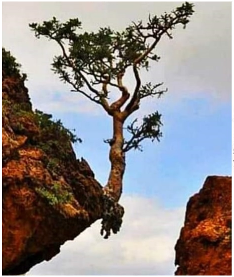
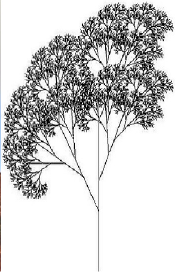
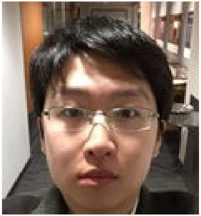
Exponential recursive trees: Profile study

Rafik Aguech
Faculté des Sciences de Monastir (Tunisia) & King Saud Univ. (Saudi Arabia)
Seminar @ Univ. Sorbonne Paris Nord, 6 Dec. 2022

Based on joint works with





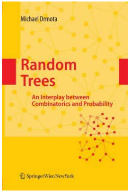

Sudip Bose Hosam Mahmoud Yi Zhang Mehri Javanian

George Washington Univ. (USA) Zanjan Univ. (Iran)

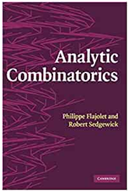
Aguech, Bose, Mahmoud, Zhang: *Some properties of exponential trees*, International Journal of Computer Mathematics: Computer Systems Theory, 2021.

Aguech, Javanian: *Protected nodes in exponential recursive trees*, submitted, 2022.

Motivations



Classic litterature on growing trees and graphs deals mostly with objects growing "slowly" by adding a small number of nodes and edges at each step.



Modern (social/professional/political) networks, such as Facebook, Instagram, LinkedIn and Twitter, grow very quickly and exhibit degree distributions not compatible with models like Erdős–Renyi random graphs, Galton–Watson random trees. . .

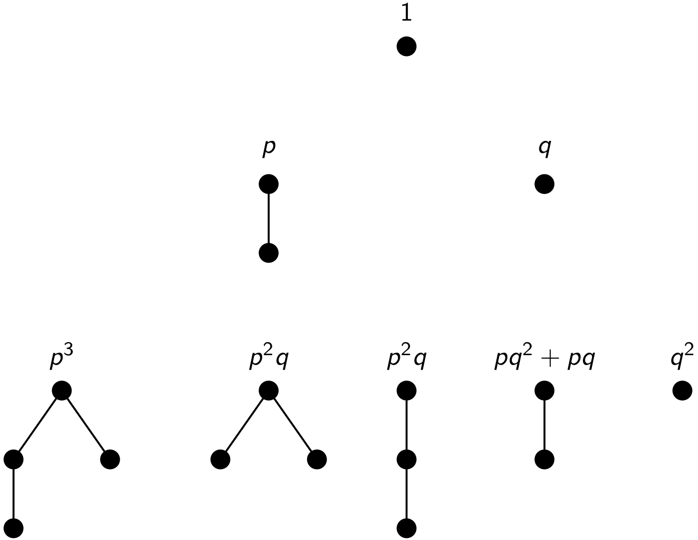
To cope with the need to model certain aspects of fast growing stuctures, some interesting *tree models* were introduced:

[Feng, Mahmoud 2018]: *Profile of random exponential binary trees*
[Mahmoud 2021]: *Profile of random exponential recursive trees*

Table of contents

Exponential recursive trees

To each node is attached a new child with probability p.



$$S_n = S_{n-1} + \mathrm{Bin}\Big(S_{n-1},\, p\Big).$$

Let $\mathcal{F}_n$ be the sigma field generated by the first n steps of evolution. We have

$$\mathbb{E}[S_n \mid \mathcal{F}_{n-1}] = S_{n-1} + pS_{n-1} = (1+p)\, S_{n-1}.$$

So $S_n/\,(1+p)^n$ is an L^1 bounded martingale,
therefore it converges almost surely to some random variable S_*.

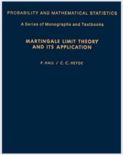
What is the distribution of S_*?

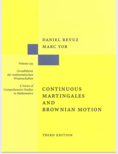
Martingales

Definition

A discrete time process $(X_n)_n$ is called a martingale relative to the sigma field (aka, filtration) $(\mathcal{F}_n)_n$ if for all $n \geq 0$:

- X_n is $\mathcal{F}_n$ mesurable;
- $\mathbb{E}[|X_n|] < +\infty$;
- $\mathbb{E}[X_n \,|\, \mathcal{F}_{n-1}] = X_{n-1}$ almost surely.





Martingale convergence theorem

If $(X_n, F_n)_n$ is an L^1 bounded martingale
(i.e. there is c such that for all n, $\mathbb{E}[|X_n|] \leq c$),
then there exists a real valued random variable X defined on the same probability space as $(X_n)_n$ such that $\lim_n X_n = X$ almost surely.

Remark: X is also in L^1.

Size of exponential recursive trees

Let $\mathbb{I}$ be the indicator of success in recruiting at the root in the first step.

$$S_n \stackrel{D}{=} S_{n-1}\mathbb{I} + \widetilde{S}_{n-1}$$

Size of exponential recursive trees

Proposition [Mahmoud 2020]

Let S_n be the size (number of nodes) of an exponential recursive tree after n steps. Then

$$\frac{S_n}{(p+1)^n} \xrightarrow{a.s.} S_*,$$

where the limiting random variable S_* has moments $a_m := \mathbb{E}[S_*^m]$ defined inductively by

$$a_m = \frac{p}{(p+1)^m - (p+1)} \sum_{i=1}^{m-1} \binom{m}{i} a_i a_{m-i}, \qquad \text{for } m \geq 2,$$

with $a_1 = \mathbb{E}[S_*] = 1$.

Leaves in exponential recursive trees

For an exponential recursive tree $\mathcal{T}_n$ of age n, let L_n be its number of leaves.

- If $\mathbb{I} = 0$ the tree $\mathcal{T}'_{n-1}$ in the following $n-1$ steps behaves as a tree of size $n-1$, with L'_{n-1} leaves, and $L'_{n-1} \overset{D}{=} L_{n-1}$.

Leaves in exponential recursive trees

For an exponential recursive tree $\mathcal{T}_n$ of age n, let L_n be its number of leaves.

- If $\mathbb{I} = 0$ the tree $\mathcal{T}'_{n-1}$ in the following $n-1$ steps behaves as a tree of size $n-1$, with L'_{n-1} leaves, and $L'_{n-1} \overset{D}{=} L_{n-1}$.
- If $\mathbb{I} = 1$ at step 1 we obtain a root and a child, each acting independently to construct its own exponential tree in $n-1$ steps. The child will develop a tree $\mathcal{T}''_{n-1}$ with $L''_{n-1} \overset{D}{=} L_{n-1}$ leaves. The root continues to recruit and will father a tree $\mathcal{T}'''_{n-1}$ with $L'''_{n-1} \overset{D}{=} L_{n-1}$ leaves.

Leaves in exponential recursive trees

For an exponential recursive tree $\mathcal{T}_n$ of age n, let L_n be its number of leaves.

- If $\mathbb{I} = 0$ the tree $\mathcal{T}'_{n-1}$ in the following $n-1$ steps behaves as a tree of size $n-1$, with L'_{n-1} leaves, and $L'_{n-1} \overset{D}{=} L_{n-1}$.
- If $\mathbb{I} = 1$ at step 1 we obtain a root and a child, each acting independently to construct its own exponential tree in $n-1$ steps. The child will develop a tree $\mathcal{T}''_{n-1}$ with $L''_{n-1} \overset{D}{=} L_{n-1}$ leaves. The root continues to recruit and will father a tree $\mathcal{T}'''_{n-1}$ with $L'''_{n-1} \overset{D}{=} L_{n-1}$ leaves.
- When the tree $\mathcal{T}'''_{n-1}$ is hooked to $\mathcal{T}''_{n-1}$ to construct $\widetilde{\mathcal{T}}_n$ (a tree distributed like $\mathcal{T}_n$), its contribution to the total number of leaves in $\widetilde{\mathcal{T}}_n$ is potentially reduced by 1, if $\mathcal{T}'''_{n-1}$ is a single node (also a leaf), as the hooking changes the outdegree of the root to 1. Let us indicate the event that $\mathcal{T}'''_{n-1}$ is a leaf by the indicator variable $\mathbb{J}_{n-1}$.

Leaves in exponential recursive trees

$\mathbb{J}_{n-1}$ = the indicator of the event that $\mathcal{T}'''_{n-1}$ is a leaf.

Remark

We have

$$\mathbb{P}(\mathbb{J}_{n-1} = 1) = q^{n-1}.$$

Leaves in exponential recursive trees

Proposition

$$L_n \overset{D}{=} L'_{n-1}(1-\mathbb{I}) + (L''_{n-1} + L'''_{n-1} - \mathbb{J}_{n-1})\,\mathbb{I},$$

where L'_k, L''_k and L'''_k are independent copies of L_k.

Remark

$(\mathbb{I}, L'_{n-1}, L''_{n-1})$ is a block of independent random variables. This block is independent of of the block $(\mathbb{J}_{n-1}, L'''_{n-1})$.

Leaves in exponential recursive trees

Proposition

Let L_n be the number of leaves in an exponential recursive tree at age n. The mean and variance of L_n are

$$\mathbb{E}[L_n] = \frac{1}{2}(p+1)^n + \frac{1}{2}(1-p)^n \sim \frac{1}{2}(p+1)^n,$$

$$\begin{aligned}\mathbb{V}\mathrm{ar}[L_n] &= \frac{1}{4}(1-p)(p+1)^{2n-1} - \frac{1}{6-2p}(1-p)^2(p+1)^{n-1} \\ &\quad + \frac{1}{2}(1-p)^n - \frac{1}{4(3-p)}(1-p)^{2n+1} - \frac{1}{2}(1-p^2)^{2n} \\ &\sim \frac{1}{4}(1-p)(p+1)^{2n-1}.\end{aligned}$$

Leaves in exponential recursive trees

Theorem

Let L_n be the number of leaves in an exponential recursive tree at age n. We have the convergence

$$\frac{L_n}{(p+1)^n} \xrightarrow{D} L_*,$$

where the limiting random variable L_* has moments $b_m := \mathbb{E}[L_*^m]$ defined inductively by

$$b_m = \frac{p}{(p+1)^m - (p+1)} \sum_{i=1}^{m-1} \binom{m}{i} b_i b_{m-i}, \qquad \text{for } m \geq 2,$$

with $b_1 = \mathbb{E}[L_*] = \frac{1}{2}$.

Leaf profile in exponential recursive trees

Let $X_{n,k}$ the number of nodes at level k in $\mathcal{T}_n$. [Mahmoud 2020] finds

$$\mathbb{E}[X_{n,k}] = p^k \binom{n}{k}.$$

How many nodes among these are leaves?
Let $L_{n,\,k}$ be this number of leaves.

Lemma

In distribution , we have, for $1 \le k \le n$

$$L_{n,k} = \text{Bin}(X_{n-1,k-1},\, p) + \text{Bin}(L_{n-1,\,k},\, q).$$

Leaf profile in exponential recursive trees

Theorem

Let $L_{n,k}$ be the number of leaves at level k in an exponential recursive tree at age n.
We then have, for $1 \leq k \leq n$

$$\mathbb{E}[L_{n,k}] = p^k \sum_{\ell=0}^{n-k} q^\ell \binom{n-1-\ell}{k-1}.$$

Internal path length in exponential recursive trees

Definition

The distance of a node to the root of a tree (measured in the number of edges on the joining path) is called the *depth* of the node in the tree.

After n steps of evolution, the tree has S_n nodes at various depths.
Label the nodes of the tree $\mathcal{T}_n$ arbitrarily with labels $1, 2, \ldots, S_n$.
We note D_i the depth of node i.

Definition

The *internal path length* I_n is the sum of the depths of all the nodes:

$$I_n = \sum_{i=1}^{S_n} D_i.$$

One has the stochastic equality:

$$I_n = \sum_{i=1}^{S_{n-1}} D_i + (D_i + 1)[\![\text{proba that one adds a child to node } i]\!].$$

Taking an expectation conditioned of $\mathcal{F}_{n-1}$, we get

$$\mathbb{E}[I_n \,|\, \mathcal{F}_{n-1}] = \sum_{i=1}^{S_{n-1}} D_i + p \sum_{i=1}^{S_{n-1}} (D_i + 1) = (p+1) I_{n-1} + p S_{n-1}.$$

Lemma

With respect to the filtration $(\mathcal{F}_n)_n$, the process

$$M_n = \frac{1}{(p+1)^n}\, I_n - \frac{pn}{(p+1)^{n+1}}\, S_n$$

is a martingale.

With $\alpha = p/(1+p)$, we have

$$\mathbb{E}[I_n] = np(p+1)^{n-1} = \alpha n \mathbb{E}[S_n],$$

which suggests that $I_n/(n(p+1)^n)$ converges in L^1 to αS^*. Toward this end, we use [Mahmoud 2020], which asserts that a large number of nodes falls at depths around αn, where the tree is widest.
We evaluate the difference

$$M_n = \frac{I_n}{(p+1)^n} - \frac{\alpha n S_n}{(p+1)^n} = \frac{1}{(p+1)^n} \sum_{k=1}^{S_n} (D_i - \alpha n).$$

Recall that $X_{n,k}$ is the number of nodes at level k, we have

$$M_n = (p+1)^{-n} \sum_{k=1}^{n} (k - \alpha n) X_{n,k}.$$

Splitting this sum according to whether $|k - \alpha n| \leq n^{3/4}$ or not, we get

$$\begin{aligned} M_n &= \frac{1}{(p+1)^n} \sum_{\substack{k=1 \\ |k-\alpha n| \leq n^{3/4}}}^{n} (k - \alpha n) X_{n,k} \\ &+ \frac{1}{(p+1)^n} \sum_{\substack{k=1 \\ |k-\alpha n| > n^{3/4}}}^{n} (k - \alpha n) X_{n,k} \\ &=: M_n^{(1)} + M_n^{(2)}. \end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\big[|M_n^{(1)}|\big] &= \frac{1}{(p+1)^n} \sum_{\substack{k=1 \\ |k-\alpha n| \le n^{3/4}}}^{n} |k - \alpha n| \, p^k \binom{n}{k} \\
&= \frac{n^{3/4}}{(1-\alpha)^n (p+1)^n} \sum_{\substack{k=1 \\ |k-\alpha n| \le n^{3/4}}}^{n} \alpha^k (1-\alpha)^{n-k} \binom{n}{k} \\
&\le n^{3/4},
\end{aligned}
$$

providing the limit $\lim_{n\to\infty} \mathbb{E}\big[\frac{M_n^{(1)}}{n}\big] = 0$.

$$
\begin{aligned}
\mathbb{E}\left[|M_n^{(2)}|\right] &= \frac{1}{(p+1)^n} \sum_{\substack{k=1 \\ |k-\alpha n|>n^{3/4}}}^{n} |k-\alpha n|\, p^k \binom{n}{k} \\
&\leq \frac{1}{(1-\alpha)^n (p+1)^n} \\
&\times \sum_{\substack{k=1 \\ |k-\alpha n|>n^{3/4}}}^{n} (n-\alpha n)\left((1-\alpha)p\right)^k (1-\alpha)^{n-k} \binom{n}{k} \\
&= (1-\alpha)n \sum_{\substack{k=1 \\ |k-\alpha n|>n^{3/4}}}^{n} \alpha^k (1-\alpha)^{n-k} \binom{n}{k}.
\end{aligned}
$$

Internal path length in exponential recursive trees

Convergence of the martingale M_n/n: Second step

Introduce the random variables $G_n := \text{Bin}(n, \alpha)$ and $Z := \mathcal{N}(0,1)$. We have

$$\begin{aligned}
\mathbb{E}\big[|M_n^{(2)}|\big] &\le (1-\alpha)n\,\mathbb{P}\big(|G_n - \alpha n| > n^{3/4}\big) \\
&= (1-\alpha)n\,\mathbb{P}\Big(\Big|\frac{G_n - \alpha n}{\sqrt{\alpha(1-\alpha)n}}\Big| > \frac{n^{3/4}}{\sqrt{\alpha(1-\alpha)n}}\Big) \\
&= (1-\alpha)n\,\mathbb{P}\Big(|Z| > \frac{n^{1/4}}{\sqrt{\alpha(1-\alpha)}}\Big)\big(1+o(1)\big).
\end{aligned}$$

In the last line we applied the central limit theorem approximation. By Markov's inequality $\mathbb{P}(X \ge a) \le \frac{\mathbb{E}[X]}{a}$, we get, as $n \to \infty$

$$\mathbb{E}\Big[\Big|\frac{M_n^{(2)}}{n}\Big|\Big] \le \frac{\mathbb{E}\big[|Z|\big](1-\alpha)\sqrt{\alpha(1-\alpha)}}{n^{1/4}}\big(1+o(1)\big) \to 0.$$

Theorem

Let I_n be the internal path length of an exponential recursive tree at age n. As $n \to \infty$, we have

$$I_n \sim np(1+p)^{n-1} S_*,$$

that is, more rigorously,

$$\frac{I_n}{n\,(1+p)^n} \xrightarrow{L^1} \frac{p}{p+1}\, S_*.$$

Protected nodes

A node is called *protected* if it is at distance ≥ 2 of any leaf.

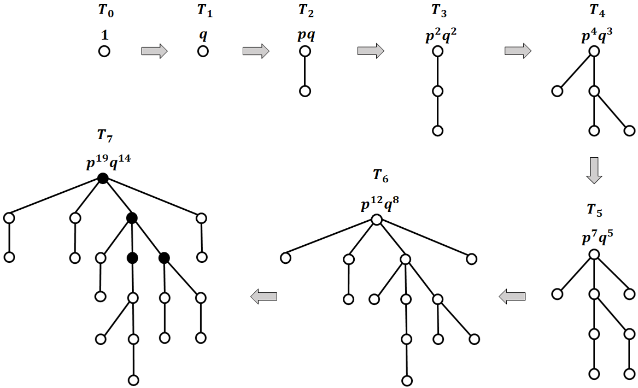


Figure: **(a)** An evolution (first 7 seps) of an exponential recursive tree, with the corresponding probabilities. **(b)** In $\mathcal{T}_7$, protected nodes are shown in black.

Protected nodes

Theorem

Let X_n be the number of protected nodes in an exponential recursive tree of age n. We have the convergence in distribution

$$\frac{X_n}{(p+1)^n} \xrightarrow{D} X_*,$$

where the limiting random variable X_* has moments $b_m := \mathbb{E}[X_*^m]$ defined inductively by

$$b_m = \frac{p}{(p+1)^m - (p+1)} \sum_{i=1}^{m-1} \binom{m}{i} b_i b_{m-i}, \qquad m \geq 2,$$

with $b_1 = \mathbb{E}[X_*] = \mu_p = \lim_{n\to\infty} \mu_{n,p}$.

Conclusion, open problems

Natural model: exponential recursive trees
$=$ each (internal or external) node gets a new child with probability p.

We got the asymptotic behaviour for

- ✓ size after n iterations
- ✓ # leaves (and number $L_{n,k}$ of leaves at depth k)
- ✓ internal path length
- ✓ # protected nodes

Open problems:

- law of the (maximal) height H_n?
- *law* of the profile (not just the mean): limit of the joint law $(L_{n,1}, \ldots, L_{n,H_n})$?
- law of the location of the largest width?

Conclusion, open problems

Natural model: exponential recursive trees
$=$ each (internal or external) node gets a new child with probability p.

We got the asymptotic behaviour for

- ✓ size after n iterations
- ✓ # leaves (and number $L_{n,k}$ of leaves at depth k)
- ✓ internal path length
- ✓ # protected nodes

Open problems:

- law of the (maximal) height H_n?
- *law* of the profile (not just the mean): limit of the joint law $(L_{n,1}, \ldots, L_{n,H_n})$?
- law of the location of the largest width?

exp(exp(exp(exp(thanks))))!