Almost giant clusters and fluctuations for percolation on large trees

Jean Bertoin

Institut für Mathematik
Universität Zürich

Erdős-Rényi random graph model in supercritical regime

Bond percolation on complete graph with n vertices

Super-critical regime, i.e. with parameter

$$p(n) \sim c/n \quad \text{with } c > 1.$$

There is a unique giant component with size

$$G_n(c) \sim \theta(c)n$$

The second, third, etc. largest clusters are almost microscopic, i.e. size of order $\ln n$.

Central Limit Theorem (Stepanov, Pittel, ...)

$$\frac{G_n(c) - \theta(c)n}{\sqrt{n}} \Longrightarrow \mathcal{N}(0, \sigma_c^2).$$

Percolation on a finite rooted tree

T_n with vertices $\{0, 1, \ldots, n\}$.

Percolation parameter $p(n)$ depends on n.

Basic problem : Identify the supercritical regimes as $n \to \infty$

(existence of a giant component)

$C^0_{p(n)}$ = size of cluster contains root 0.

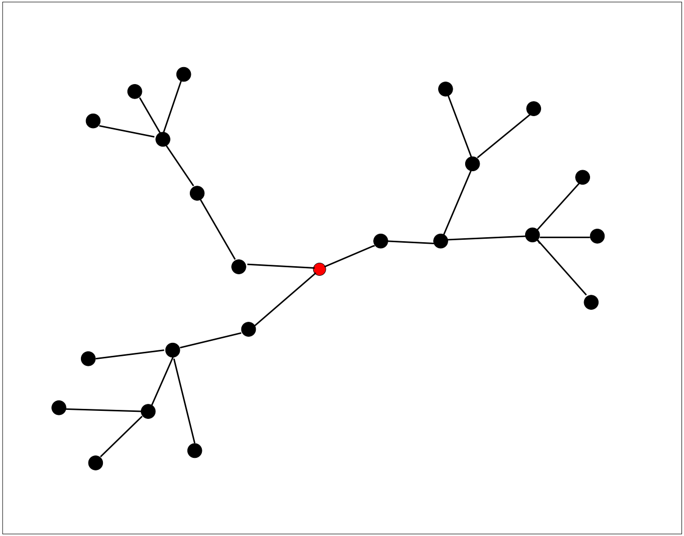
Consider the regime

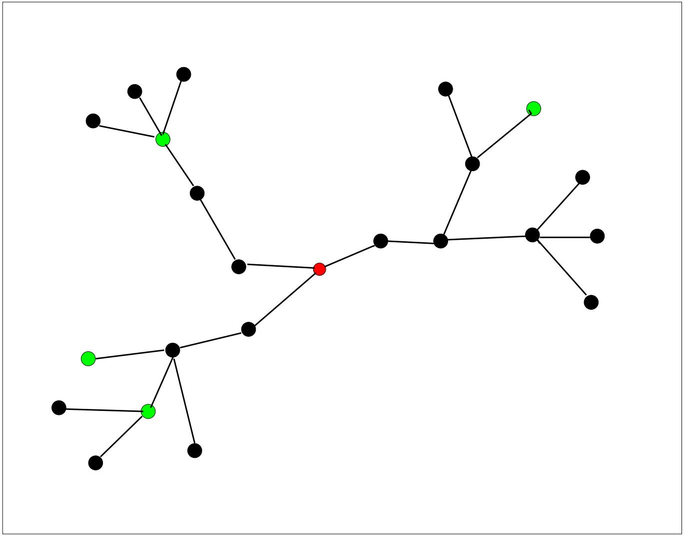
$$p(n) = 1 - \frac{c}{\ell(n)} + o(1/\ell(n)). \qquad (R_c)$$

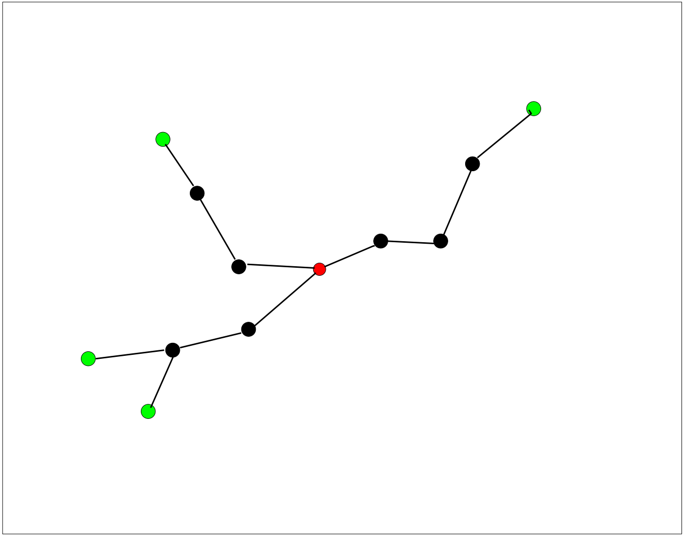
where $\lim_{n\to\infty} \ell(n) = \infty$, and $c \geq 0$.

$V_1, V_2, \ldots$ a sequence of i.i.d. uniform vertices.

$L_{k,n}$ = length of the tree reduced to $V_1, \ldots, V_k$ and the root 0







Lemma

If

$$\frac{L_{1,n}}{\ell(n)} \Rightarrow L_1 \text{ and } \frac{L_{2,n}}{\ell(n)} \Rightarrow L_1 + L_2$$

where L_1 *and* L_2 *are i.i.d., then in the regime* (R_c)

$$n^{-1} C^0_{p(n)} \longrightarrow \mathbb{E}(\exp(-cL)).$$

This is applies with $\ell(n) = \ln n$ to e.g.:

- d-regular trees, $L_i = 1/\ln d$

- random recursive trees, binary search trees, ... , $L_i = 1$.

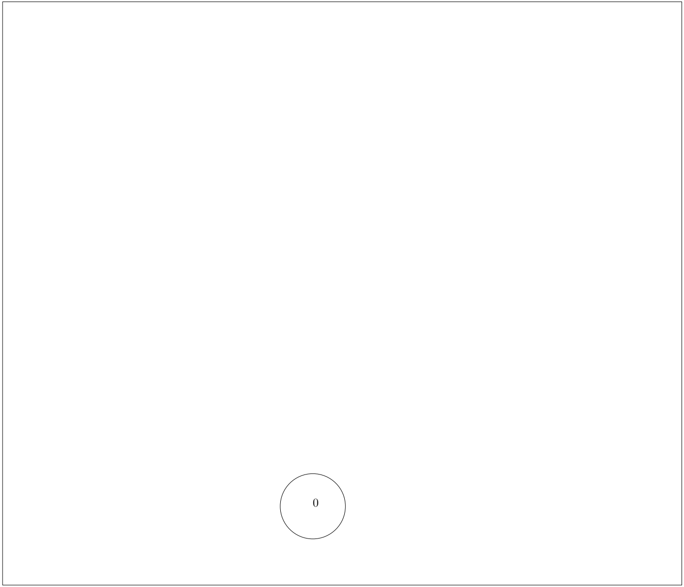
Recursive trees

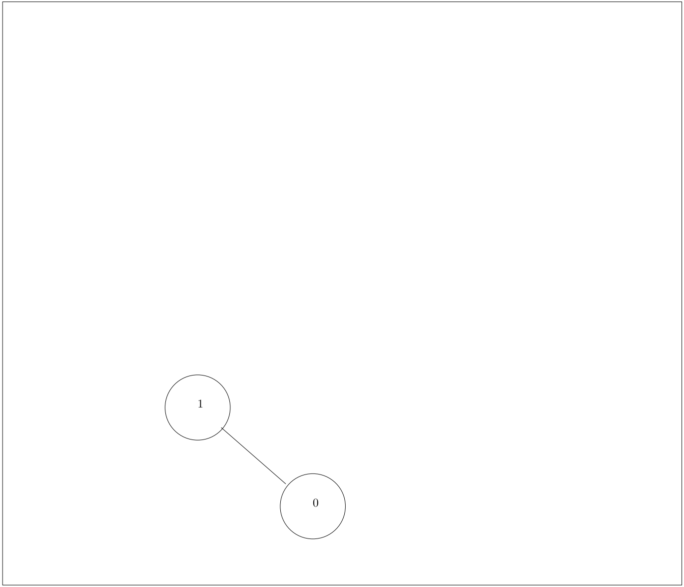
A tree on $\{0, 1, \ldots, n\}$ is called recursive if the sequence of vertices along any branch from the root 0 to a leaf is increasing.

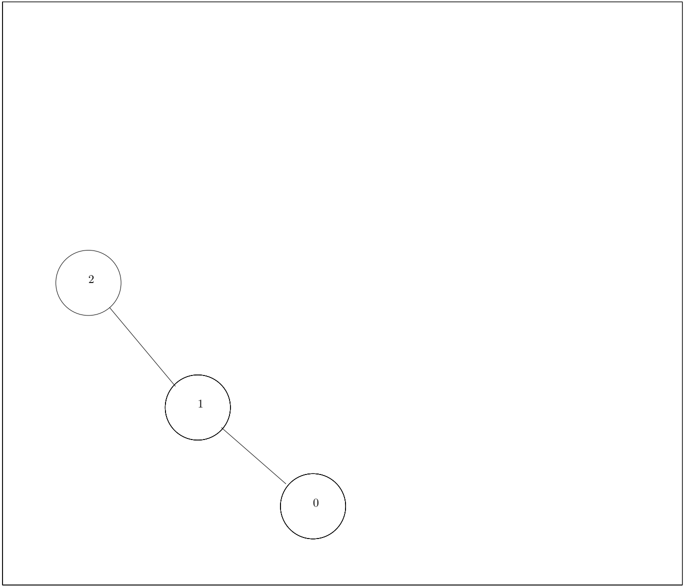
There are $n!$ such recursive trees, we pick one of them uniformly at random, denote it by T_n.

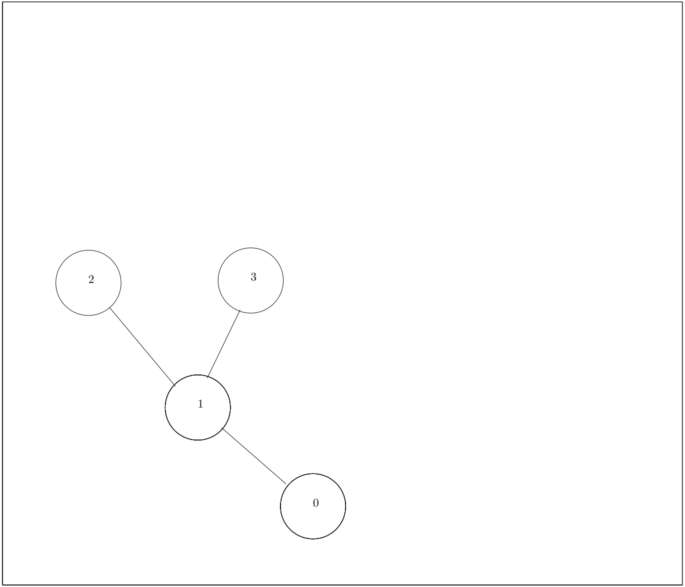
Simple algorithm to construct T_n:

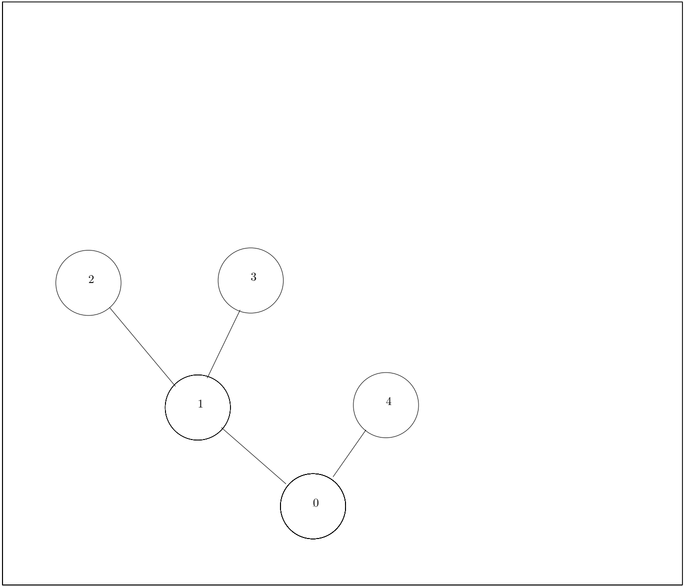
For $i = 1, 2, \ldots$, create an edge between i and $U(i)$ randomly chosen in $\{0, \ldots, i-1\}$, independently of the $U(j)$ for $j \neq i$.

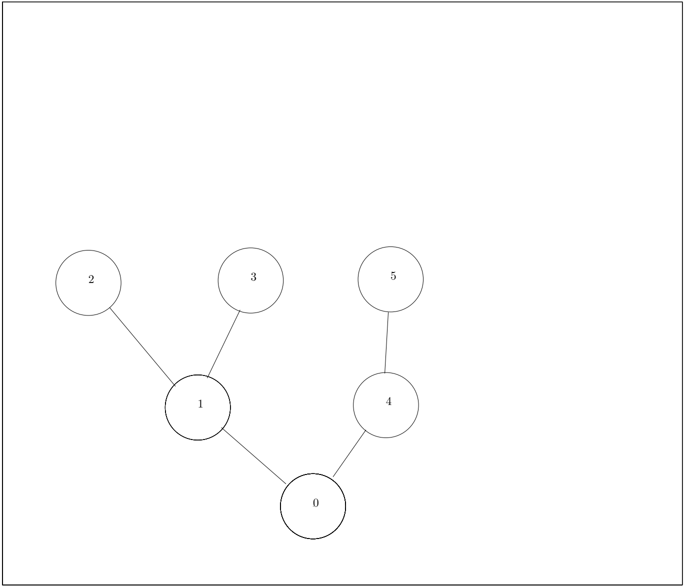

0

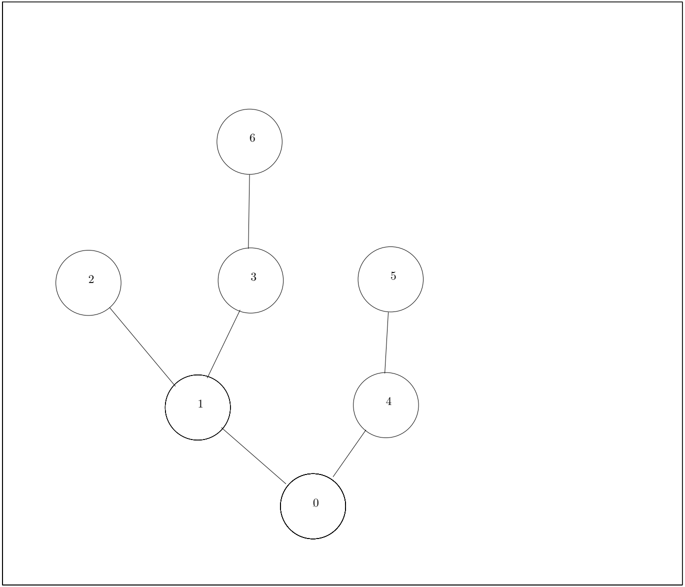

1
0


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

We know

$$C^0_{p(n)} \sim \mathrm{e}^{-c} n\,,$$

in the regime

$$p(n) = 1 - \frac{c}{\ln n} + o(1/\ln n)$$

Denote by

$$C^1_{p(n)} \geq C^2_{p(n)} \geq \dots$$

the sequence of the sizes of the other clusters ranked in the decreasing order.

Almost giant clusters

Theorem

For every fixed $j \geq 1$,

$$\left(\frac{\ln n}{n} C^1_{p(n)}, \ldots, \frac{\ln n}{n} C^j_{p(n)}\right) \Rightarrow (\mathbf{x}_1, \ldots, \mathbf{x}_j)$$

where $\mathbf{x}_1 > \mathbf{x}_2 > \ldots$ denotes the sequence of the atoms of a Poisson random measure on $(0, \infty)$ with intensity

$$c\mathrm{e}^{-c} x^{-2} \mathrm{d}x \,.$$

Some remarks

- The 2nd, 3rd, ... clusters are almost giant (only fail to be giant by a logarithmic factor).
- $1/\mathbf{x}_1, 1/\mathbf{x}_2 - 1/\mathbf{x}_1, \ldots, 1/\mathbf{x}_j - 1/\mathbf{x}_{j-1}$ are i.i.d. exponential variables with parameter $c\mathrm{e}^{-c}$.
  In particular $1/\mathbf{x}_j$ has the gamma distribution with parameter $(j, c\mathrm{e}^{-c})$.
- The parameter c only appears through a constant factor in the intensity measure. Maximal intensity for $c = 1$.

Some remarks

- The 2nd, 3rd, ... clusters are almost giant (only fail to be giant by a logarithmic factor).
- $1/\mathbf{x}_1, 1/\mathbf{x}_2 - 1/\mathbf{x}_1, \ldots, 1/\mathbf{x}_j - 1/\mathbf{x}_{j-1}$ are i.i.d. exponential variables with parameter $c\mathrm{e}^{-c}$.
  In particular $1/\mathbf{x}_j$ has the gamma distribution with parameter $(j, c\mathrm{e}^{-c})$.
- The parameter c only appears through a constant factor in the intensity measure. Maximal intensity for $c = 1$.

Some remarks

- The 2nd, 3rd, ... clusters are almost giant (only fail to be giant by a logarithmic factor).
- $1/\mathbf{x}_1, 1/\mathbf{x}_2 - 1/\mathbf{x}_1, \ldots, 1/\mathbf{x}_j - 1/\mathbf{x}_{j-1}$ are i.i.d. exponential variables with parameter $c\mathrm{e}^{-c}$.
  In particular $1/\mathbf{x}_j$ has the gamma distribution with parameter $(j, c\mathrm{e}^{-c})$.
- The parameter c only appears through a constant factor in the intensity measure. Maximal intensity for $c = 1$.

Percolation and isolation of the root

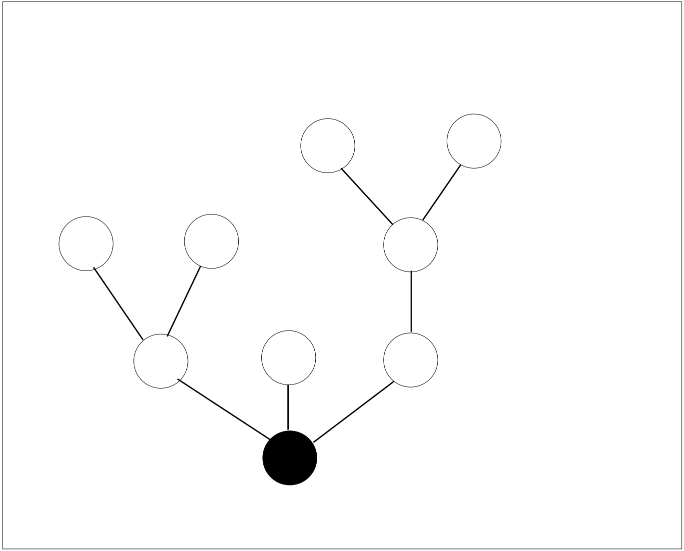
Basic ideas for the proof:

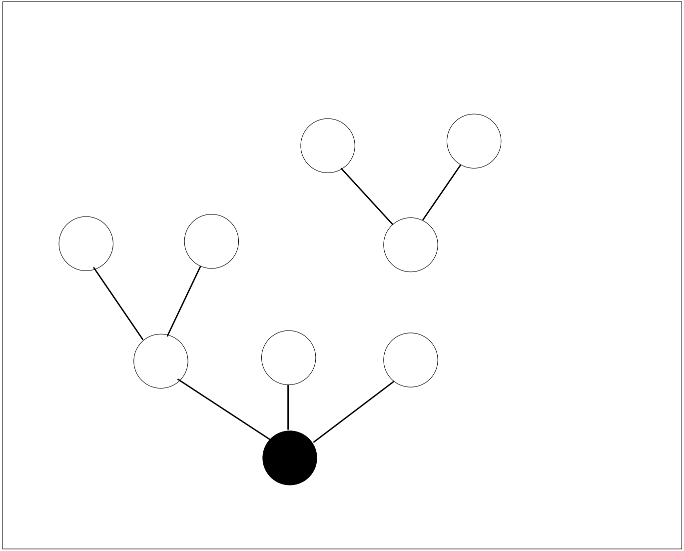
• relate percolation to an algorithm in combinatorics (Meir and Moon) for isolating the root in a tree,

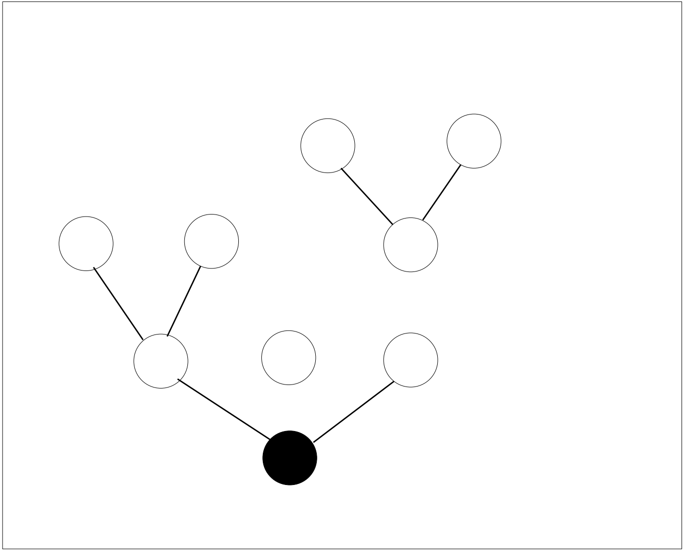
• use a coupling (Iksanov and Möhle) with a remarkable random walk.

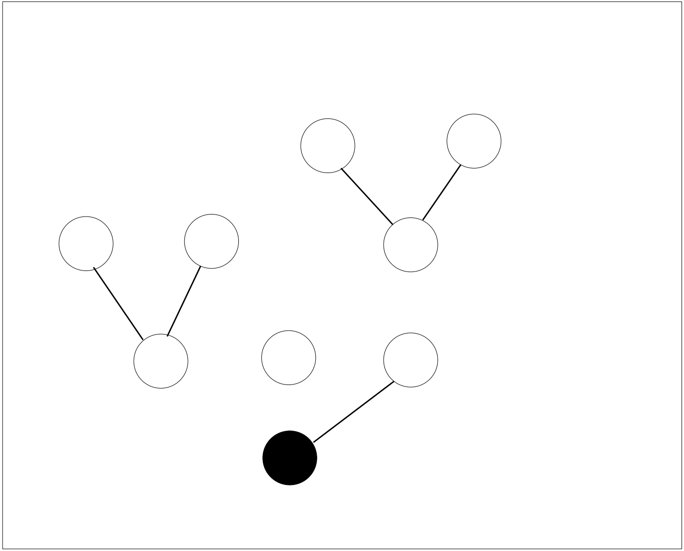
Pick an edge uniformly at random in the tree, remove it and then discard the entire subtree generated by that edge.

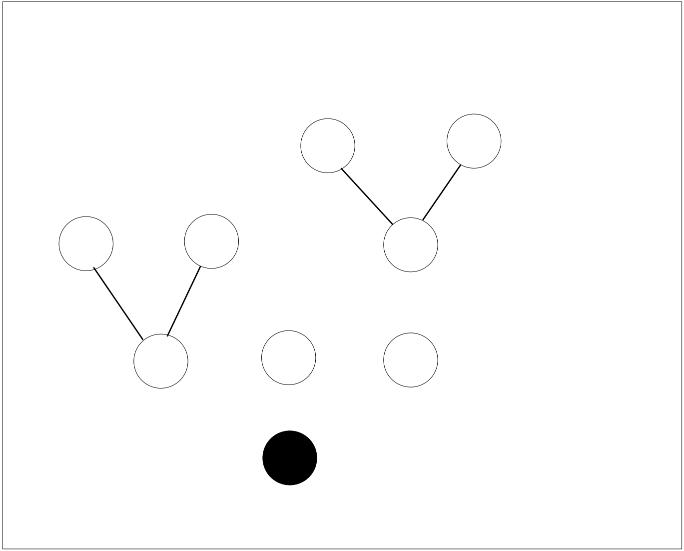
Iterate until the root has been isolated.











Iksanov-Möhle coupling with a random walk

Consider the random walk

$$S_j = \xi_1 + \ldots + \xi_j \,, \qquad j \in \mathbb{N}$$

where

$$\mathbb{P}(\xi = \ell) = \frac{1}{\ell(\ell+1)} \,.$$

Introduce the first passage time

$$N(n) = \min\{j \geq 1 : S_j > n\} \,.$$

Lemma

[Iksanov and Möhle]

One can couple S and the isolation of the root algorithm such that: For every $k < N(n)$,

$$(|V_1|, \dots, |V_k|, |V'_k|) = (\xi_1, \dots, \xi_k, n + 1 - S_k)\,,$$

where $|V_i|$ denotes the size of the component removed at the i-th step, and $|V'_k|$ the size of the root-component after k steps.

In turn, there is a dynamical version of percolation which can then be coupled with the preceding algorithm.

Remove each edge e of T_n at an independent exponential variable with parameter $1/\ln n$.

At time

$$t(n) = -\ln n \times \ln p(n) \sim c$$

one observes the percolation with parameter $p(n)$.

Then modify by instantaneously freezing clusters that do not contain the root (i.e. edges are only removed when they belong to the cluster that contains the root).

We obtain a continuous time version of the algorithm for isolating the root.

We recover percolation from the root-isolation algorithm by performing additional percolation on components which have been frozen.

This coupling enables us to reduce the study of the component sizes in the isolation of the root algorithm to extreme values theory for large sequences of i.i.d. variables.

Anomalous fluctuations of the giant component

The series $\sum \mathtt{x}_i$ diverges a.s.

Fluctuations of the giant cluster cannot be deduced from the behavior of the 2nd, 3rd, ... largest clusters.

However the coupling with the random walk S can be used again.

Theorem

There is the weak convergence

$$\left(n^{-1}C^0_{p(n)} - \mathrm{e}^{-c}\right)\ln n - c\mathrm{e}^{-c}\ln\ln n \Longrightarrow -c\mathrm{e}^{-c}(Z + \ln c)\ ,$$

where

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

This result was proven first by J. Schweinsberg (2012) in terms of the Bolthausen-Sznitman coalescent, using delicate estimates for the evolution of the number of blocks in the latter.

Goldschmidt and Martin observed that isolation of the root in a recursive tree is closely related to the Bolthausen-Sznitman coalescent.

This result was proven first by J. Schweinsberg (2012) in terms of the Bolthausen-Sznitman coalescent, using delicate estimates for the evolution of the number of blocks in the latter.

Goldschmidt and Martin observed that isolation of the root in a recursive tree is closely related to the Bolthausen-Sznitman coalescent.

The coupling with a random walk enables to see first the birth of the anomalous fluctuations of the giant cluster by interrupting the construction of the Random Recursive Tree when it has size $\ln^4 n$.

Then one verifies the regular spread of those fluctuations when one resumes the recursive construction, using arguments from branching processes.

The coupling with a random walk enables to see first the birth of the anomalous fluctuations of the giant cluster by interrupting the construction of the Random Recursive Tree when it has size $\ln^4 n$.

Then one verifies the regular spread of those fluctuations when one resumes the recursive construction, using arguments from branching processes.