The distribution of the maximum protection number in random trees

Analytic and Probabilistic Combinatorics Workshop BIRS

Joint work with Clemens Heuberger and Stephan Wagner

17 November 2022

FWF
Der Wissenschaftsfonds.

UNIVERSITÄT KLAGENFURT

Modeling - Analysis - Optimization
DOC.FUNDS DOCTORAL SCHOOL

Bootstrapping and Double-Exponential Limit Laws (Prodinger & Wagner, 2015)

"It is a very typical situation that an extremal parameter in a combinatorial structure follows a discrete double-exponential distribution, and that fluctuations in the average occur."

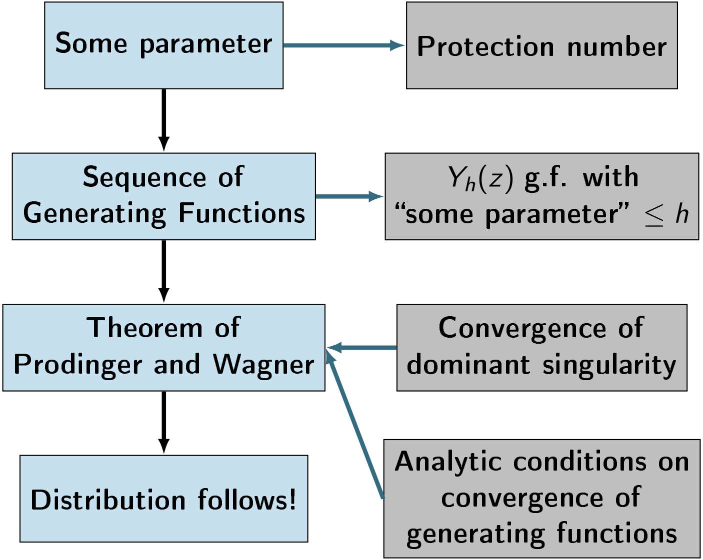
Some examples:

- Longest sequence of 1's in a 0-1 sequence.
- Longest sequence of $a \in \mathcal{A}$, $\mathcal{A}$ is an alphabet of size k.
- Longest horizontal segment in a Motzkin path.
- Maximum outdegree in planted plane trees.

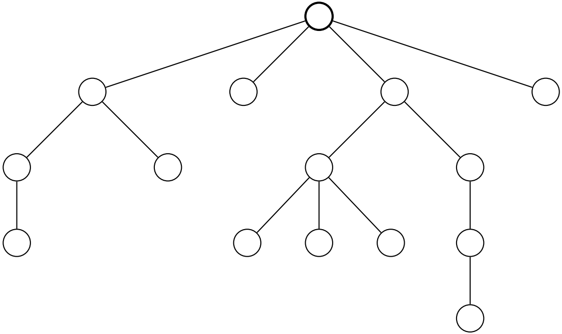
$$\mathbb{P}(X_n \leq h) = \exp(-An\rho^h)(1 + o(1)),$$

$$\mathbb{E}(X_n) = \log_b n + \log_b A + \frac{\gamma}{\log b} + \frac{1}{2} + \phi_b(\log_b An) + o(1).$$

Summary


Some parameter
Protection number
Sequence of Generating Functions
$Y_h(z)$ g.f. with "some parameter" $\leq h$
Theorem of Prodinger and Wagner
Convergence of dominant singularity
Analytic conditions on convergence of generating functions
Distribution follows!

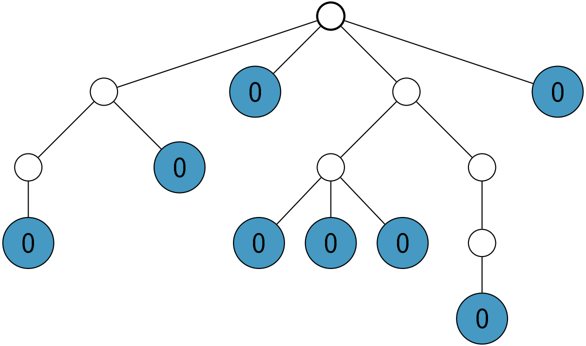
The protection number of a vertex



Definition

The protection number of a vertex v is the length of the shortest path from v to any leaf contained in the maximal subtree where v is a root.

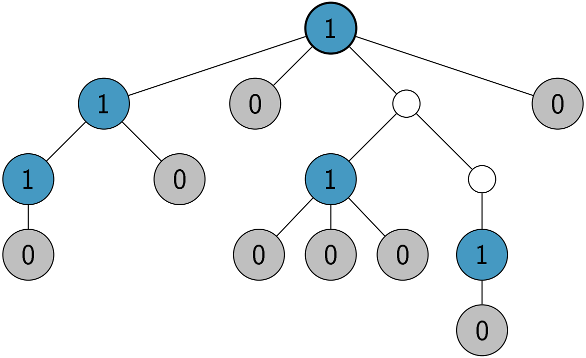
The protection number of a vertex



Definition

The protection number of a vertex v is the length of the shortest path from v to any leaf contained in the maximal subtree where v is a root.

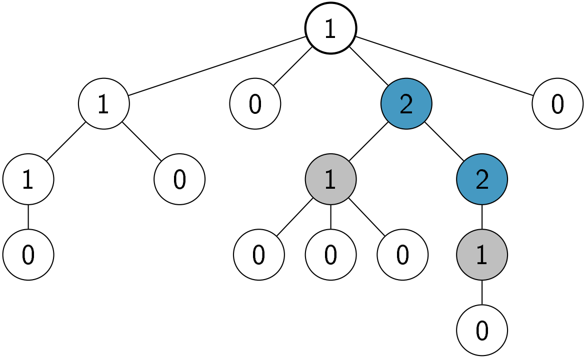
The protection number of a vertex



Definition

The protection number of a vertex v is the length of the shortest path from v to any leaf contained in the maximal subtree where v is a root.

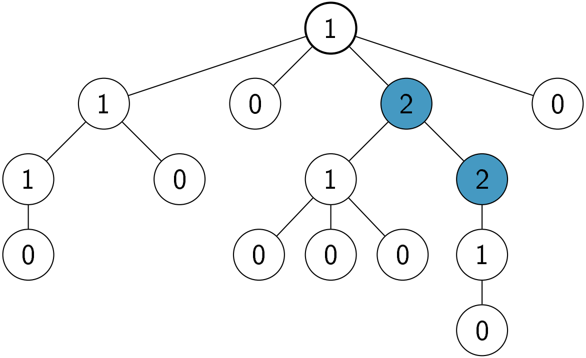
The protection number of a vertex



Definition

The protection number of a vertex v is the length of the shortest path from v to any leaf contained in the maximal subtree where v is a root.

The protection number of a vertex

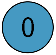


Definition

The protection number of a vertex v is the length of the shortest path from v to any leaf contained in the maximal subtree where v is a root.

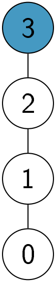
Maximum protection number: Some examples

- A maximum protection number of 0 means the tree is a single vertex.



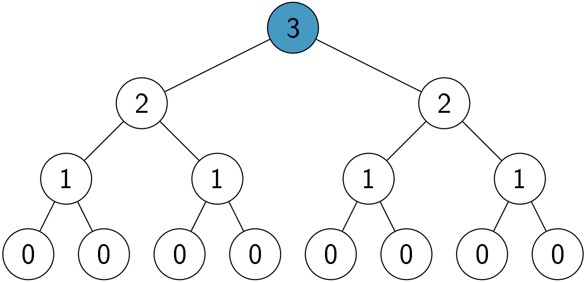
Maximum protection number: Some examples

- A maximum protection number of 0 means the tree is a single vertex.
- Paths (vertices are a leaf or have exactly one child) have a very high ratio of protection number to number of vertices.



Maximum protection number: Some examples

- A maximum protection number of 0 means the tree is a single vertex.
- Paths (vertices are a leaf or have exactly one child) have a very high ratio of protection number to number of vertices.
- Trees where vertices generally have more than one child have a low ratio of protection number to number of vertices.



Timeline of work on protection number of trees

- Number of vertices with protection number at least 2:
  - in ordered trees. *Cheon and Shapiro (2008).*
  - in k-ary trees, digital search trees, binary search trees, tries and suffix trees, random recursive trees.
    Devroye, Du, Gaither, Holmgren, Homma, Janson, Mahmoud, Mansour, Prodinger, Sellke, Ward (2010–2015).
- Number of vertices with protection number at least k, again in various types of trees.
  Bóna, Copenhaver, Devroye, Heuberger, Janson, Prodinger, Pittel (2014–2017).
- Protection number of the root. Plane trees, simply generated trees, Pólya trees.
  Gittenberger, Gołębiewski, Heuberger, Klimczak, Larcher, Prodinger, Sulkowska (2017–2021).

Simply generated trees

Definition

A simply generated tree has a generating function Y which satisfies the functional equation $Y(x) = x\Phi(Y(x))$ where Φ is a weight generating function $\Phi(x) = \sum_{n\geq 0} w_n x^n$, $w_n \geq 0$.

- Complete binary trees: $B(x) = x + xB(x)^2 = x(1+B(x)^2)$.
- Plane trees: $P(x) = x + xP(x) + xP(x)^2 + \cdots = x\frac{1}{1-P(x)}$.

Some things to note:

- $w_n \neq 0$ means the tree **can** have vertices with exactly n children.
- ρ is the (finite) radius of convergence or dominant singularity of $Y(x)$.
- $\tau = Y(\rho)$, so that $\Phi(\tau) = \tau\Phi'(\tau)$ and $\rho = \tau/\Phi(\tau) = 1/\Phi'(\tau)$.

Protection number of simply generated trees

Let $Y_{h,k}$ be the generating function for simply generated trees with:

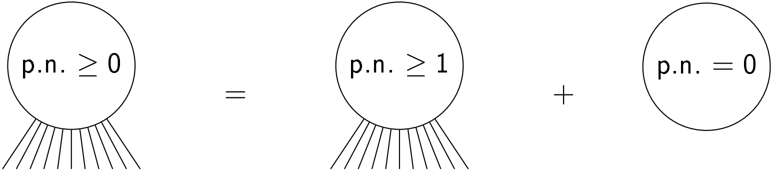
- the maximum protection number of any vertex is $\leq h$,
- the protection number of the root is $\geq k$.

Protection number of simply generated trees

Let $Y_{h,k}$ be the generating function for simply generated trees with:

- the maximum protection number of any vertex is $\leq h$,
- the protection number of the root is $\geq k$.

$$Y_{h,0}(x) = Y_{h,1}(x) + x,$$

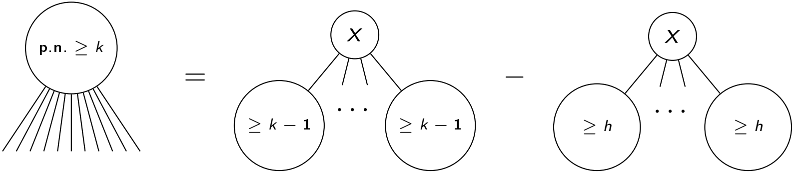


Protection number of simply generated trees

Let $Y_{h,k}$ be the generating function for simply generated trees with:

- the maximum protection number of any vertex is $\leq h$,
- the protection number of the root is $\geq k$.

$$
\begin{aligned}
Y_{h,0}(x) &= Y_{h,1}(x) + x, \\
Y_{h,k}(x) &= x\Phi(Y_{h,k-1}(x)) - x\Phi(Y_{h,h}(x)), \qquad 1 \leq k \leq h.
\end{aligned}
$$



Protection number of simply generated trees

The system of functional equations:

$$\begin{aligned} Y_{h,0}(x) &= Y_{h,1}(x) + x, \\ Y_{h,k}(x) &= x\Phi(Y_{h,k-1}(x)) - x\Phi(Y_{h,h}(x)), \qquad 1 \leq k \leq h. \end{aligned}$$

We set $x := \rho_h$ (common radius of convergence of system for fixed h) and $\eta_{h,k} := Y_{h,k}(\rho_h)$, so the system becomes

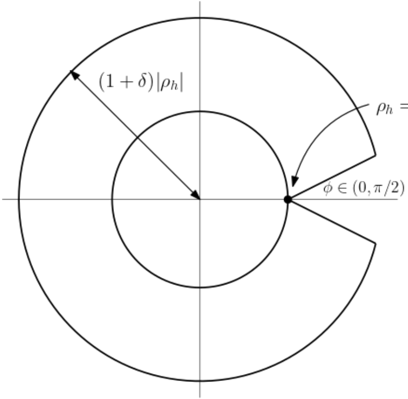
$$\begin{aligned} \eta_{h,0} &= \eta_{h,1} + \rho_h, \\ \eta_{h,k} &= \rho_h\Phi(\eta_{h,k-1}) - \rho_h\Phi(\eta_{h,h}), \qquad 1 \leq k \leq h. \end{aligned}$$

Determinant of Jacobian:

$$0 = \prod_{j=1}^{h} \left(\rho_h \Phi'(\eta_{h,j})\right) + \left(1 - \rho_h \Phi'(\eta_{h,0})\right)\left(1 + \sum_{k=2}^{h} \prod_{j=k}^{h} \left(\rho_h \Phi'(\eta_{h,j})\right)\right).$$

Theorem of Prodinger and Wagner

$$Y_h(z) = \sum_{n \geq 0} y_{h,n} z^n \to Y(z)$$



$\rho_h = \rho + c\zeta^h + o(\zeta^h), \zeta \in (0, 1)$

$Y_h(z) = A_h(z) + C_h(1 - z/\rho_h)^\alpha + o((1 - z/\rho_h)^\alpha)$
uniformly in h

$Y(z) = A(z) + C(1 - z/\rho)^\alpha + o((1 - z/\rho)^\alpha)$

For details: Helmut Prodinger and Stephan Wagner. Bootstrapping and double-exponential limit laws. DMTCS, 2015.

Goal: Apply the Theorem of Prodinger and Wagner

Problem 1

Show that the dominant singularity for $Y_{h,0}$ is $\rho_h \in \mathbb{R}$, where

$$\rho_h = \rho + c\zeta^h + o(\zeta^h)$$

as $h \to \infty$ for some constants $\rho > 0$, $c > 0$ and $0 < \zeta < 1$.

The system that we must use to obtain this result is the following:

$$\eta_{h,0} = \eta_{h,1} + \rho_h,$$

$$\eta_{h,k} = \rho_h \Phi(\eta_{h,k-1}) - \rho_h \Phi(\eta_{h,h})$$

$$0 = \prod_{j=1}^{h} \left(\rho_h \Phi'(\eta_{h,j})\right) + \left(1 - \rho_h \Phi'(\eta_{h,0})\right)\left(1 + \sum_{k=2}^{h} \prod_{j=k}^{h} \left(\rho_h \Phi'(\eta_{h,j})\right)\right).$$

Aim: $\rho_h = \rho + c\zeta^h + o(\zeta^h)$

Show:

1. $\rho_h \to \rho$.

Aim: $\rho_h = \rho + c\zeta^h + o(\zeta^h)$

Show:

1. $\rho_h \to \rho$.
2. $\eta_{h,k} \to \eta_k$.

Aim: $\rho_h = \rho + c\zeta^h + o(\zeta^h)$

Show:

1. $\rho_h \to \rho$.
2. $\eta_{h,k} \to \eta_k$.
3. $\prod_{j=1}^{h} \left(\rho_h \Phi'(\eta_{h,j})\right) = O((\rho\Phi'(0))^h)$ and

$$1 + \sum_{k=2}^{h} \prod_{j=k}^{h} \left(\rho_h \Phi'(\eta_{h,j})\right) \to \frac{1}{1 - \rho\Phi'(0)}.$$

Aim: $\rho_h = \rho + c\zeta^h + o(\zeta^h)$

Show:

1. $\rho_h \to \rho$.
2. $\eta_{h,k} \to \eta_k$.
3. $\prod_{j=1}^{h} \left(\rho_h \Phi'(\eta_{h,j})\right) = O((\rho\Phi'(0))^h)$ and
$$1 + \sum_{k=2}^{h} \prod_{j=k}^{h} \left(\rho_h \Phi'(\eta_{h,j})\right) \to \frac{1}{1 - \rho\Phi'(0)}.$$
4. $\eta_{h,0} = \tau + O({B_1}^h)$ and $\rho_h = \rho + O({B_1}^h)$.

Aim: $\rho_h = \rho + c\zeta^h + o(\zeta^h)$

Show:

1. $\rho_h \to \rho$.
2. $\eta_{h,k} \to \eta_k$.
3. $\prod_{j=1}^{h} \left(\rho_h \Phi'(\eta_{h,j})\right) = O((\rho\Phi'(0))^h)$ and
$$1 + \sum_{k=2}^{h} \prod_{j=k}^{h} \left(\rho_h \Phi'(\eta_{h,j})\right) \to \frac{1}{1-\rho\Phi'(0)}.$$
4. $\eta_{h,0} = \tau + O({B_1}^h)$ and $\rho_h = \rho + O({B_1}^h)$.
5. $\prod_{j=1}^{h} \left(\rho_h \Phi'(\eta_{h,j})\right) = (\rho\Phi'(0))^h \lambda_2 (1 + O({B_2}^h))$ and
$$1 + \sum_{k=2}^{h} \prod_{j=k}^{h} \left(\rho_h \Phi'(\eta_{h,j})\right) = \frac{1}{1-\rho\Phi'(0)(1+O({B_3}^h))}.$$

Asymptotic behaviour of the singularity

Lemma (Heuberger, SJS, Wagner, 2022+)

As $h \to \infty$, we have that

$$\rho_h = \rho + \frac{1}{\Phi(\tau)}(\rho\Phi'(0))^{h+1}\lambda_1(1-\rho\Phi'(0)) + o((\rho\Phi'(0))^h),$$

where

$$\lambda_1 = \eta_0 \prod_{i\geq 1} \frac{\eta_i}{\rho\Phi'(0)\eta_{i-1}}.$$

Asymptotic behaviour of the singularity

Lemma (Heuberger, SJS, Wagner, 2022+)

As $h \to \infty$, we have that

$$\rho_h = \rho + \frac{1}{\Phi(\tau)} (\rho\Phi'(0))^{h+1} \lambda_1 (1 - \rho\Phi'(0)) + o((\rho\Phi'(0))^h),$$

where

$$\lambda_1 = \eta_0 \prod_{i \geq 1} \frac{\eta_i}{\rho\Phi'(0)\eta_{i-1}}.$$

With some additional analysis. . .

Result

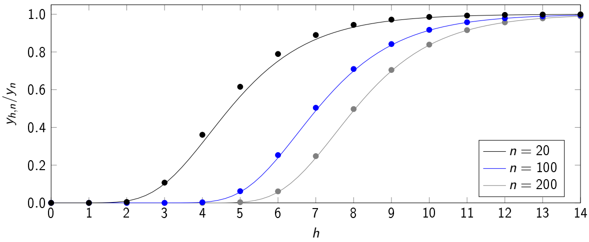
Theorem (Heuberger, SJS, Wagner, 2022+)

The probability that a random tree of size n has maximum protection number $\leq h$ is

$$\frac{y_{h,n}}{y_n} = exp\Big(-\frac{1}{\tau}\Phi'(0)\lambda_1(1-\rho\Phi'(0))n(\rho\Phi'(0))^h\Big)(1+o(1))$$

as $n \to \infty$ and $h = \log_{(\rho\Phi'(0))^{-1}} n + O(1)$.

Binary trees, $\Phi(x) = (1+x)^2$: Actual data (marks) plotted with the distribution (line)



Result

Theorem (Heuberger, SJS, Wagner, 2022+)

The expected value of the maximum protection number in a random tree of size n is

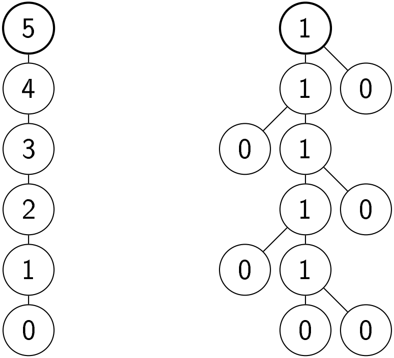
$$\log_b(n) + \log_b\Big(\frac{\lambda_2\Phi'(0)}{\Phi(\tau)}(1-\rho\Phi'(0))\Big) + \frac{\gamma}{\log(b)} + \frac{1}{2} + \psi_b\Big(\log_b\Big(\frac{\lambda_2\Phi'(0)}{\Phi(\tau)}(1-\rho\Phi'(0))n\Big)\Big) + o(1),$$

where γ denotes the Euler-Mascheroni constant, $\lambda_2 = \eta_0\prod_{i\geq 1}\frac{\eta_i}{\rho\Phi'(0)\eta_{i-1}}$, and ψ_b is the 1-periodic function that is defined by the Fourier series

$$\psi_b(x) = -\frac{1}{\log(b)}\sum_{k\neq 0}\Gamma\Big(-\frac{2k\pi i}{\log(b)}\Big)e^{2k\pi i x}.$$

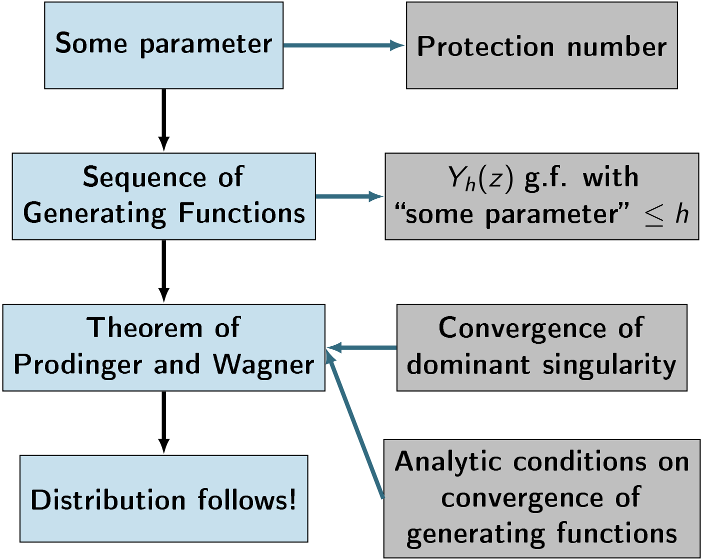
There's more!

Proofs and results depend on $\Phi'(0) \neq 0$. So we must consider the case where $\Phi'(0) = w_1 = 0$ separately.



- Set $r = \min\{s \in \mathbb{N} : \Phi^{(s)}(0) \neq 0\}$, $r \geq 2$.
- $\rho_h = \rho + c\zeta^{r^h} + o(\zeta^{r^h})$.

Thank you!


Some parameter
Protection number
Sequence of Generating Functions
$Y_h(z)$ g.f. with "some parameter" $\leq h$
Theorem of Prodinger and Wagner
Convergence of dominant singularity
Distribution follows!
Analytic conditions on convergence of generating functions