Almost giant clusters for percolation on large trees

Jean Bertoin

Institut für Mathematik
Universität Zürich

Erdős-Rényi random graph model in supercritical regime

$G_n =$ complete graph with n vertices

Bond percolation with parameter

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

There is a unique giant component with size $\sim \theta(c)n$

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

Percolation on trees much simpler than on general graphs or lattices because unique path between two vertices.

E.g. percolation on the infinite regular k-tree
$\sim$ branching process with Bin(k,p) reproduction law.

Exist infinite clusters if and only if $kp > 1$.

Here, we consider percolation on a finite tree T_n with size $n \gg 1$

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

Two questions:

1) What are the supercritical regimes (existence of a giant component) ?

2) Estimate the sizes of the 2nd, 3rd, ..., largest clusters.

Characterization of supercritical regimes

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

Bernoulli bond percolation on T_n with parameter $p(n)$.

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

$C^0_{p(n)}$ is giant if $n^{-1} C^0_{p(n)} \Longrightarrow G \not\equiv 0$.

$\ell : \mathbb{N} \to \mathbb{R}_+$ with $\lim_{n\to\infty} \ell(n) = \infty$, and $c \geq 0$.

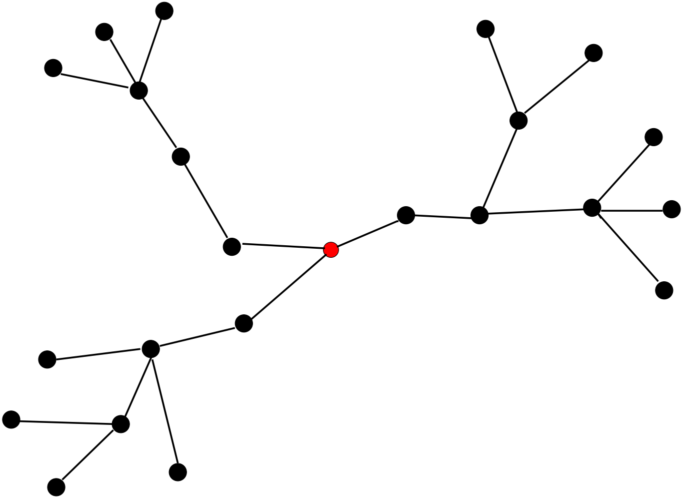
Consider the regime

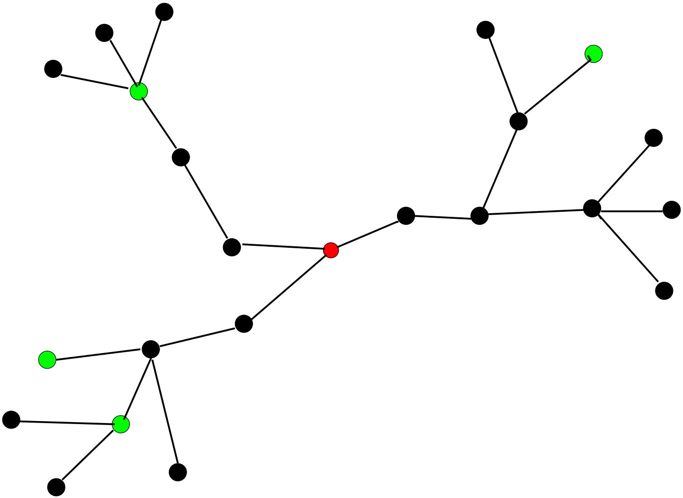
$$p(n) = 1 - \frac{c}{\ell(n)} + o(1/\ell(n)). \qquad (R_c)$$

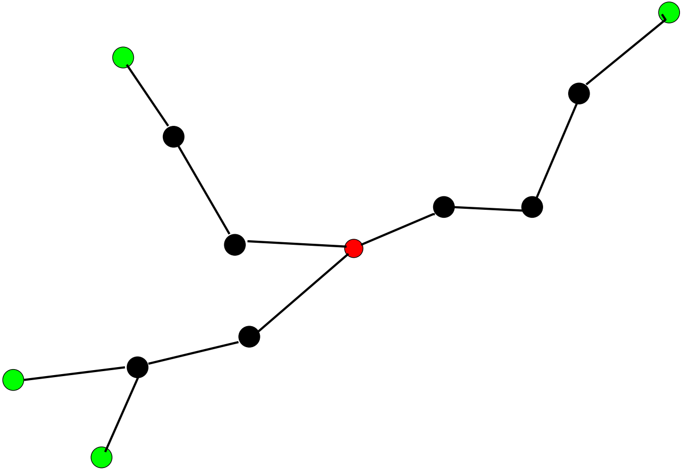
$V_1, V_2, \ldots$ a sequence of i.i.d. uniform vertices.

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

$$\frac{1}{\ell(n)} L_{k,n} \Longrightarrow L_k. \qquad (H_k)$$







Theorem

(i) *If* (H_k) *holds for all* k*, then in the regime* (R_c)

$$n^{-1} C^0_{p(n)} \Rightarrow G(c)\,, \tag{1}$$

where

$$\mathbb{E}(G(c)^k) = \mathbb{E}(\mathrm{e}^{-cL_k}). \tag{2}$$

(ii) *If* (1) *holds in the regime* (R_c) *for all* $c > 0$ *with* $\lim_{c\to 0+} G(c) = 1$,
then (H_k) *is fulfilled for all* k *and* (2) *holds.*

Sketch of proof :

$$\mathbb{E}\left(\left((n+1)^{-1}C^0_{p(n)}\right)^k\right) = \mathbb{E}\left(p(n)^{L_{k,n}}\right).$$

In the regime (R_c), $p(n) \sim \exp(-c/\ell(n))$ and (H_k) yields

$$\lim_{n\to\infty} \mathbb{E}\left(\exp\left(-\frac{c}{\ell(n)}L_{k,n}\right)\right) = \mathbb{E}(\mathrm{e}^{-cL_k}).$$

(H_k) is known to hold for a number of families of (random) trees:

- Cayley trees (Aldous) with $\ell(n) = \sqrt{n}$ and $L_k \sim \mathrm{Chi}(2k)$.

- d-regular trees with $\ell(n) = \ln n$ and $L_k = k / \ln d$

- random recursive trees, binary search trees, ... with $\ell(n) = \ln n$ and $L_k = k$.

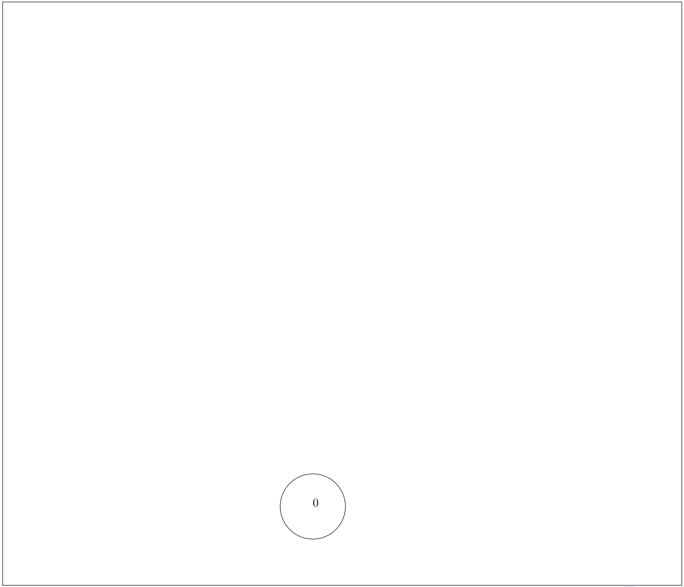
Recursive trees

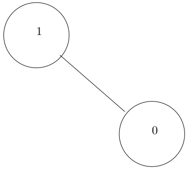
A tree on $\{0, 1, \ldots, n\}$ is called recursive if the sequence of vertices along any branch from the root 0 to a leaf is increasing.

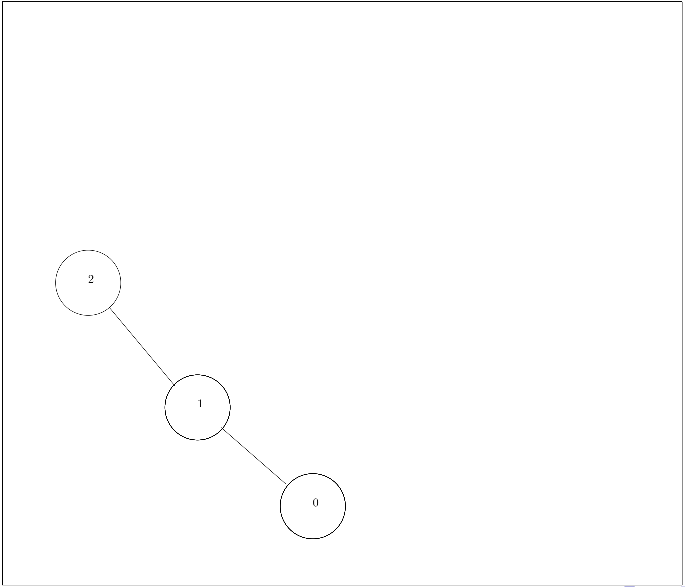
There are $n!$ such recursive trees, we pick one of them uniformly at random, denote it by T_n.

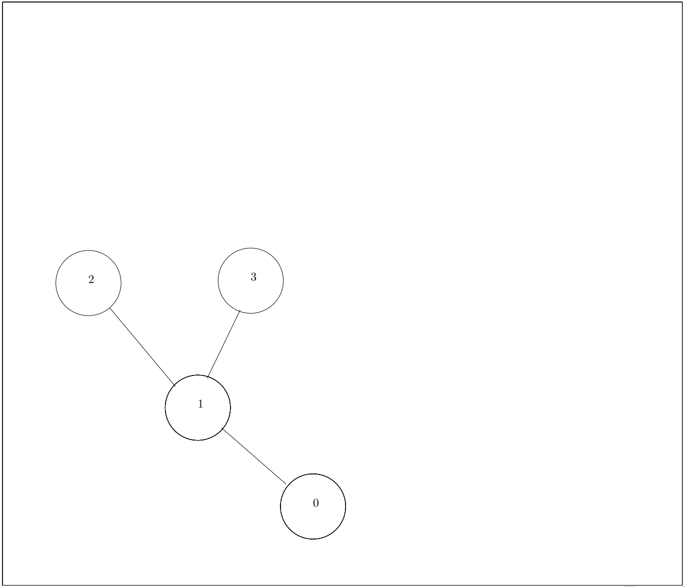
Simple algorithm to construct T_n:

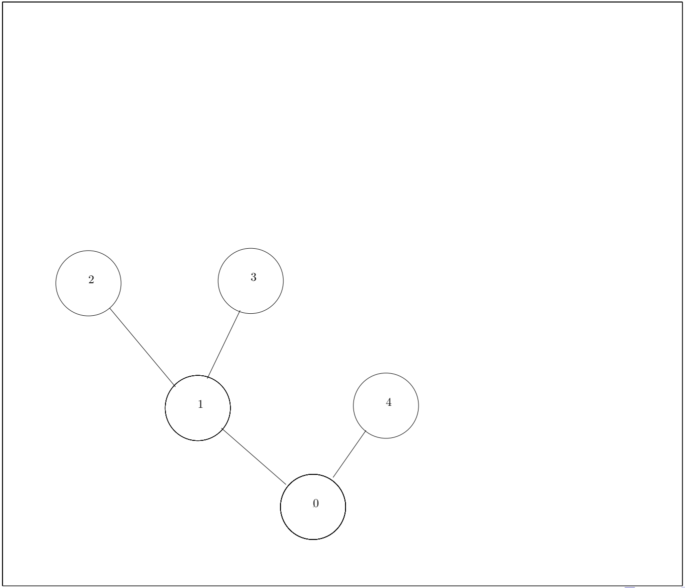
For $i = 1, 2, \ldots$, create an edge between i and $U(i)$ randomly chosen in $\{0, \ldots, i-1\}$, independently of the $U(j)$ for $j \neq i$.

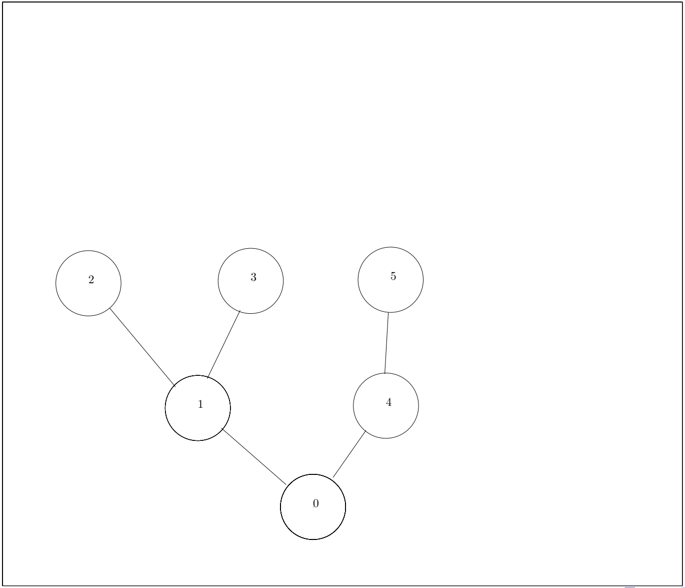

0

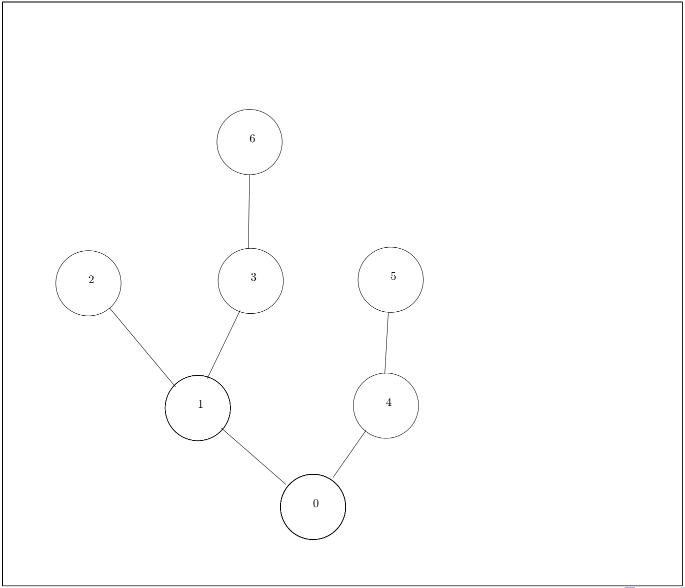

1
0


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

(H_k) holds with $\ell(n) = n$ and then

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

in the regime

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

Denote by

$$C_1(n) \geq C_2(n), \ldots$$

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

Almost giant clusters

Theorem

For every fixed $j \geq 1$,

$$\left(\frac{\ln n}{n}C_1, \ldots, \frac{\ln n}{n}C_j\right) \Rightarrow (\mathbf{x}_1, \ldots, \mathbf{x}_j)$$

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

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

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:

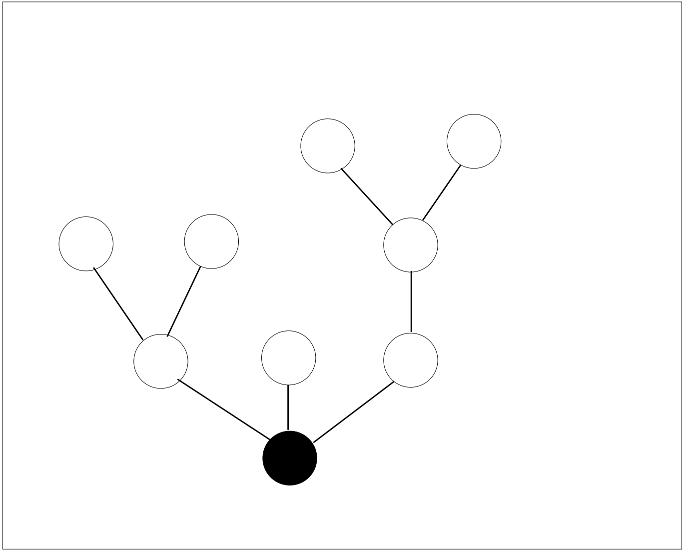
• relate percolation to an algorithm in combinatorics for isolating the root in a tree,

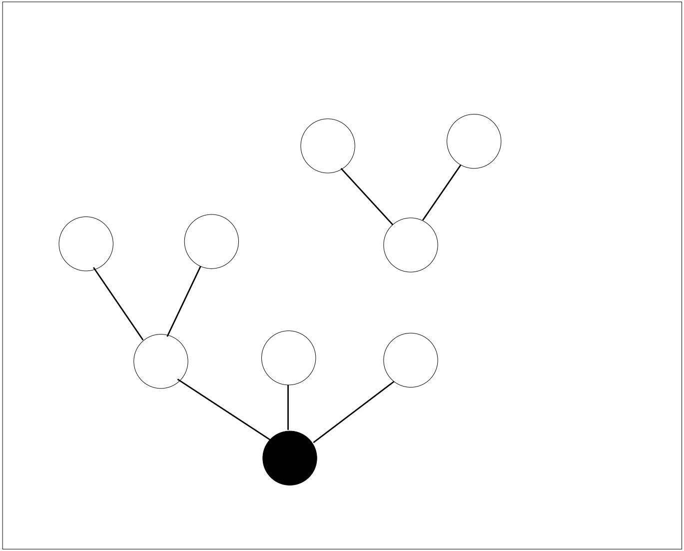
• use a coupling with a certain random walk which was pointed at by Iksanov and Möhle.

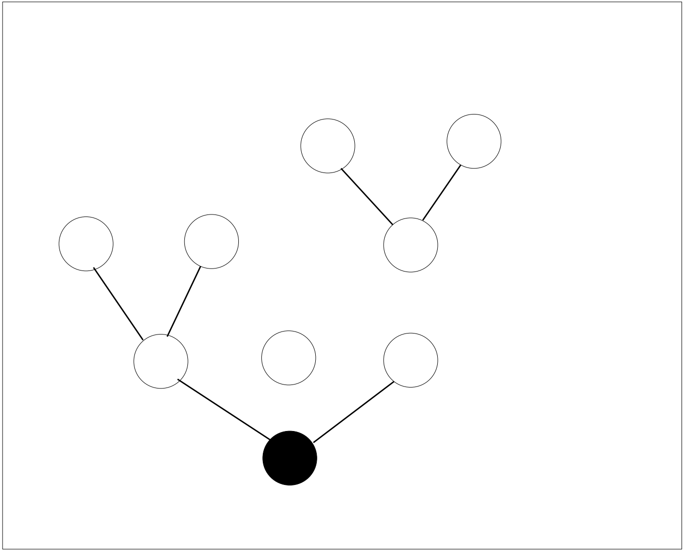
Meir and Moon (1970+) introduced the following random algorithm on rooted trees.
See also Janson, Panholzer, Holmgren, Iksanov and Möhle, ...

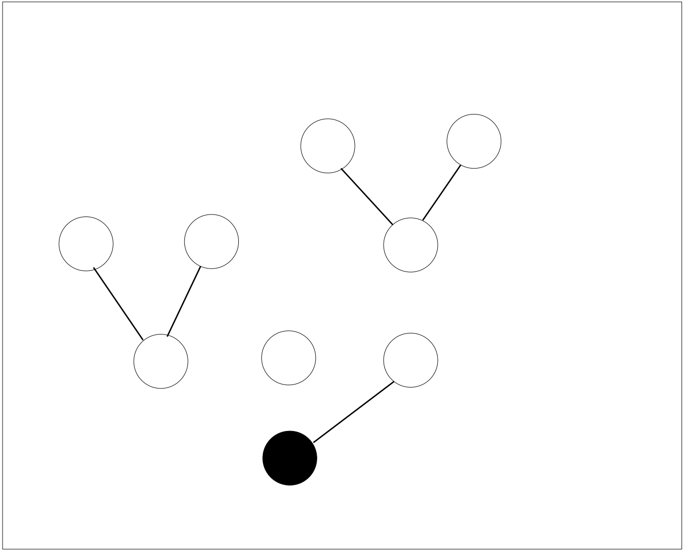
Pick an edge uniformly at random in the tree, remove it and then discard the entire subtree generated by that edge.

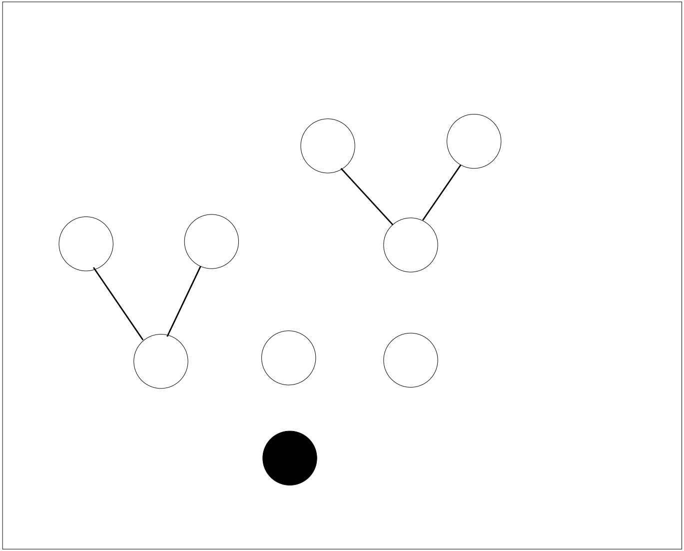
Iterate until the root has been isolated.











We will use a dynamical version of percolation which yields a natural coupling with the preceding algorithm.

To distinguish between the two structures, we use the term clusters (respectively components) for the connected subsets of vertices which arise from percolation (respectively from the root-isolation algorithm).

Attach to each edge e of T_n an independent exponential variable with parameter $1/\ln n$, say $\varepsilon(e)$.

If we remove e at time $\varepsilon(e)$, then we observe at time

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

a Bernoulli bond-percolation on T_n with parameter $p(n)$.

The choice of parametrization is such that $t(n) \to t$.

Now modify this dynamical percolation 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.

Conversely, we can recover percolation from the root-isolation algorithm by performing additional percolation on components which have been frozen:

In short, the largest percolation clusters can be recovered from the largest components in the isolation of the root algorithm, together with the times at which they appear.

Information on the latter can be derived from a coupling with a certain random walk (Iksanov and Möhle).

Coupling with a random walk

The coupling relies on a couple of basic properties of random recursive trees.

Fractal property of T_n :
removing any given edge disconnects T_n into two subtrees which, conditionally on their sizes, are independent random recursive trees.

Second, remove from T_n an edge uniformly at random.
Let $\tilde{T}$ = subtree $\not\ni 0$. Then

$$\mathbb{P}(|\tilde{T}| = \ell) = \mathbb{P}(\xi = \ell \mid \xi \leq n)\,, \qquad \ell = 1, \ldots, n\,,$$

where

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

This incites us to introduce

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

with ξ_i i.i.d. copies of ξ, and 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 $k < N(n)$,

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

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.

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.

Scale free trees

Scale-free random trees grow via preferential attachment algorithm (Barabási-Albert).

Fix $\beta > -1$ and suppose that $T_n^{(\beta)}$ has been constructed.

Denote by $d_n(i)$ the degree of the vertex i in $T_n^{(\beta)}$.

Then incorporate $n+1$ with an edge linking to a random vertex $v_n \in \{0, \ldots, n\}$ with law

$$\mathbb{P}(v_n = i) = \frac{d_n(i) + \beta}{2n + \beta(n+1)}, \qquad i \in \{0, \ldots, n\}.$$

One checks that $C^0_{p(n)} \sim \mathrm{e}^{-c(1+\beta)/(2+\beta)} n$ and

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*

$$c\mathrm{e}^{-c(1+\beta)/(2+\beta)} x^{-2} \mathrm{d}x \, .$$