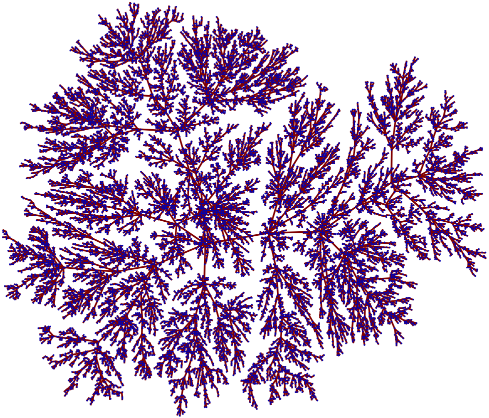
Asymptotic Regimes in the Destruction of Large Random Recursive Trees

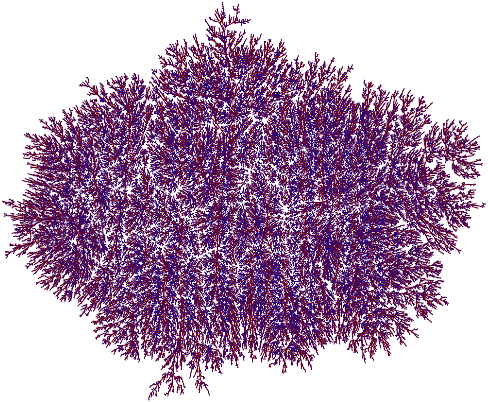
Jean Bertoin

based on a series of joint works with
Erich BAUR (ENS Lyon)

Kraków, AofA'16



RRT $n = 15000$ (Courtesy of Igor Kortchemski)



RRT $n = 50000$ (Courtesy of Igor Kortchemski)

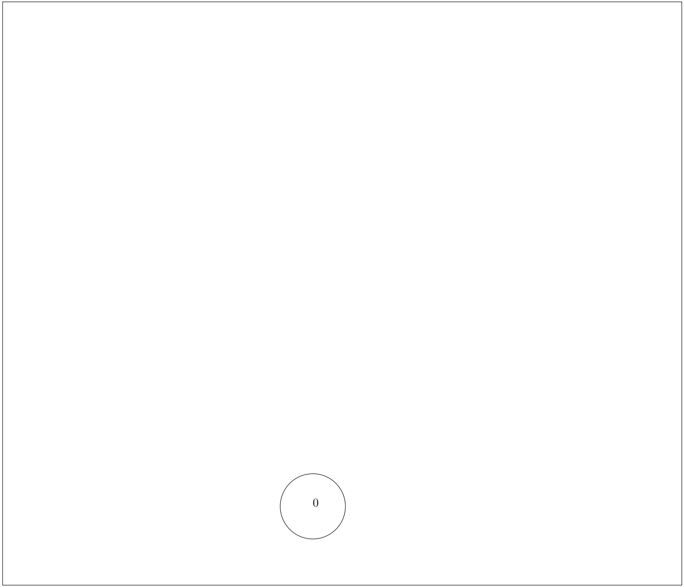
Recursive construction

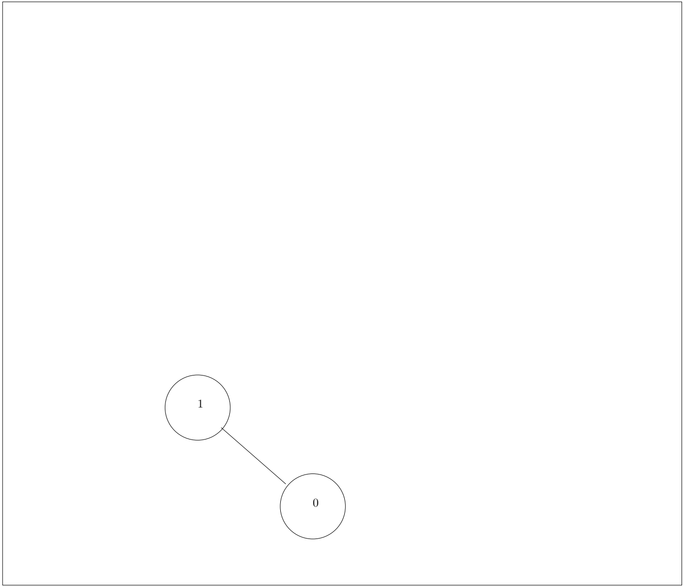
Random Recursive Trees can be viewed as the genealogical tree induced by a Yule process[1].

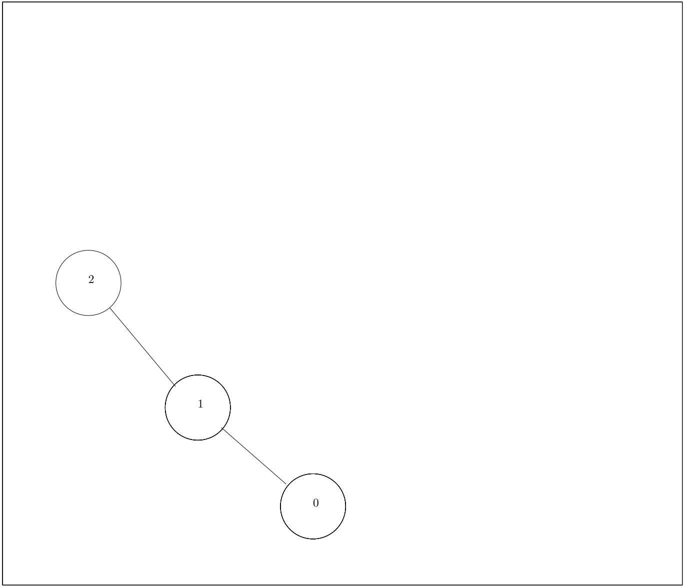
They can be constructed by a simple algorithm :

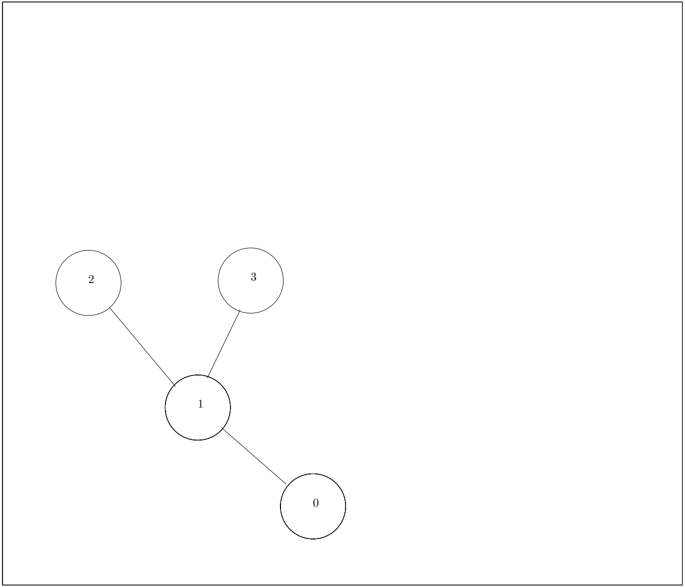
For each vertex $i \geq 1$, the parent of i has the uniform distribution on $\{0, \ldots, i-1\}$, independently of the other vertices.

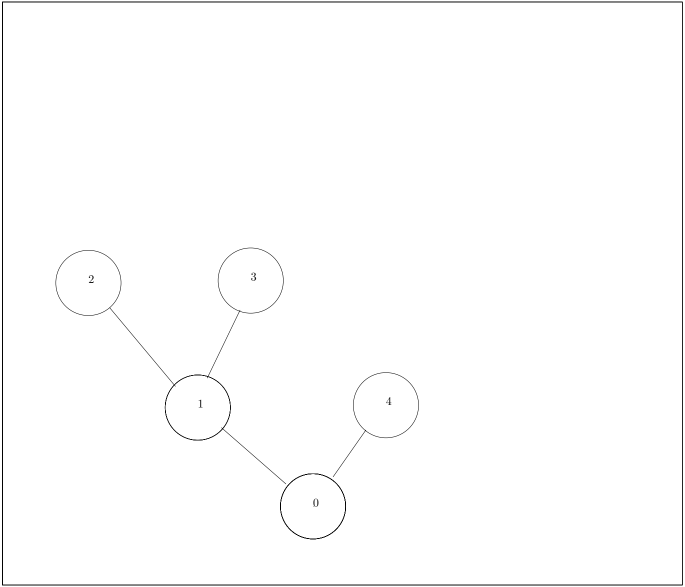
[1] population model in which each individual gives birth to a child at a constant (unit) rate.

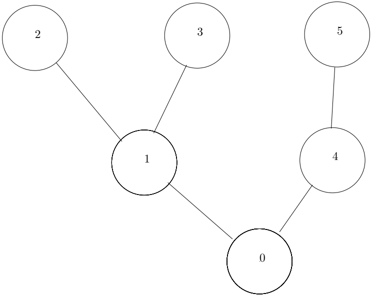

0

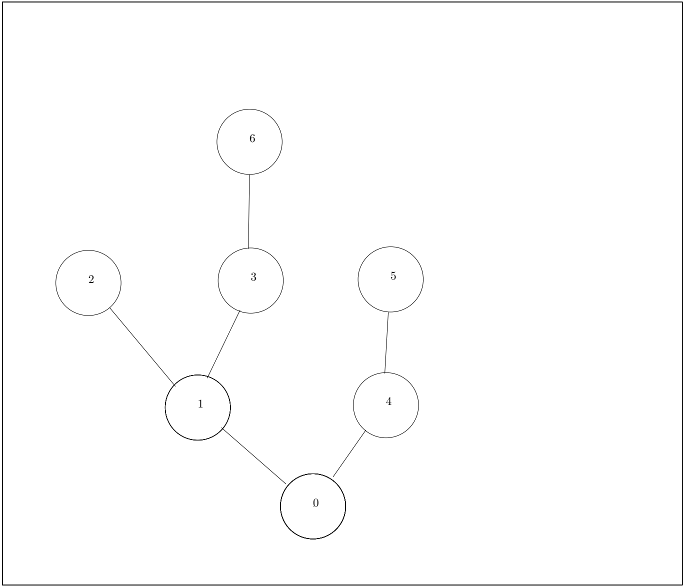

1
0


2
1
0


2
3
1
0


2
3
1
4
0


2
3
5
1
4
0

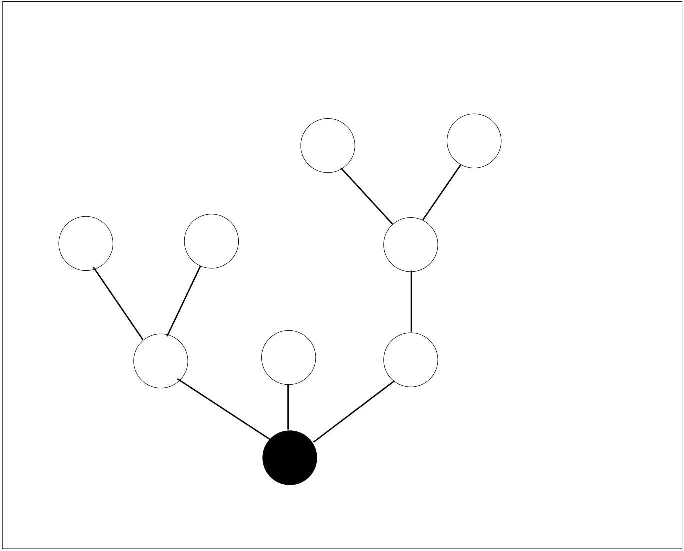

6
2
3
5
1
4
0

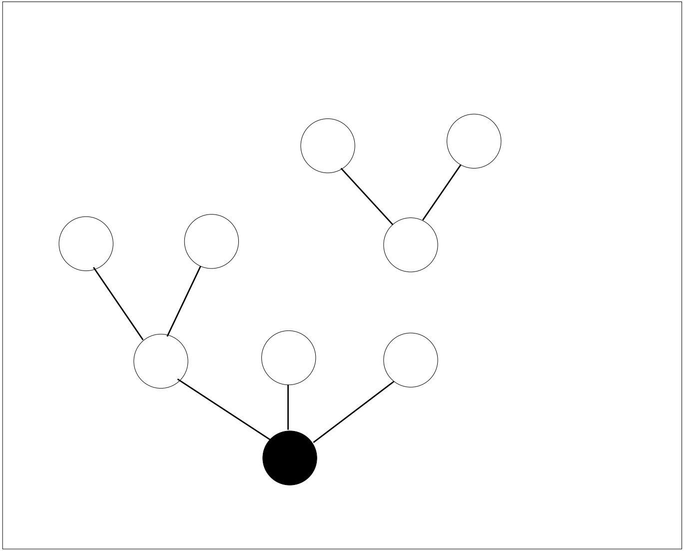
RRTs fulfill the remarkable splitting property (Meir & Moon):

Imagine that we remove an edge uniformly at random.

RRT then splits into two subtrees, say T' and T''.

The law of the sizes of T' and T'' is know explicitly, and conditionally on the latter, T' and T'' are two independent RRTs.





In the mid 70's, Meir & Moon considered the destruction of RRT, by cutting its edges one after the other, in uniformly random order.

They were interested in the number of cuts X_n occurring in the cluster containing the root 0, until the root is isolated and proved

$$X_n \sim \frac{n}{\ln n}.$$

Drmota, Iksanov, Möhle & Rösler (2009) proved that fluctuations are not Gaussian, but rather involve a totally asymmetric Cauchy law .

Iksanov and Möhle (2007) used the splitting property to construct a beautiful coupling between:

- the size of the root cluster during the destruction process,

- a random walk on $\mathbb{N}$ with step distribution

$$\frac{1}{j(j+1)}, \qquad j = 1, 2, \ldots .$$

(in the domain of attraction of a totally asymmetric Cauchy law).

Our purpose is to investigate the whole family of cluster sizes which arise during the destruction of a RRT with large size.

Just as for the Random Graph model $\mathcal{G}(n, p)$ with $p = p(n)$, each edge of RRT_n is kept with probability p and removed with probability $1 - p$, independently of the other edges.

This yields a random forest $\mathcal{T}(n, p)$. We shall identify various natural regimes.

Supercritical Regime

We first consider the regime in which the size of the root cluster starts decreasing:

$$1 - p(n) \sim \frac{a}{\log n}. \tag{1}$$

Then the root-cluster has size

$$C_0 \sim \mathrm{e}^{-a} n,$$

It is the unique giant component, and has non-Gaussian fluctuations.

Theorem (J. Schweinsberg (2012), J.B. (2014))

In the regime (1)

$$\left(n^{-1}C_0 - \mathrm{e}^{-a}\right)\ln n - a\mathrm{e}^{-a}\ln\ln n \Longrightarrow -a\mathrm{e}^{-a}(Z + \ln a)\ ,$$

where Z has a completely asymmetric Cauchy distribution:

$$\mathbb{E}(\mathrm{e}^{i\theta Z}) = \exp\left(-\frac{\pi}{2}|\theta| - i\theta\ln|\theta|\right)\ , \qquad \theta\in\mathbb{R}\,.$$

The next largest clusters, $C_1 \geq C_2 \geq \ldots$, fail short to be giant as well.

Theorem (J.B., RSA 2014)

In the regime (1)*, for every* $j \geq 1$*,*

$$\left(\frac{\ln n}{n} C_1, \dots, \frac{\ln n}{n} C_j\right)$$

converges in distribution as $n \to \infty$ *towards*

$$(\mathbf{x}_1, \dots, \mathbf{x}_j)$$

where $\mathbf{x}_1 > \mathbf{x}_2 > \dots$ *denotes the sequence of the atoms of a Poisson random measure on* $(0, \infty)$ *with intensity* $a\mathrm{e}^{-a} x^{-2} \mathrm{d}x$ *.*

Note that $x^{-2}\mathrm{d}x$ is the Lévy measure of a Cauchy process.

This was then considerably extended by Erich Baur, who considered the entire the supercritical regime

$$1/n \ll 1 - p(n) \ll 1$$

(in the case $1 - p(n) = o(1/n)$, no edge are cut).

The root cluster has then size $C_0 \sim n^{p(n)}$, and ...

Theorem (E. Baur, RSA 2016)

Suppose $(1-p(n))n^{p(n)} \to \infty$. *Then for every* $j \geq 1$,

$$\left(\frac{n^{-p(n)}}{1-p(n)}C_1, \dots, \frac{n^{-p(n)}}{1-p(n)}C_j\right)$$

converges in distribution as $n \to \infty$ *towards*

$$(\mathbf{x}_1, \dots, \mathbf{x}_j)$$

where $\mathbf{x}_1 > \mathbf{x}_2 > \dots$ *denotes the sequence of the atoms of a Poisson random measure on* $(0, \infty)$ *with intensity* $x^{-2}\mathrm{d}x$.

Subcritical Regime

The subcritical regime is

$$1/n \ll p(n) \ll 1$$

(in the case $p(n) = o(1/n)$, all the edges are cut).

One can characterize the regime for which the largest clusters have fixed a size $\ell \in \mathbb{N}$.

Theorem (E. Baur & J. B., 2016+)

Fix $\ell \geq 1$ and $a > 0$, and assume $p(n) \sim a n^{-1/\ell}$.
As $n \to \infty$, the number of clusters of size $\ell + 1$ is approximately

$$\mathrm{Poisson}(\ell! a^{\ell}).$$

Further, w.h.p., there are no clusters of size $> \ell + 1$.

Another interesting regime belonging to the sub-critical case is

$$p(n) \sim a/\ln n, \tag{2}$$

because then the root-cluster C_0 just starts growing, in the sense

$$C_0 \Longrightarrow \mathrm{Geometric}(\mathrm{e}^{-a}).$$

It is natural to ask about the size of the largest clusters in that regime.

Theorem (E. Baur & J. B., 2016+)

In the regime (2), w.h.p., the largest cluster has size

$$t^* \ln n - \frac{\ln \ln n}{2(\ln(at^*) - \ln(1+at^*))} + O(1)$$

where $t^* > 0$ *is the unique solution to*

$$f_a(t^*) = 1 + t^* \ln(at^*) - \frac{1+at^*}{a} \ln(1+at^*) = 0.$$

Actually, there is a sharper limit theorem, which involves a periodicity phenomenon (weak convergence only along certain sub-sequences).

Critical regime

The critical regime corresponds simply to constant percolation parameter.

Consider a dynamical setting: each edge has an exponential lifetime (after which it is cut), independently of the other edges.

Thus at time $t \geq 0$, we observe a Bernoulli bond percolation with parameter $p = \mathrm{e}^{-t}$.

The largest clusters have size of order $n^p = n^{\mathrm{e}^{-t}}$:

$$n^{-\mathrm{e}^{-t}} C_i \longrightarrow X_i(t) \qquad \text{for } i = 0, 1, 2, \ldots$$

We set $Y_i(t) = \ln X_i(t)$ and consider the family of points

$$\mathcal{Y}_t = \{Y_i(t) : i \geq 0\}.$$

Theorem (E. Baur & J.B. (EJP, 2015))

$\mathcal{Y} = (\mathcal{Y}_t : t \geq 0)$ *is a branching Ornstein-Uhlenbeck process:*
The law of $\mathcal{Y}_{s+t}$ *given* $\mathcal{Y}_s = \{y_i : i \geq 0\}$ *is that of*

$$\{\mathrm{e}^{-t} y_i + \zeta_j^{(i)} : i \geq 0, j \geq 0\}$$

where the families $\{\zeta_j^{(i)} : j \geq 0\}$ *are i.i.d. with the same law as* $\mathcal{Y}_t$*.*
Furthermore, for all $q > \mathrm{e}^t$*,*

$$\begin{aligned} \mathbb{E}\left(\sum \mathrm{e}^{qY_i(t)}\right) &= \mathbb{E}\left(\sum X_i^q(t)\right) \\ &= \frac{(q-1)}{(\mathrm{e}^{-t}q - 1)} \frac{\Gamma(q)}{\Gamma(\mathrm{e}^{-t}q)}. \end{aligned}$$

The process $\mathcal{X}_t = \{X_i(t) : i \geq 0\}$, $t \geq 0$, is a remarkable instance of a growth-fragmentation process.

The process of the size of the root cluster $(X_0(t))_{t\geq 0}$ has Mittag-Leffler one-dimensional marginals

$$\mathbb{P}(X_0(t) \in \mathrm{d}x)/\mathrm{d}x = \frac{\mathrm{e}^t}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^{k+1}}{k!} \Gamma(k\mathrm{e}^{-t} + 1) x^{k-1} \sin(\pi k \mathrm{e}^{-t});$$

and has been studied independently by M. Möhle (Alea, 2015).

Baur, E. Percolation on random recursive trees. *RSA* (2016).

Baur, E. & Bertoin, J. The fragmentation process of an infinite recursive tree and Ornstein-Uhlenbeck type processes. *EJP* (2015)

Baur, E. & Bertoin, J. Weak limits for the largest subpopulations in Yule processes with high mutation probabilities. ArXiV.

Bertoin, J. Non-Gaussian fluctuations of the giant cluster for percolation on random recursive trees. *EJP* (2014)

Bertoin, J. Sizes of the largest clusters for supercritical percolation on random recursive trees. *RSA* (2014)

Drmota, M., Iksanov, A., Möhle, M. and Rösler, U. A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *RSA* (2009)

Iksanov, A. & Möhle, M. A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *ECP* (2007)

Meir, A. and Moon, J. W. Cutting down recursive trees. *Math. Biosciences* (1974)

Möhle, M. The Mittag-Leffler process and a scaling limit for the block counting process of the Bolthausen-Sznitman coalescent. *ALEA* (2015)

Schweinsberg, J. Dynamics of the evolving Bolthausen-Sznitman coalescent. *EJP* (2012)