Statistical Mechanics Seminar, Warwick, 18th February 2010

The scaling limit of critical random graphs

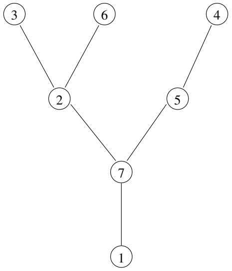
Christina Goldschmidt

Joint work with Louigi Addario-Berry (McGill University) and Nicolas Broutin (INRIA Rocquencourt)

Part I : Trees

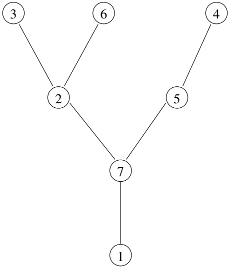
A warm-up: uniform random trees

Take a uniform random tree T_m on vertices labelled by $[m] = \{1, 2, \ldots, m\}$.



A warm-up: uniform random trees

Take a uniform random tree T_m on vertices labelled by $[m] = \{1, 2, \ldots, m\}$.



What happens as m grows?

Useful link to branching processes

A uniform random tree on m vertices has the same distribution as

Useful link to branching processes

A uniform random tree on m vertices has the same distribution as

- the family tree of a Galton-Watson branching process

Useful link to branching processes

A uniform random tree on m vertices has the same distribution as

- the family tree of a Galton-Watson branching process
- with Poisson(1) offspring distribution

Useful link to branching processes

A uniform random tree on m vertices has the same distribution as

- the family tree of a Galton-Watson branching process
- with Poisson(1) offspring distribution
- conditioned to have precisely m vertices

Useful link to branching processes

A uniform random tree on m vertices has the same distribution as

- the family tree of a Galton-Watson branching process
- with Poisson(1) offspring distribution
- conditioned to have precisely m vertices
- and with a uniformly-chosen labelling.

Useful link to branching processes

A uniform random tree on m vertices has the same distribution as

- the family tree of a Galton-Watson branching process
- with Poisson(1) offspring distribution
- conditioned to have precisely m vertices
- and with a uniformly-chosen labelling.

(The following theory also works for any Galton-Watson branching process having offspring mean 1 and finite offspring variance.)

Two ways of encoding a tree

It will be convenient to encode our trees in terms of discrete functions which are easier to manipulate.

Two ways of encoding a tree

It will be convenient to encode our trees in terms of discrete functions which are easier to manipulate.

We will do this is two different ways:

- the height function
- the depth-first walk.

Height function

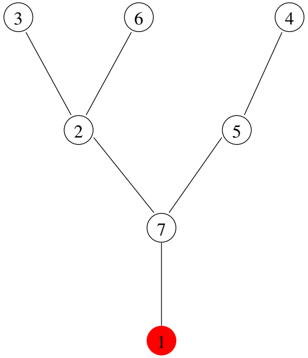
We will think of the lowest-labelled vertex as the root.

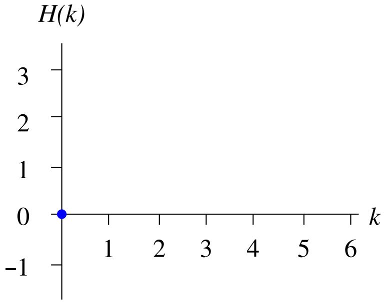
Height function

We will think of the lowest-labelled vertex as the root.

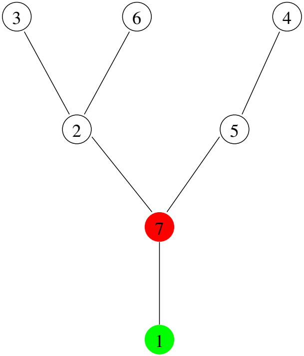
Consider the vertices in depth-first order and sequentially record the distance from the root.

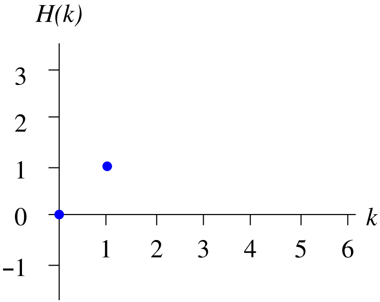
Height function



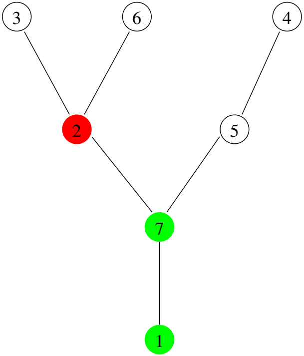


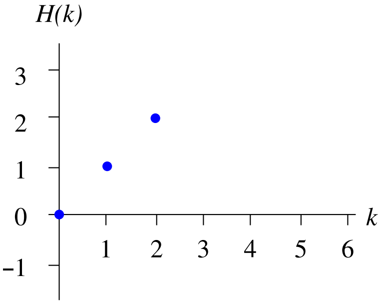
Height function



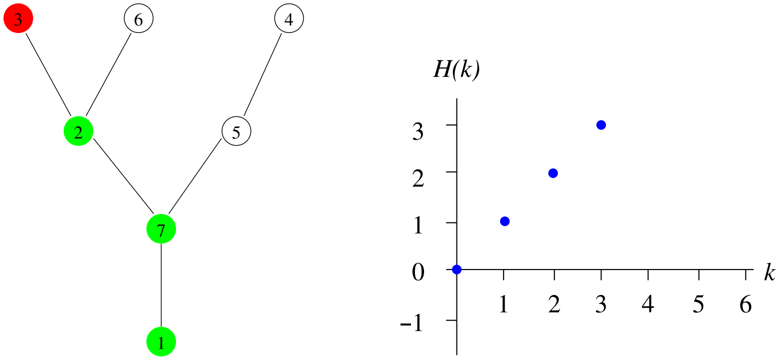


Height function

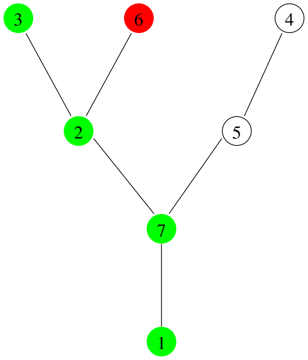


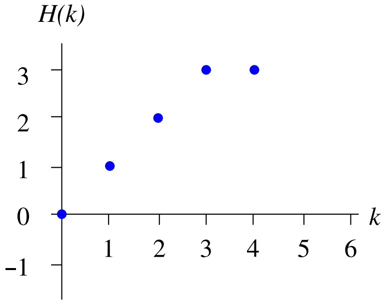


Height function

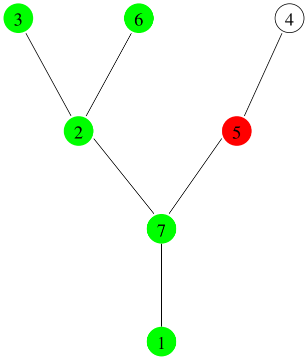


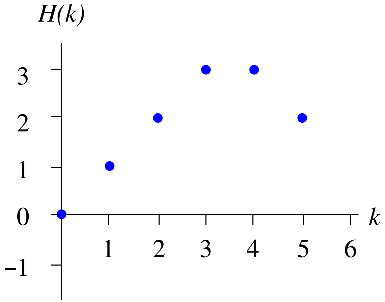
Height function



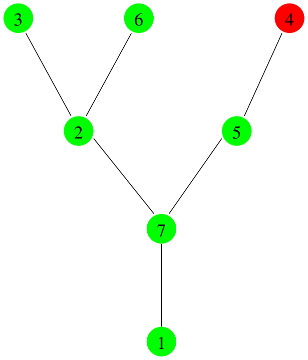


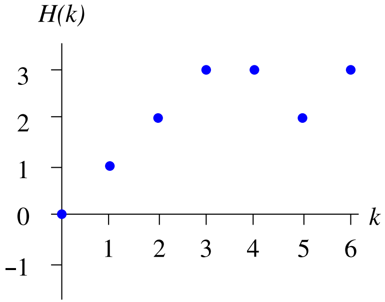
Height function





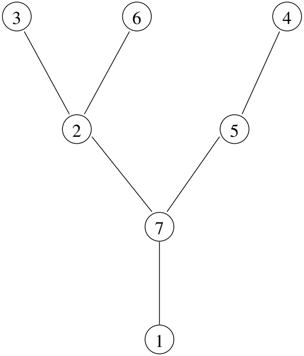
Height function

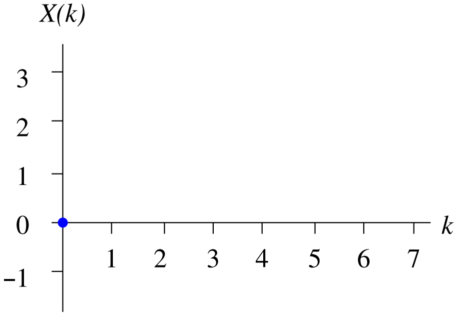




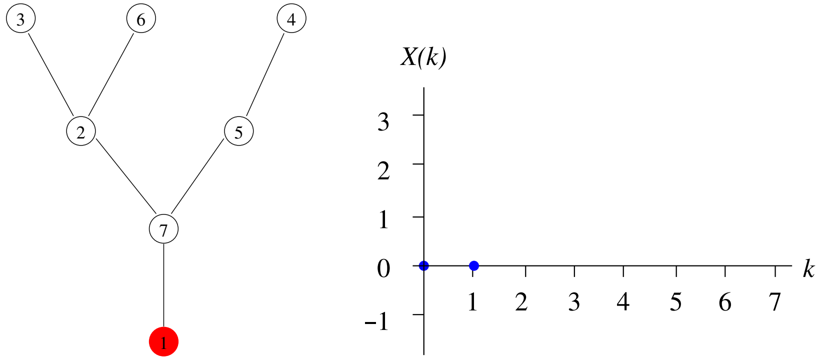
Depth-first walk

We again consider the vertices in depth-first order but now at each step we add an increment consisting of the number of children minus 1. The walk starts from 0.

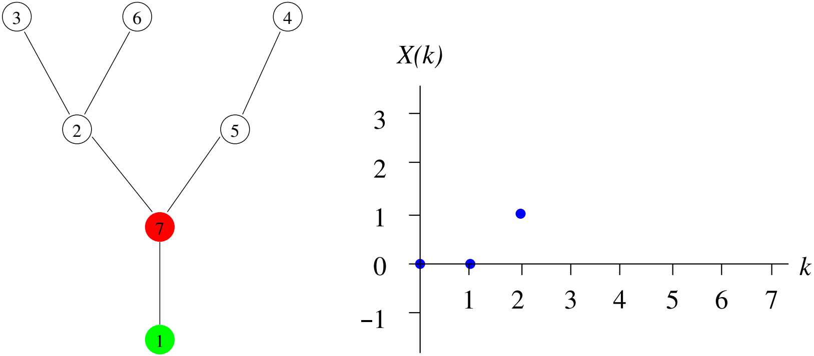




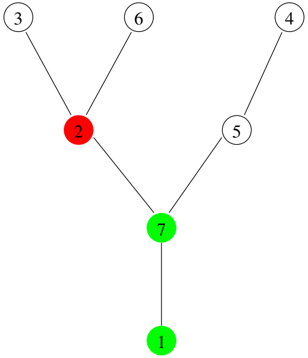
Depth-first walk

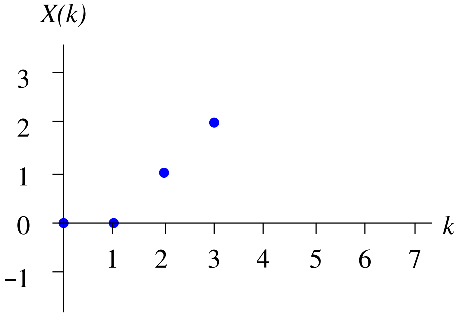


Depth-first walk

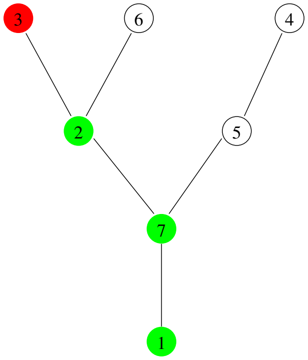


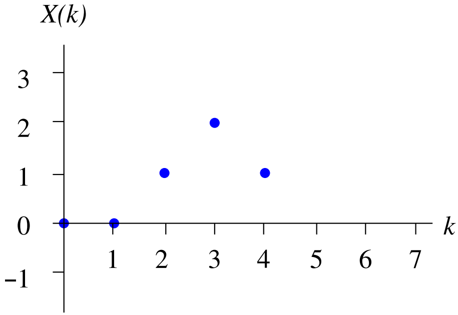
Depth-first walk



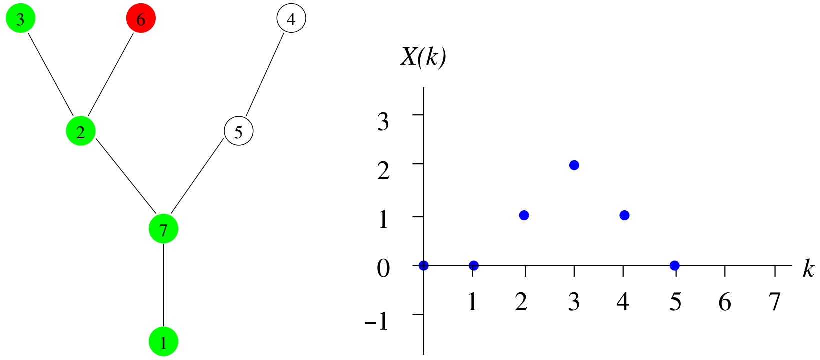


Depth-first walk

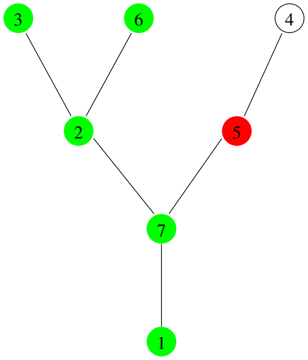


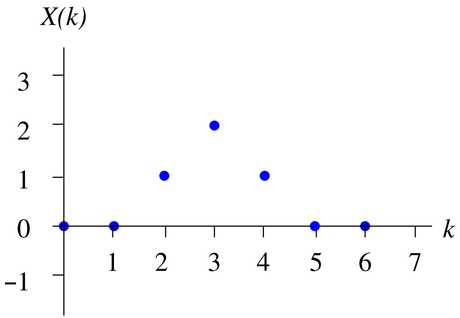


Depth-first walk

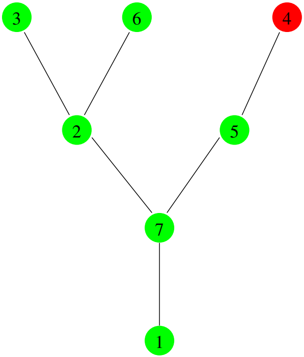


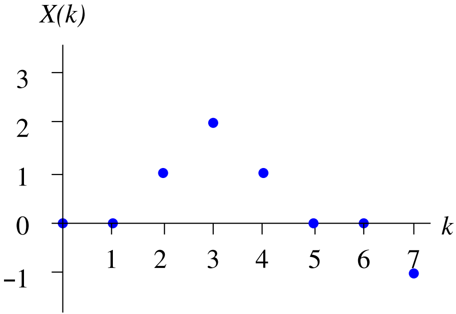
Depth-first walk





Depth-first walk





Comparing encodings

It is fairly straightforward to see that the height function encodes the topology of the tree (although not its labels).

Comparing encodings

It is fairly straightforward to see that the height function encodes the topology of the tree (although not its labels).

It is less easy to see that the depth-first walk also encodes the topology. In fact,

$$H(i) = \#\left\{0 \leq j \leq i-1 : X(j) = \min_{j \leq k \leq i} X(k)\right\}.$$

Comparing encodings

It is fairly straightforward to see that the height function encodes the topology of the tree (although not its labels).

It is less easy to see that the depth-first walk also encodes the topology. In fact,

$$H(i) = \# \left\{ 0 \le j \le i - 1 : X(j) = \min_{j \le k \le i} X(k) \right\}.$$

The advantage of the depth-first walk is that we can more easily understand its distribution.

Distribution of the depth-first walk

Suppose that we had a Poisson-Galton-Watson(1) branching process without any condition on the total progeny. Then at each step of the depth-first walk we would add an independent increment whose distribution is

$$\text{Poisson}(1) - 1$$

until the first time T that the walk hits -1 (which signals the end of the component).

Distribution of the depth-first walk

Suppose that we had a Poisson-Galton-Watson(1) branching process without any condition on the total progeny. Then at each step of the depth-first walk we would add an independent increment whose distribution is

$$\text{Poisson}(1) - 1$$

until the first time T that the walk hits -1 (which signals the end of the component).

In other words, we have a random walk with step-sizes having mean 0 and finite variance. The only complication is that we have to condition it on $T = m$.

Taking limits

By Donsker's theorem, the unconditioned walk run for n steps and with space rescaled by $1/\sqrt{n}$ converges to a Brownian motion run for time 1.

Taking limits

By Donsker's theorem, the unconditioned walk run for n steps and with space rescaled by $1/\sqrt{n}$ converges to a Brownian motion run for time 1.

It turns out to be also true that the random walk conditioned on $T = m$, with space rescaled by $1/\sqrt{m}$, converges in distribution to a limit, called a Brownian excursion.

Taking limits

By Donsker's theorem, the unconditioned walk run for n steps and with space rescaled by $1/\sqrt{n}$ converges to a Brownian motion run for time 1.

It turns out to be also true that the random walk conditioned on $T = m$, with space rescaled by $1/\sqrt{m}$, converges in distribution to a limit, called a Brownian excursion.

Intuitively, this is a Brownian motion started from 0, conditioned to leave 0 immediately and to stay positive until it returns to 0 at time 1.

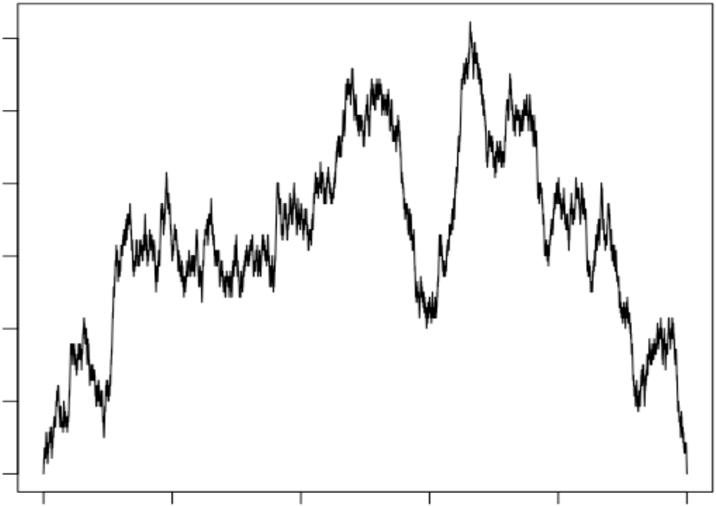
Taking limits

By Donsker's theorem, the unconditioned walk run for n steps and with space rescaled by $1/\sqrt{n}$ converges to a Brownian motion run for time 1.

It turns out to be also true that the random walk conditioned on $T = m$, with space rescaled by $1/\sqrt{m}$, converges in distribution to a limit, called a Brownian excursion.

Intuitively, this is a Brownian motion started from 0, conditioned to leave 0 immediately and to stay positive until it returns to 0 at time 1. (Of course, some work is necessary to make good sense of this, since the conditioning is singular!)

Brownian excursion



Taking limits

Formally, we have

$$(m^{-1/2}X^m(\lfloor mt \rfloor), 0 \leq t < 1) \xrightarrow{d} (e(t), 0 \leq t < 1)$$

as $m \to \infty$.

Taking limits

Formally, we have

$$(m^{-1/2}X^m(\lfloor mt \rfloor), 0 \leq t < 1) \xrightarrow{d} (e(t), 0 \leq t < 1)$$

as $m \to \infty$.

It is also possible to prove that

$$(m^{-1/2}H^m(\lfloor mt \rfloor), 0 \leq t < 1) \xrightarrow{d} (e(t), 0 \leq t < 1)$$

as $m \to \infty$.

Taking limits

Formally, we have

$$(m^{-1/2}X^m(\lfloor mt \rfloor), 0 \leq t < 1) \xrightarrow{d} (e(t), 0 \leq t < 1)$$

as $m \to \infty$.

It is also possible to prove that

$$(m^{-1/2}H^m(\lfloor mt \rfloor), 0 \leq t < 1) \xrightarrow{d} (e(t), 0 \leq t < 1)$$

as $m \to \infty$.

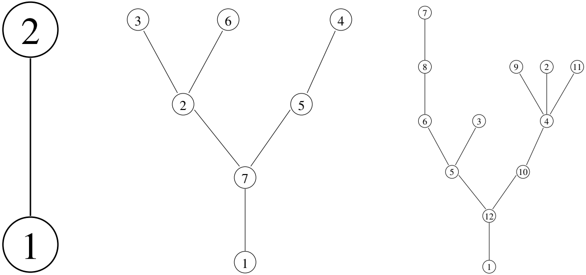
This suggests that there is some sort of limiting object for the tree itself, which should somehow be encoded by the Brownian excursion.

Scaling limit for the tree

Consider the tree as a metric space with the natural metric being given by the graph distance.

Scaling limit for the tree

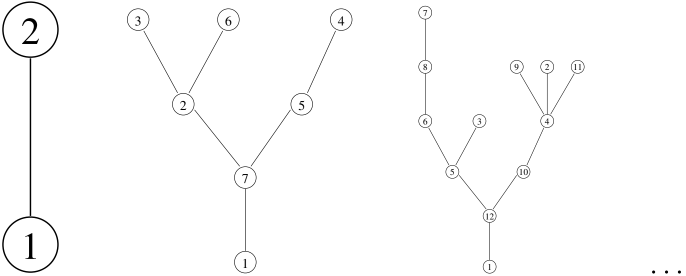
Consider the tree as a metric space with the natural metric being given by the graph distance. Rescale the edge-lengths by $1/\sqrt{m}$:



. . .

Scaling limit for the tree

Consider the tree as a metric space with the natural metric being given by the graph distance. Rescale the edge-lengths by $1/\sqrt{m}$:



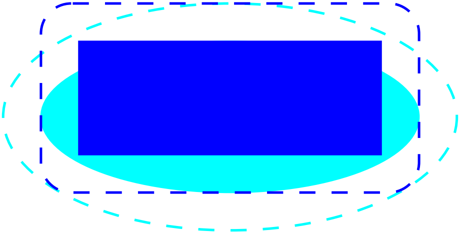
We need a notion of convergence for metric spaces.

Measuring the distance between metric spaces

The Hausdorff distance between two compact subsets K and K' of a metric space (M, δ) is

$$d_H(K, K') = \inf\{\epsilon > 0 : K \subseteq F_\epsilon(K'), K' \subseteq F_\epsilon(K)\},$$

where $F_\epsilon(K) := \{x \in M : \delta(x, K) \leq \epsilon\}$ is the ϵ-fattening of K.



Measuring the distance between metric spaces

To measure the distance between two compact metric spaces (X, d) and (X', d'), the idea is to embed them (isometrically) into a single larger metric space and then compare them using the Hausdorff distance.

Measuring the distance between metric spaces

To measure the distance between two compact metric spaces (X, d) and (X', d'), the idea is to embed them (isometrically) into a single larger metric space and then compare them using the Hausdorff distance.

So define the Gromov-Hausdorff distance

$$d_{GH}(X, X') = \inf\{d_H(\phi(X), \phi'(X'))\},$$

where the infimum is taken over all choices of metric space (M, δ) and all isometric embeddings $\phi : X \to M$, $\phi' : X' \to M$.

Scaling limit

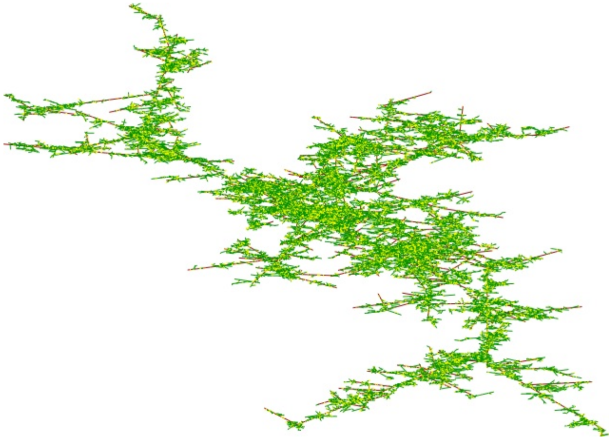
Theorem. (Aldous (1993), Le Gall (2005)) As $m \to \infty$,

$$\frac{1}{\sqrt{m}} T_m \xrightarrow{d} \mathcal{T},$$

where the convergence is in the Gromov-Hausdorff distance.

The limit $\mathcal{T}$ is called the Brownian continuum random tree.

The Brownian continuum random tree



[Picture by Grégory Miermont]

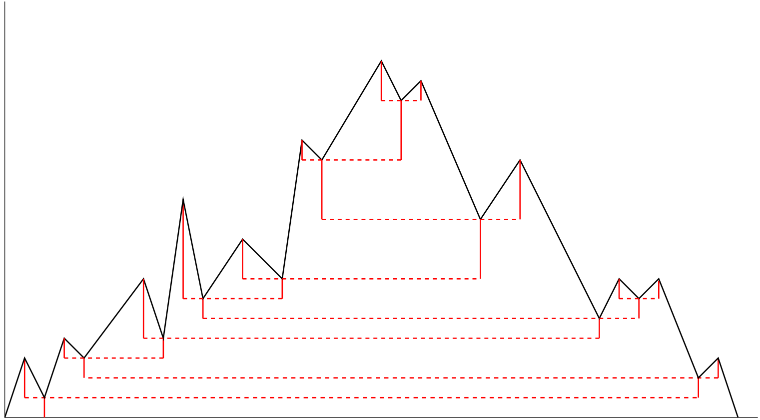
Trees from excursions

Let $h : [0, 1] \to \mathbb{R}^+$ be an excursion, that is a continuous function such that $h(0) = h(1) = 0$ and $h(x) > 0$ for $x \in (0, 1)$.

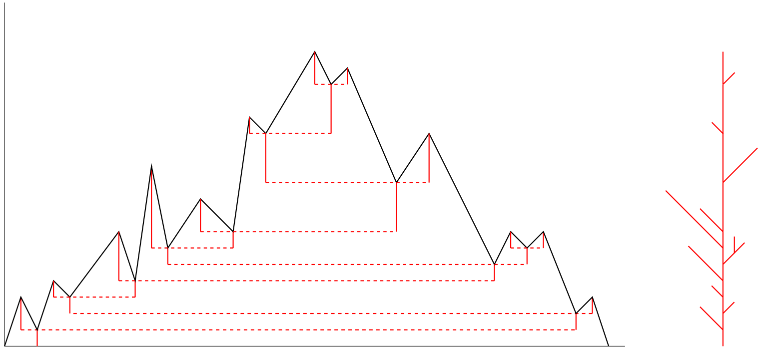
Trees from excursions

Define a distance d on $[0,1]$ via

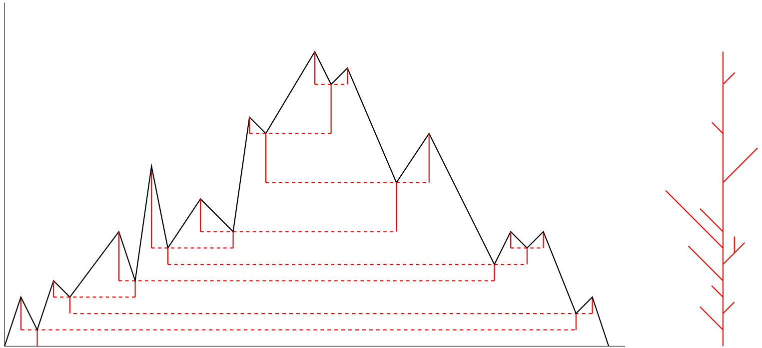
$$d(x,y) = h(x) + h(y) - 2 \inf_{x \wedge y \leq z \leq x \vee y} h(z).$$



Trees from excursions

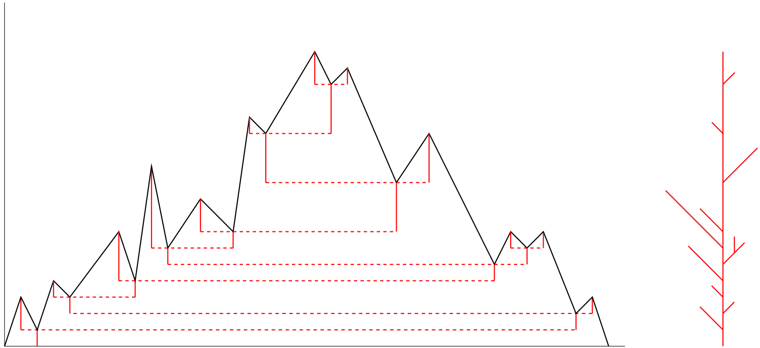


Trees from excursions



Define an equivalence relation $\sim$ by $x \sim y$ if $d(x, y) = 0$ and take the quotient $\mathcal{T}_h = [0, 1]/\sim$.

Trees from excursions



Define an equivalence relation $\sim$ by $x \sim y$ if $d(x, y) = 0$ and take the quotient $\mathcal{T}_h = [0, 1]/\sim$.

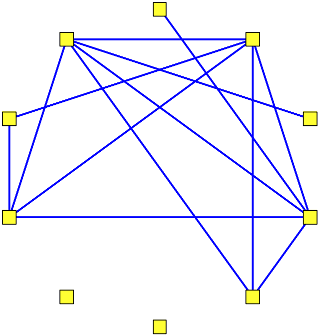
The Brownian continuum random tree is $\mathcal{T}_h$ with $h(x) = 2e(x)$ and $(e(x), 0 \leq x \leq 1)$ a standard Brownian excursion.

Part II : Graphs

The Erdős-Rényi random graph

Take n vertices labelled by $[n] := \{1, 2, \ldots, n\}$ and put an edge between any pair independently with probability p. Call the resulting model $G(n, p)$.

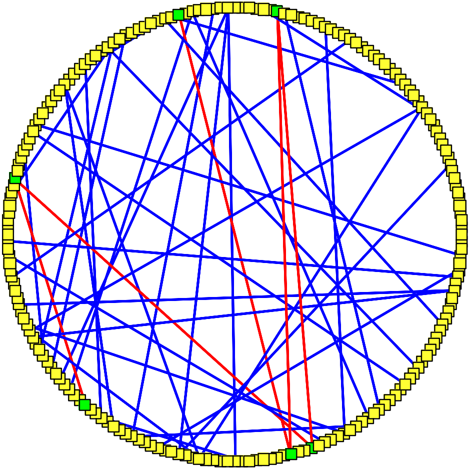
Example: $n = 10$, $p = 0.4$ (vertex labels omitted).



The phase transition

Let $p = c/n$ and consider the largest component (vertices in green, edges in red).

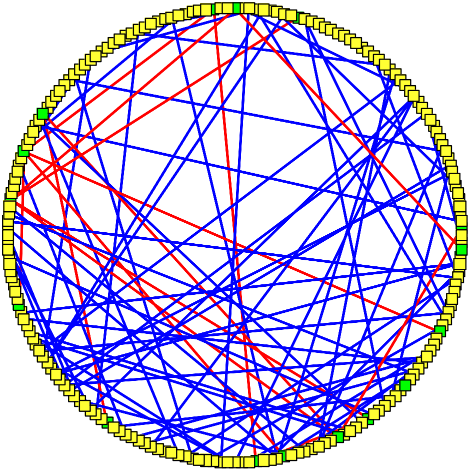
$n = 200$, $c = 0.4$



The phase transition

Let $p = c/n$ and consider the largest component (vertices in green, edges in red).

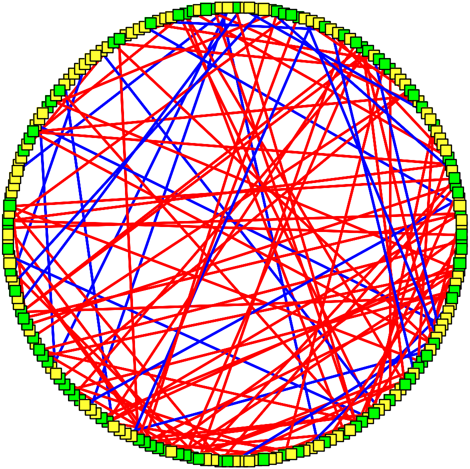
$n = 200$, $c = 0.8$



The phase transition

Let $p = c/n$ and consider the largest component (vertices in green, edges in red).

$n = 200$, $c = 1.2$



The phase transition (Erdős and Rényi (1960))

Consider $p = c/n$.

- For $c < 1$, the largest connected component has size $O(\log n)$;
- for $c > 1$, the largest connected component has size $\Theta(n)$ (and the others are all $O(\log n)$).

The critical random graph

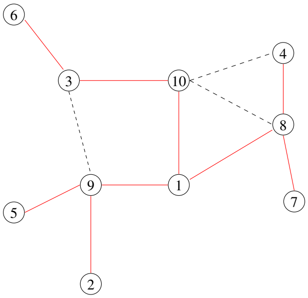
The critical window: $p = \frac{1}{n} + \frac{\lambda}{n^{4/3}}$, where $\lambda \in \mathbb{R}$. For such p, the largest components have size $\Theta(n^{2/3})$.

The critical random graph

The critical window: $p = \frac{1}{n} + \frac{\lambda}{n^{4/3}}$, where $\lambda \in \mathbb{R}$. For such p, the largest components have size $\Theta(n^{2/3})$.

We will also be interested in the surplus of a component, the number of edges more than a tree that it has.

A component with surplus 3:



Convergence of the sizes and surpluses

Fix λ and let $C_1^n, C_2^n, \ldots$ be the sequence of component sizes in decreasing order, and let $S_1^n, S_2^n, \ldots$ be their surpluses.

Write $\mathbf{C}^n = (C_1^n, C_2^n, \ldots)$ and $\mathbf{S}^n = (S_1^n, S_2^n, \ldots)$.

Convergence of the sizes and surpluses

Fix λ and let $C_1^n, C_2^n, \ldots$ be the sequence of component sizes in decreasing order, and let $S_1^n, S_2^n, \ldots$ be their surpluses.

Write $\mathbf{C}^n = (C_1^n, C_2^n, \ldots)$ and $\mathbf{S}^n = (S_1^n, S_2^n, \ldots)$.

Theorem. (Aldous (1997)) As $n \to \infty$,

$$(n^{-2/3}\mathbf{C}^n, \mathbf{S}^n) \xrightarrow{d} (\mathbf{C}, \mathbf{S}).$$

Convergence of the sizes and surpluses

Fix λ and let $C_1^n, C_2^n, \ldots$ be the sequence of component sizes in decreasing order, and let $S_1^n, S_2^n, \ldots$ be their surpluses.

Write $\mathbf{C}^n = (C_1^n, C_2^n, \ldots)$ and $\mathbf{S}^n = (S_1^n, S_2^n, \ldots)$.

Theorem. (Aldous (1997)) As $n \to \infty$,

$$(n^{-2/3}\mathbf{C}^n, \mathbf{S}^n) \xrightarrow{d} (\mathbf{C}, \mathbf{S}).$$

Here, convergence in the first co-ordinate takes place in

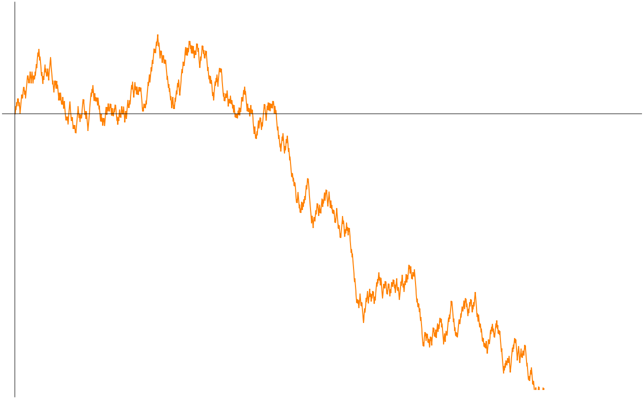
$$\ell^2_{\searrow} := \left\{ \mathbf{x} = (x_1, x_2, \ldots) : x_1 \geq x_2 \geq \ldots \geq 0, \sum_{i=1}^{\infty} x_i^2 < \infty \right\}.$$

Limiting sizes and surpluses

Let $W^{\lambda}(t) = W(t) + \lambda t - \frac{t^2}{2}$, $t \geq 0$, where $(W(t), t \geq 0)$ is a standard Brownian motion.

Limiting sizes and surpluses

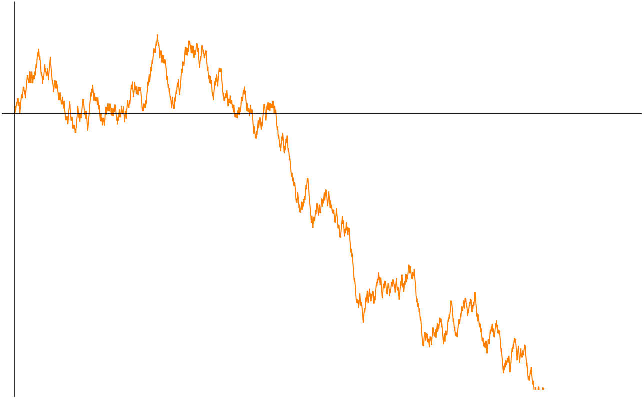
Let $W^{\lambda}(t) = W(t) + \lambda t - \frac{t^2}{2}$, $t \geq 0$, where $(W(t), t \geq 0)$ is a standard Brownian motion.



Limiting sizes and surpluses

Let $W^\lambda(t) = W(t) + \lambda t - \frac{t^2}{2}$, $t \geq 0$, where $(W(t), t \geq 0)$ is a standard Brownian motion.

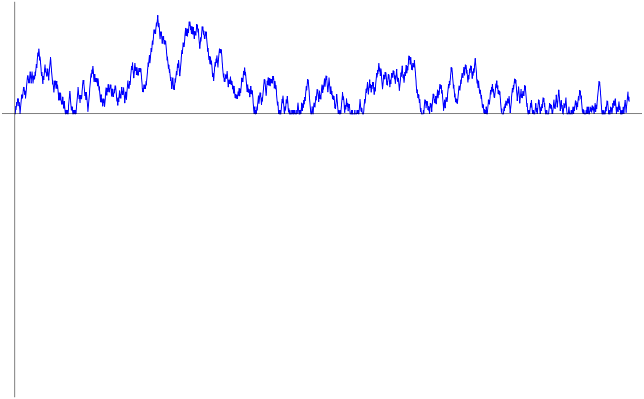
Let $B^\lambda(t) = W^\lambda(t) - \min_{0 \leq s \leq t} W^\lambda(s)$ be the process reflected at its minimum.

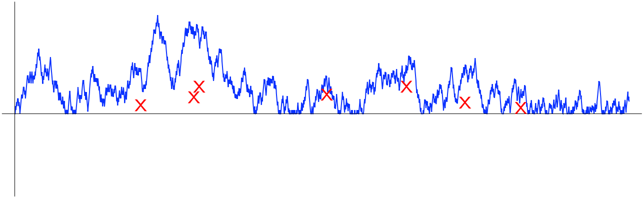


Limiting sizes and surpluses

Let $W^{\lambda}(t) = W(t) + \lambda t - \frac{t^2}{2}$, $t \geq 0$, where $(W(t), t \geq 0)$ is a standard Brownian motion.

Let $B^{\lambda}(t) = W^{\lambda}(t) - \min_{0 \leq s \leq t} W^{\lambda}(s)$ be the process reflected at its minimum.





Decorate the picture with the points of a rate one Poisson process which fall above the x-axis and below the graph.

C is the sequence of excursion-lengths of this process, in decreasing order.

S is the sequence of numbers of points falling in the corresponding excursions.

Question

What do the limiting components look like?

Question

What do the limiting components look like?

The vertex-labels are irrelevant: we are really interested in what distances look like in the limit. So we will give a metric space answer.

Our approach

Simple but important fact: a component of $G(n, p)$ conditioned to have m vertices and s surplus edges is a uniform connected graph on those m vertices with $m + s - 1$ edges.

Our approach

Simple but important fact: a component of $G(n, p)$ conditioned to have m vertices and s surplus edges is a uniform connected graph on those m vertices with $m + s - 1$ edges.

Our general approach is to pick out a (well-chosen) spanning tree, and then to put in the surplus edges.

Our approach

Simple but important fact: a component of $G(n, p)$ conditioned to have m vertices and s surplus edges is a uniform connected graph on those m vertices with $m + s - 1$ edges.

Our general approach is to pick out a (well-chosen) spanning tree, and then to put in the surplus edges.

There is one case which we already understand very well: when the surplus of a component is 0 and so we have a uniform random tree.

The limit of the random graph

In the tree case, we rescaled distances by $1/\sqrt{m}$, where m was the number of vertices. This is the correct distance rescaling for all of the big components in the random graph.

The limit of the random graph

In the tree case, we rescaled distances by $1/\sqrt{m}$, where m was the number of vertices. This is the correct distance rescaling for all of the big components in the random graph. Since the big components have sizes of order $n^{2/3}$, we should rescale distances by $n^{-1/3}$.

The limit of the random graph

In the tree case, we rescaled distances by $1/\sqrt{m}$, where m was the number of vertices. This is the correct distance rescaling for all of the big components in the random graph. Since the big components have sizes of order $n^{2/3}$, we should rescale distances by $n^{-1/3}$.

Each excursion of the process $(B^\lambda(t), t \geq 0)$ encodes a continuum random tree, which is a "spanning tree" for a limit component.

The limit of the random graph

In the tree case, we rescaled distances by $1/\sqrt{m}$, where m was the number of vertices. This is the correct distance rescaling for all of the big components in the random graph. Since the big components have sizes of order $n^{2/3}$, we should rescale distances by $n^{-1/3}$.

Each excursion of the process $(B^\lambda(t), t \geq 0)$ encodes a continuum random tree, which is a "spanning tree" for a limit component. These are *not* rescaled Brownian CRT's, but CRT's whose distribution has been "tilted" in a way which we will make precise in a moment.

The limit of the random graph

In the tree case, we rescaled distances by $1/\sqrt{m}$, where m was the number of vertices. This is the correct distance rescaling for all of the big components in the random graph. Since the big components have sizes of order $n^{2/3}$, we should rescale distances by $n^{-1/3}$.

Each excursion of the process $(B^{\lambda}(t), t \geq 0)$ encodes a continuum random tree, which is a "spanning tree" for a limit component. These are *not* rescaled Brownian CRT's, but CRT's whose distribution has been "tilted" in a way which we will make precise in a moment.

In the limit, surplus edges correspond to vertex-identifications (since edge-lengths have shrunk to 0). In each excursion, the points of the Poisson process tell us where these vertex-identifications should occur.

Excursions of the limit process

Consider the process $(B^{\lambda}(t), t \geq 0)$. An excursion $\tilde{e}^{(x)}$ of this process, conditioned to have length x, has a distribution specified by

$$\mathbb{E}\left[f\left(\tilde{e}^{(x)}\right)\right] = \frac{\mathbb{E}\left[f\left(e^{(x)}\right)\exp\left(\int_0^x e^{(x)}(u)du\right)\right]}{\mathbb{E}\left[\exp\left(\int_0^x e^{(x)}(u)du\right)\right]},$$

where f is any suitable test-function and $e^{(x)}$ is a Brownian excursion of length x.

Excursions of the limit process

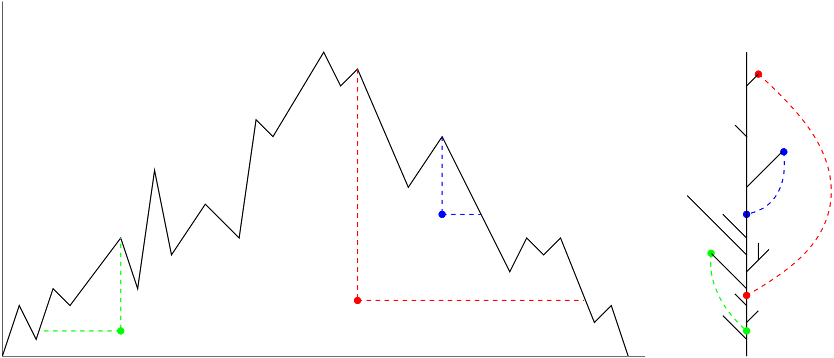
Consider the process $(B^{\lambda}(t), t \geq 0)$. An excursion $\tilde{e}^{(x)}$ of this process, conditioned to have length x, has a distribution specified by

$$\mathbb{E}\left[f\left(\tilde{e}^{(x)}\right)\right] = \frac{\mathbb{E}\left[f\left(e^{(x)}\right)\exp\left(\int_0^x e^{(x)}(u)du\right)\right]}{\mathbb{E}\left[\exp\left(\int_0^x e^{(x)}(u)du\right)\right]},$$

where f is any suitable test-function and $e^{(x)}$ is a Brownian excursion of length x.

We refer to $\tilde{e}^{(x)}$ as a tilted excursion and to the tree $\tilde{\mathcal{T}}$ that it encodes as a tilted tree.

Vertex identifications



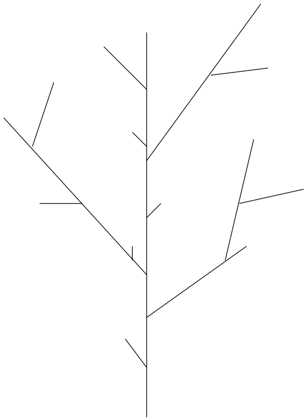
A point at (x, y) identifies the vertex v at height $h(x)$ with the vertex at distance y along the path from the root to v.

A limiting component

Note that it follows from properties of the tilted trees and of the Poisson process that we may equivalently describe the limit of a component on $\sim xn^{2/3}$ vertices as follows.

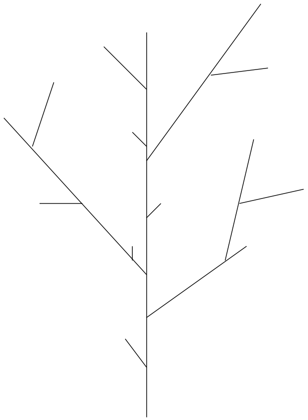
A limiting component

Sample a tilted excursion $\tilde{e}^{(x)}$ of length x and use it to create a CRT $\tilde{\mathcal{T}}$.



A limiting component

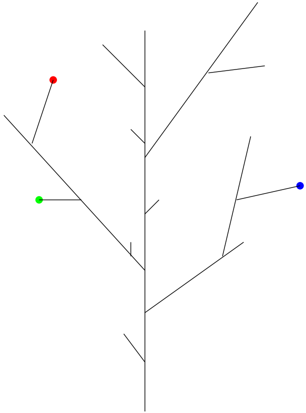
Sample a tilted excursion $\tilde{e}^{(x)}$ of length x and use it to create a CRT $\tilde{\mathcal{T}}$.



Conditional on $\tilde{e}^{(x)}$, sample a random variable P with Poisson $\left(\int_0^x \tilde{e}^{(x)}(u)du\right)$ distribution.

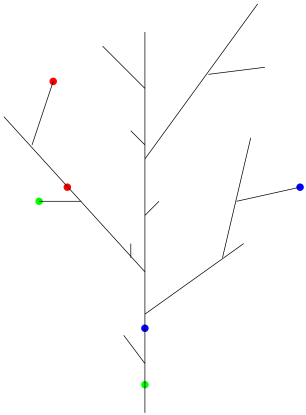
A limiting component

Conditional on $P = s$, pick s vertices of the tree $\tilde{\mathcal{T}}$ independently with density proportional to their height. (These will almost surely be leaves.)



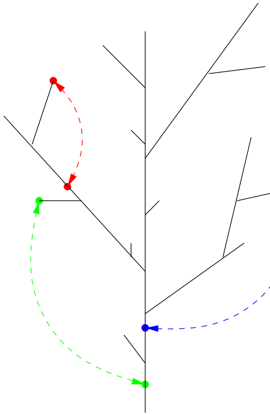
A limiting component

For each of the selected leaves, pick a uniform point on the path from the leaf to the root.



A limiting component

Identify each of the selected leaves with its chosen point.



Convergence result

Let $\mathcal{C}_1^n, \mathcal{C}_2^n, \ldots$ be the sequence of components of $G(n,p)$ in decreasing order of size, considered as metric spaces with the graph distance.

Theorem. As $n \to \infty$,

$$n^{-1/3}(\mathcal{C}_1^n, \mathcal{C}_2^n, \ldots) \xrightarrow{d} (\mathcal{C}_1, \mathcal{C}_2, \ldots),$$

where $\mathcal{C}_1, \mathcal{C}_2, \ldots$ is the sequence of metric spaces corresponding to the excursions of Aldous' marked limit process in decreasing order of length.

Here, convergence is with respect to the metric

$$d(\mathcal{A}, \mathcal{B}) := \left(\sum_{i=1}^{\infty} d_{GH}(\mathcal{A}_i, \mathcal{B}_i)^4 \right)^{1/4}.$$

Diameter

Let D_n be the diameter of $G(n,p)$ for p in the critical window, that is the largest distance between a pair of vertices lying in the same component of the graph.

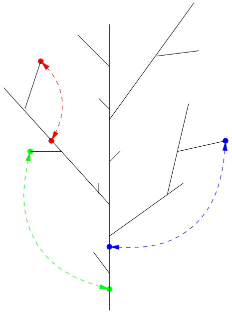
Nachmias and Peres (2008) showed that $D_n = \Theta(n^{1/3})$. (Also follows from results of Addario-Berry, Broutin and Reed.)

Our convergence result allows us to prove that

$$n^{-1/3} D_n \xrightarrow{d} D$$

as $n \to \infty$, where D is an absolutely continuous random variable with finite mean.

Papers



The continuum limit of critical random graphs

L. Addario-Berry, N. Broutin and C. Goldschmidt

arXiv:0903.4730 [math.PR].

Critical random graphs: limiting constructions and distributional properties

L. Addario-Berry, N. Broutin and C. Goldschmidt

arXiv:0908.3629 [math.PR].