eigenvalue

(Perron-Frobenius theory for non-negative matrices)

Functional equations

Idea of the proof (reduction to a single equation)

$\mathbf{a} = (A_1, \ldots, A_r) = (A_1, \overline{\mathbf{a}}),\ \mathbf{\Phi} = (\Phi_1, \ldots, \Phi_r) = (\Phi_1, \overline{\mathbf{\Phi}})$

$$\mathbf{a} = \mathbf{\Phi}(\mathbf{a}, x, u) \quad \Longleftrightarrow \quad \begin{array}{ll} A_1 & = \Phi_1(A_1, \overline{\mathbf{a}}, x, u), \\ \overline{\mathbf{a}} & = \overline{\mathbf{\Phi}}(A_1, \overline{\mathbf{a}}, x, u) \end{array}$$

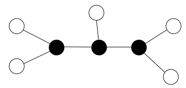
The second system has dominant eigenvalue < 1
$\Longrightarrow \overline{\mathbf{a}} = \overline{\mathbf{a}}(x, u, \boxed{A_1})$ is **analytic**

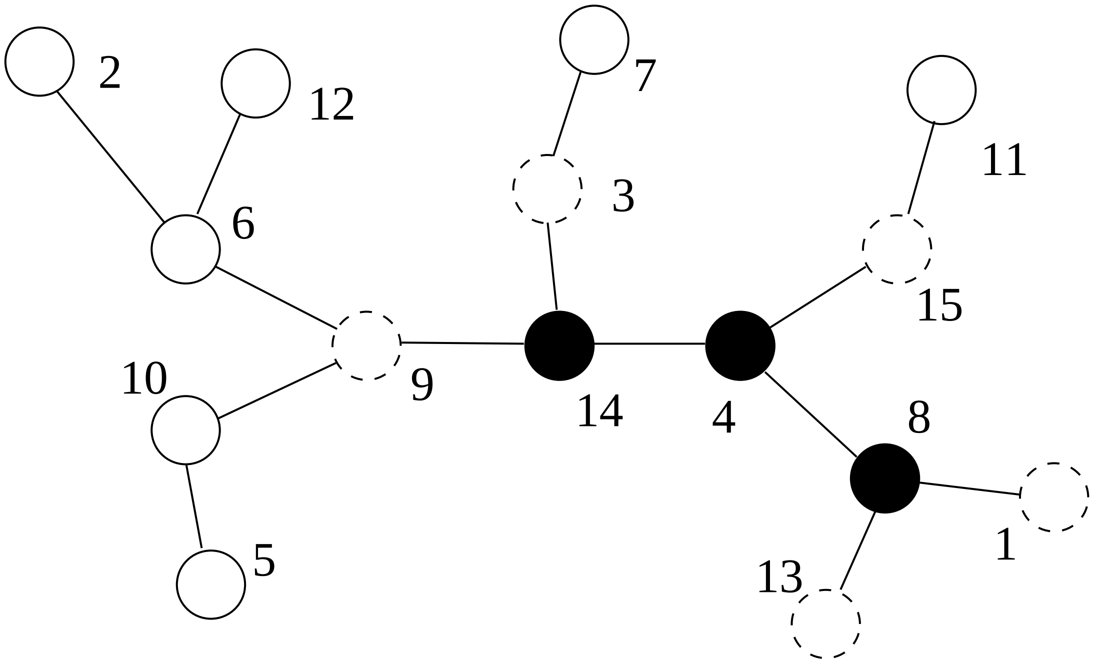
Insertion of this analytic solution into the first equation:

$$\boxed{A_1 = \Phi_1((A_1, \overline{\mathbf{a}}(x, u, A_1), x, u) = G(A_1, x, u)}$$

leads to **single equation**.

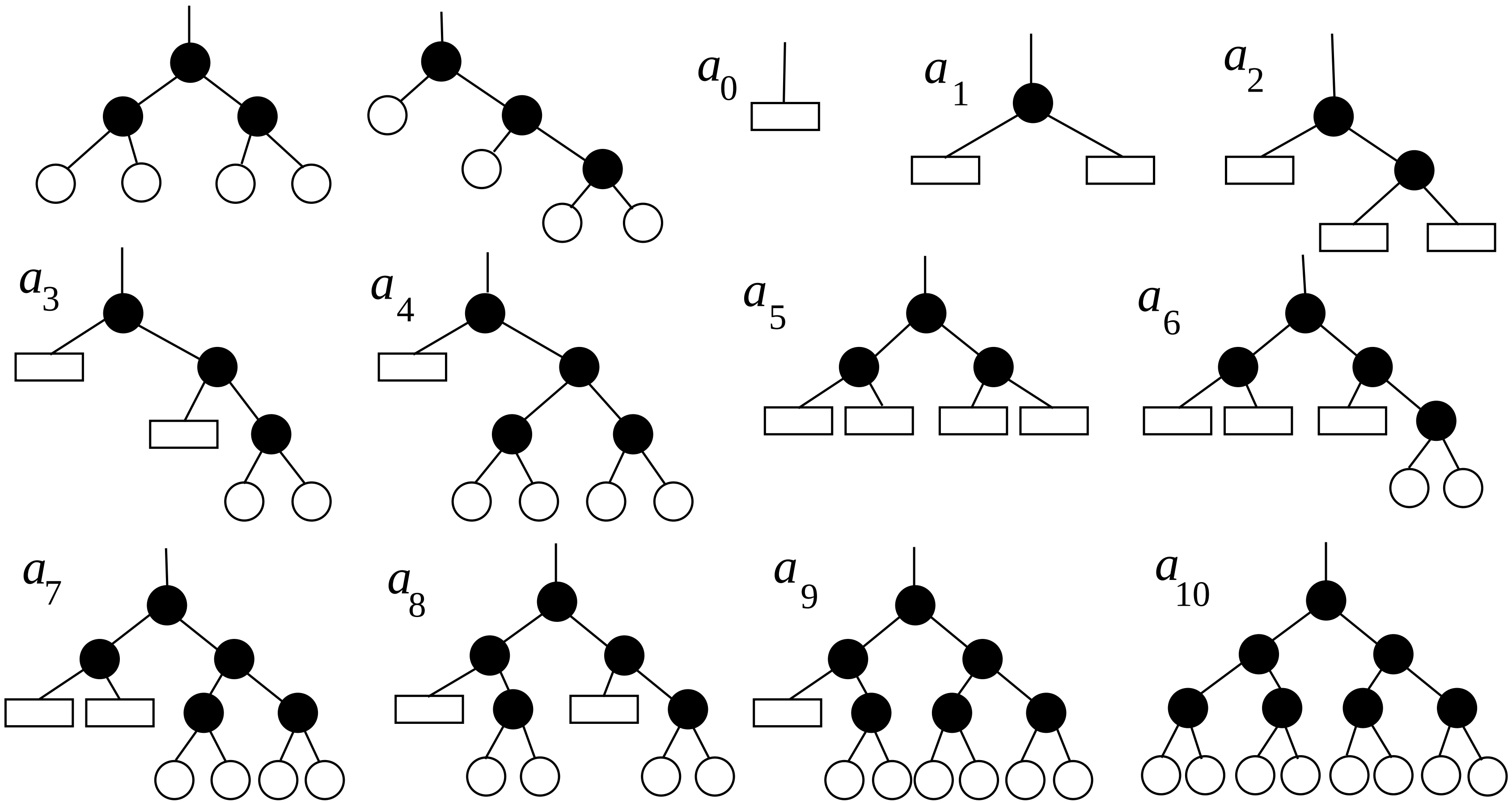
Combinatorics on Pattern in Trees

Occurence of a pattern $\mathcal{M}$  **in a labelled tree**



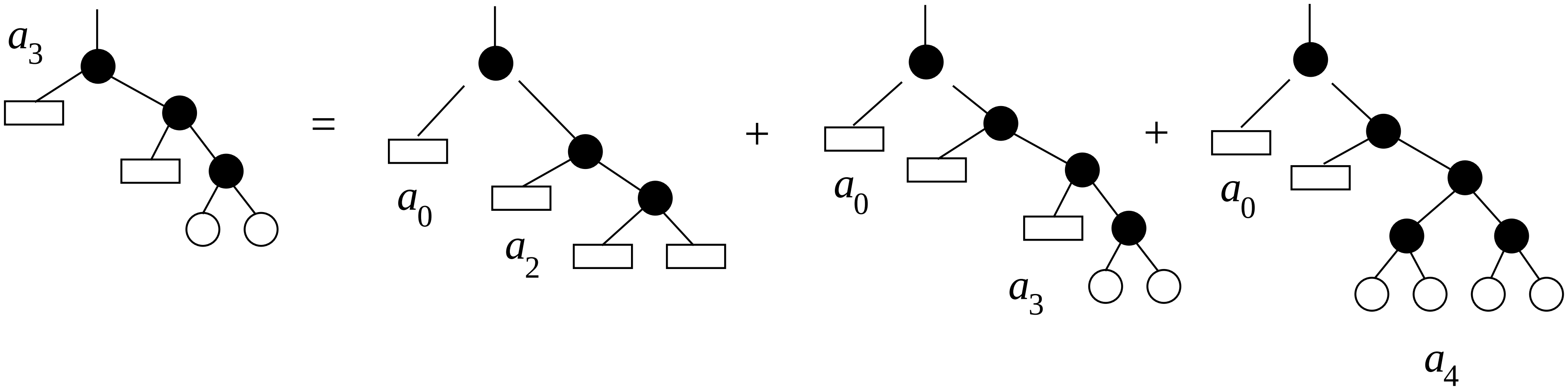
Combinatorics on Pattern in Trees

Partition of trees in classes ( ... out-degree different from 2)



Combinatorics on Pattern in Trees

Recurrences $A_3 = xA_0A_2 + xA_0A_3 + xA_0A_4$

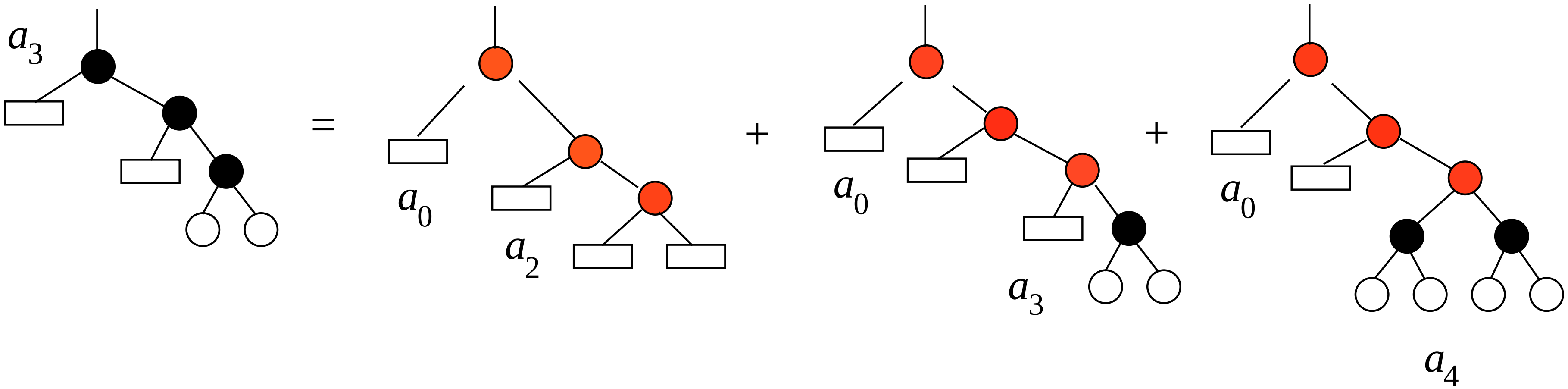


$$A_j(x) = \sum_{n,k} a_{j;n} \frac{x^n}{n!}$$

$a_{j;n}$ ... number of trees of size n in class j

Combinatorics on Pattern in Trees

Recurrences $A_3 = xuA_0A_2 + xuA_0A_3 + xuA_0A_4$



$$A_j(x,u) = \sum_{n,k} a_{j;n,m} \frac{x^n}{n!} u^m$$

$a_{j;n,m}$ ... number of trees of size n in class j with m occurences of $\mathcal{M}$

Combinatorics on Pattern in Trees

$$A_0 = A_0(x,u) = x + x\sum_{i=0}^{10} A_i + x\sum_{n=3}^{\infty}\frac{1}{n!}\left(\sum_{i=0}^{10} A_i\right)^n,$$

$$A_1 = A_1(x,u) = \frac{1}{2}xA_0^2,$$

$$A_2 = A_2(x,u) = xA_0A_1,$$

$$A_3 = A_3(x,u) = xA_0(A_2 + A_3 + A_4)u,$$

$$A_4 = A_4(x,u) = xA_0(A_5 + A_6 + A_7 + A_8 + A_9 + A_{10})u^2,$$

$$A_5 = A_5(x,u) = \frac{1}{2}xA_1^2u,$$

$$A_6 = A_6(x,u) = xA_1(A_2 + A_3 + A_4)u^2,$$

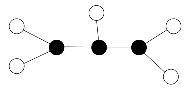
$$A_7 = A_7(x,u) = xA_1(A_5 + A_6 + A_7 + A_8 + A_9 + A_{10})u^3,$$

$$A_8 = A_8(x,u) = \frac{1}{2}x(A_2 + A_3 + A_4)^2u^3,$$

$$A_9 = A_9(x,u) = x(A_2 + A_3 + A_4)(A_5 + A_6 + A_7 + A_8 + A_9 + A_{10})u^4,$$

$$A_{10} = A_{10}(x,u) = \frac{1}{2}x(A_5 + A_6 + A_7 + A_8 + A_9 + A_{10})^2u^5.$$

Combinatorics on Pattern in Trees

Final Result for $\mathcal{M} =$ 

Central limit theorem with

$$\mu = \frac{5}{8e^3} = 0.0311169177\ldots$$

and

$$\sigma^2 = \frac{20e^3 + 72e^2 + 84e - 175}{32e^6} = 0.0764585401\ldots.$$

Perspectives

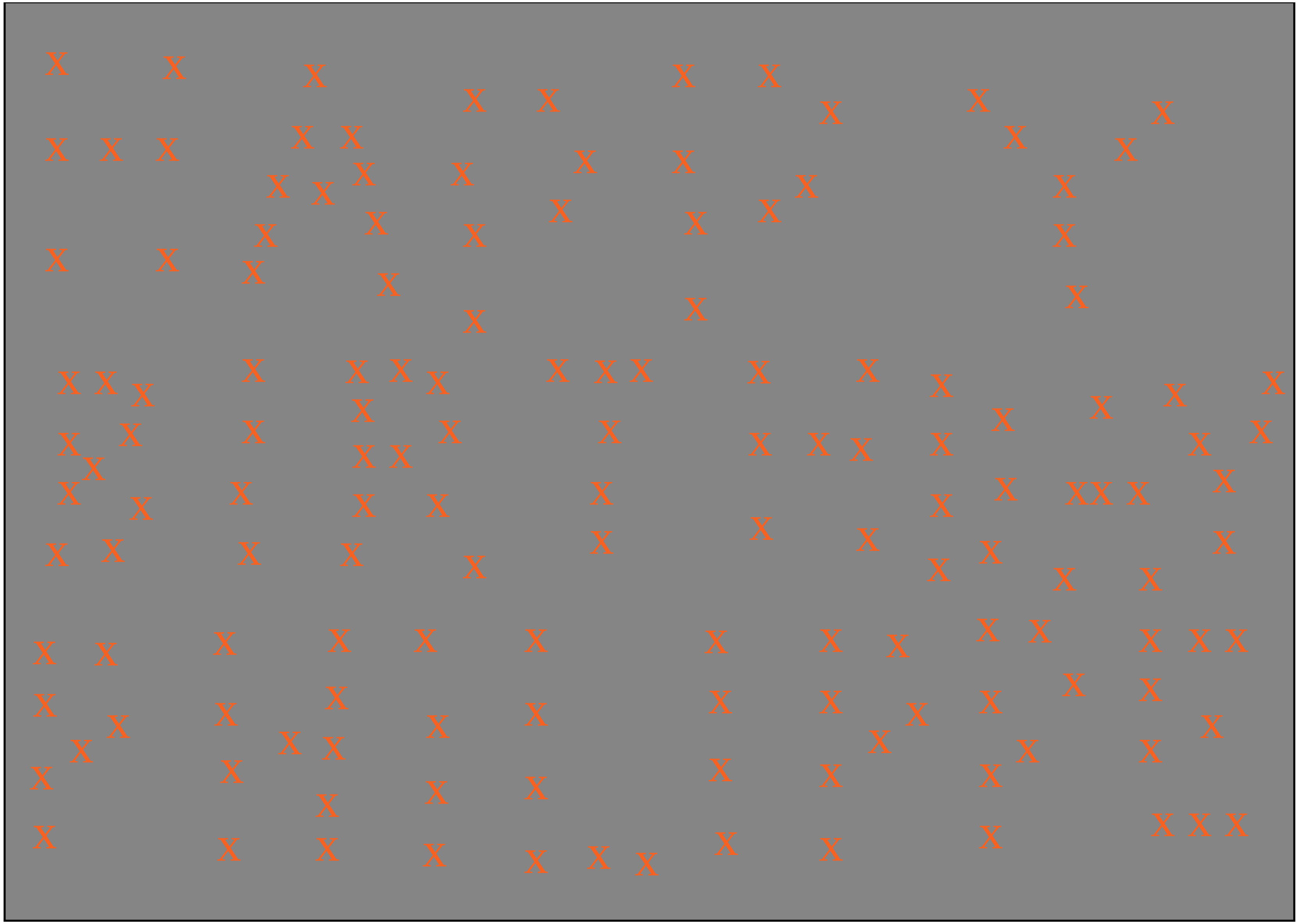
Further Applications

- Contextfree languages
- Planar graphs (with Giménez & Noy)
- Random walks on graphs (Woess)
- Random Boolean formulas (Woods, Chauvin & Flajolet & Gittenberger & Gardy)
- ...

Generalizations

- General dependency graph
- Infinite systems of equations
- ...

Patterns in Trees



Patterns in Trees