Warwick, 21st April 2009

Lecture 2: The continuum random tree (continued)

and then

The scaling limit of critical random graphs

Christina Goldschmidt

Recap from yesterday

We considered the height process H and depth-first walk X associated with a Galton-Watson forest (i.e. a sequence of i.i.d. Galton-Watson trees) and showed that

$$\frac{1}{\sqrt{n}}(X(\lfloor nt \rfloor), t \geq 0) \xrightarrow{d} \sigma(B(t), t \geq 0)$$
$$\frac{1}{\sqrt{n}}(H(\lfloor nt \rfloor), t \geq 0) \xrightarrow{d} \frac{2}{\sigma}(\beta(t), t \geq 0).$$

Recap from yesterday

We also saw that the depth first walk X^n, height process H^n and contour process C^n of a single Galton-Watson tree conditioned to have total progeny n all converge to multiples of the same standard Brownian excursion:

Theorem. (Marckert and Mokkadem (2003)) As $n \to \infty$,

$$\begin{aligned}&(n^{-1/2}X^n(\lfloor n\cdot\rfloor), n^{-1/2}H^n(\lfloor n\cdot\rfloor), n^{-1/2}C^n(\lfloor 2n\cdot\rfloor))\\ &\xrightarrow{d} \left(\sigma e, \frac{2}{\sigma}e, \frac{2}{\sigma}e\right).\end{aligned}$$

We will now discuss the limiting tree.

Real trees

Real trees

Definition. A compact metric space $(\mathcal{T}, d)$ is a real tree if for all $x, y \in \mathcal{T}$,

- there exists a unique shortest path $[[x, y]]$ from x to y (of length $d(x, y)$);
- the only non-self-intersecting path from x to y is $[[x, y]]$.

Real trees

Definition. A compact metric space $(\mathcal{T}, d)$ is a real tree if for all $x, y \in \mathcal{T}$,

- there exists a unique shortest path $[[x, y]]$ from x to y (of length $d(x, y)$);
- the only non-self-intersecting path from x to y is $[[x, y]]$.

An element $v \in \mathcal{T}$ is called a vertex. A rooted real tree has a distinguished vertex ρ called the root. The height of a vertex v is its distance $d(\rho, v)$ from the root. A leaf is a vertex v such that $v \notin [[\rho, w]]$ for any $w \neq v$.

Coding real trees

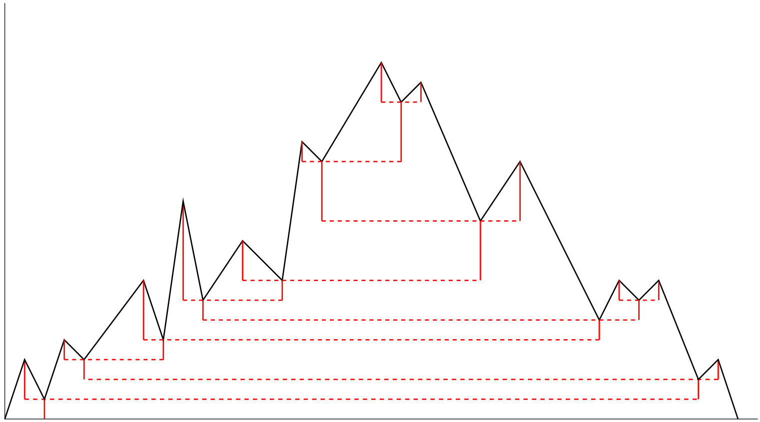
Suppose that $h : [0, \infty) \to [0, \infty)$ is a continuous function of compact support such that $h(0) = 0$. h will play the role of the height/contour process for a real tree.

Coding real trees

Coding real trees

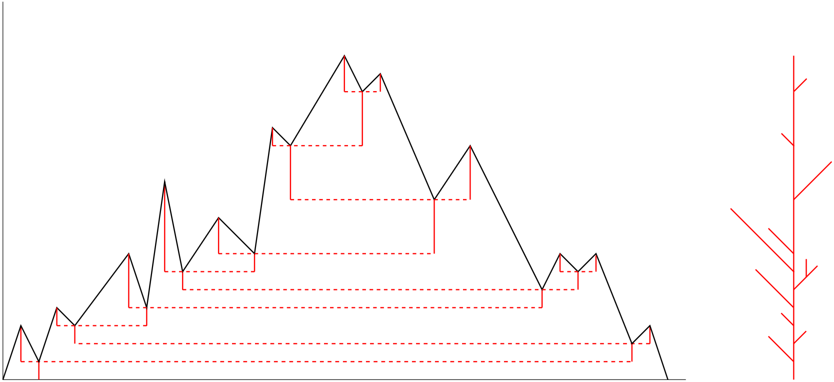
Use h to define a distance:

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



Coding real trees

Let $y \sim y'$ if $d_h(y, y') = 0$ and take the quotient $\mathcal{T}_h = [0, \infty)/\sim$.



Coding real trees

Then $(\mathcal{T}_h, d_h)$ is a real tree. We will always take the equivalence class of 0 to be the root, ρ.

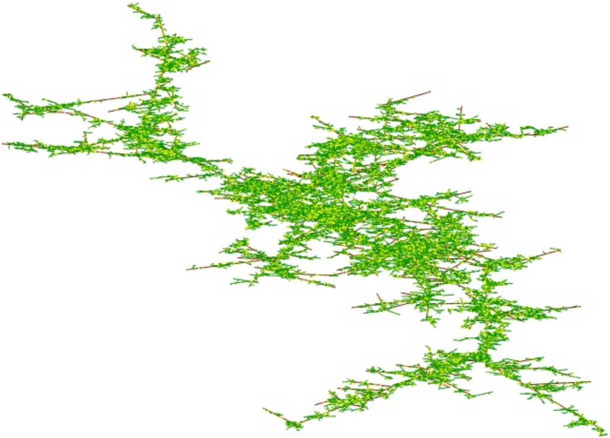
Coding real trees

Then $(\mathcal{T}_h, d_h)$ is a real tree. We will always take the equivalence class of 0 to be the root, ρ.

Aldous' Brownian continuum random tree (CRT) is $\mathcal{T}_{2e}$, where e is a standard Brownian excursion.

This will be the limit for our conditioned Galton-Watson trees.

The Brownian continuum random tree $\mathcal{T}_{2e}$



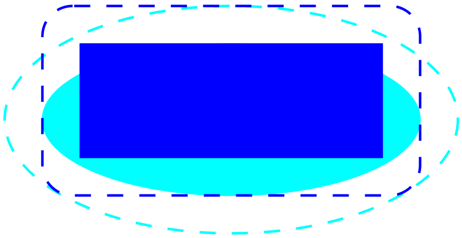
[Picture by Grégory Miermont]

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 $\phi : X \to M$, $\phi' : X' \to M$.

Measuring the distance between metric spaces

If the metric spaces are rooted, at ρ and ρ' respectively, we take

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

Measuring the distance between metric spaces

If the metric spaces are rooted, at ρ and ρ' respectively, we take

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

Fortunately, we do not have to seek an optimal embedding! In fact, it turns out that if h and h' are two coding functions then

$$d_{GH}(\mathcal{T}_h, \mathcal{T}_{h'}) \leq 2\|h - h'\|.$$

Convergence to the CRT

Let T_n be our Galton-Watson tree conditioned to have size n, thought of as the real tree coded by its contour function, C^n.

Convergence to the CRT

Let T_n be our Galton-Watson tree conditioned to have size n, thought of as the real tree coded by its contour function, C^n.

Since

$$(n^{-1/2} C^n(\lfloor 2nt \rfloor), 0 \leq t \leq 1) \xrightarrow{d} \frac{2}{\sigma}(e(t), 0 \leq t \leq 1),$$

we can deduce that

$$\frac{\sigma}{\sqrt{n}} T_n \xrightarrow{d} \mathcal{T}_{2e},$$

where convergence is in the Gromov-Hausdorff sense.

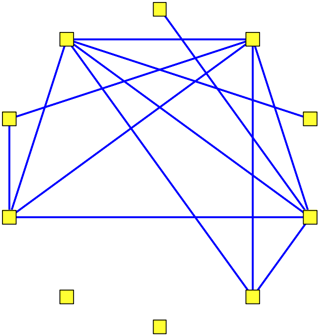
The scaling limit of critical random graphs

Joint work with Louigi Addario-Berry (Université de Montréal) and Nicolas Broutin (INRIA Rocquencourt)

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 $G(n, p)$.

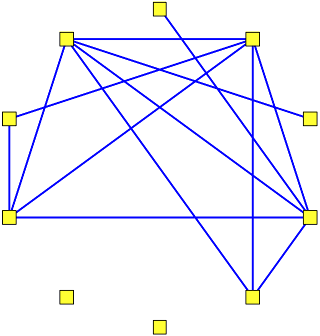
Example: $n = 10$, $p = 0.4$ (vertex labels omitted).



Connected components

We're going to be interested in connected components of these graphs.

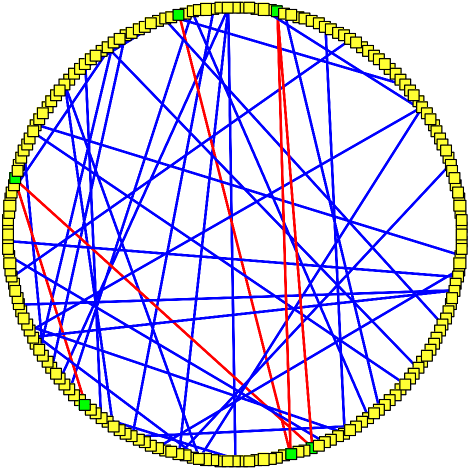
Below, there are three connected components.



The phase transition

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

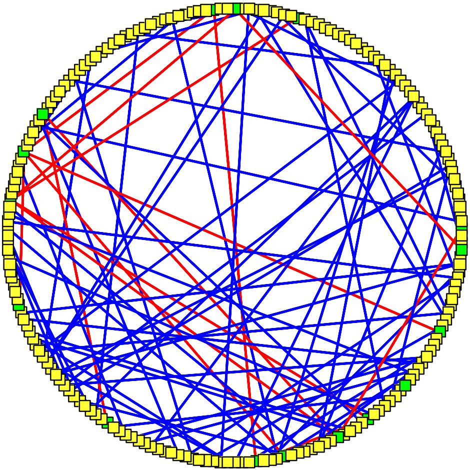
$n = 200$, $c = 0.4$



The phase transition

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

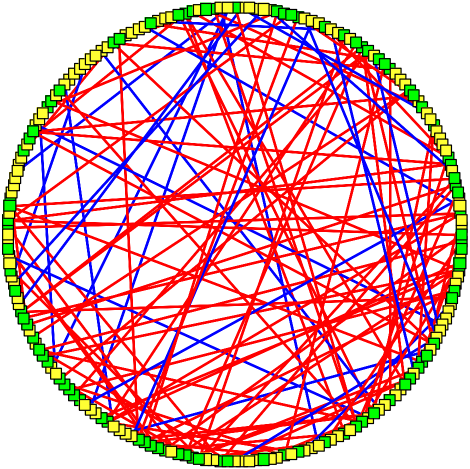
$n = 200,\ c = 0.8$



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))

By the size of a component, we mean its number of vertices.

The phase transition (Erdős and Rényi (1960))

By the size of a component, we mean its number of vertices.

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 phase transition (Erdős and Rényi (1960))

By the size of a component, we mean its number of vertices.

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 $O(\log n)$).

If $c = 1$, the largest component has size $\Theta(n^{2/3})$ and, indeed, there is a whole sequence of components of this order.

The critical random graph

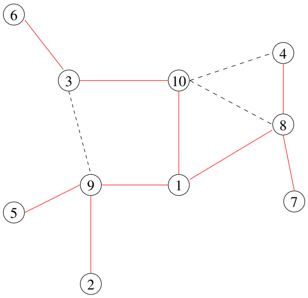
The critical window: $p = \frac{1}{n} + \frac{\lambda}{n^{4/3}}$, where $\lambda \in \mathbb{R}$. For such p, the largest components have size $\Theta(n^{2/3})$.

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

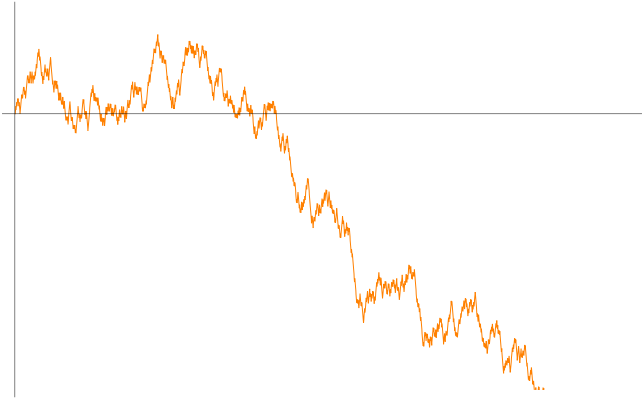
$$\ell^2_{\searrow} := \left\{ \mathbf{x} = (x_1, x_2, \ldots) : x_1 \geq x_2 \geq \ldots \geq 0, \sum_{i=1}^{\infty} x_i^2 < \infty \right\}.$$

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

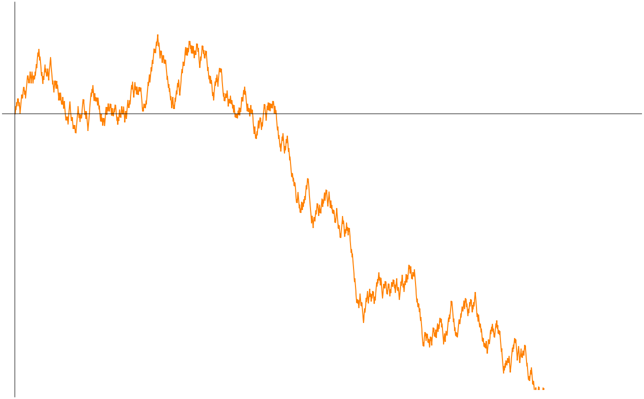
Let $W^{\lambda}(t) = W(t) + \lambda t - \frac{t^2}{2}$, $t \geq 0$, where $(W(t), t \geq 0)$ is a standard Brownian motion.



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.

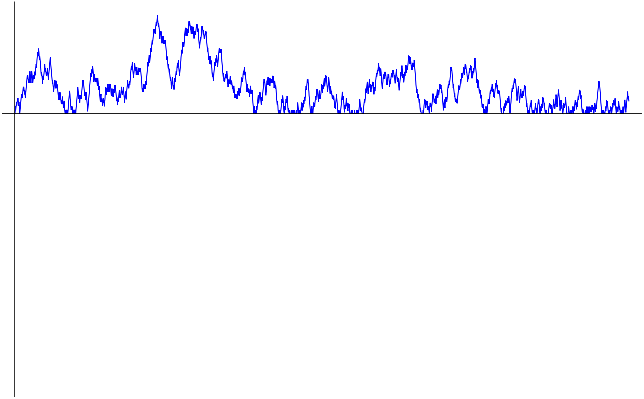
Let $B^{\lambda}(t) = W^{\lambda}(t) - \min_{0 \leq s \leq t} W^{\lambda}(s)$ be the process reflected at its minimum.

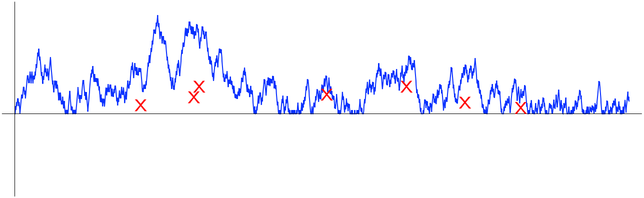


Limiting sizes and surpluses

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, and convergence will be in the Gromov-Hausdorff distance.

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

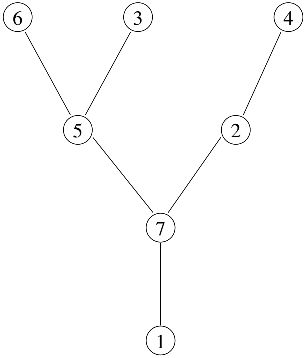
Simple but important fact: a component of $G(n,p)$ conditioned to have m vertices and s surplus edges is a uniform connected graph on those m vertices with $m+s-1$ edges.

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 is already well-understood: when the surplus of a component is 0 and so we have a uniform (unordered) random tree. It turns out that this can be made to fit into the ordered framework we have so far considered.

Labelled (unordered, unrooted) trees

We now want to be thinking about labelled trees.



Uniform random trees

Let $\mathbb{T}_{[m]}$ be the set of labelled unordered trees on labels $[m] := \{1, 2, \ldots, m\}$. Cayley's formula tells us that $|\mathbb{T}_{[m]}| = m^{m-2}$. Write T_m for a tree chosen uniformly from $\mathbb{T}_{[m]}$; this is what we mean by a uniform random tree.

Uniform random trees

Let $\mathbb{T}_{[m]}$ be the set of labelled unordered trees on labels $[m] := \{1, 2, \ldots, m\}$. Cayley's formula tells us that $|\mathbb{T}_{[m]}| = m^{m-2}$. Write T_m for a tree chosen uniformly from $\mathbb{T}_{[m]}$; this is what we mean by a uniform random tree.

Proposition. Let P be a (rooted, ordered) Galton-Watson tree, with Poisson(1) offspring distribution and total progeny N.

Uniform random trees

Let $\mathbb{T}_{[m]}$ be the set of labelled unordered trees on labels $[m] := \{1, 2, \ldots, m\}$. Cayley's formula tells us that $|\mathbb{T}_{[m]}| = m^{m-2}$. Write T_m for a tree chosen uniformly from $\mathbb{T}_{[m]}$; this is what we mean by a uniform random tree.

Proposition. Let P be a (rooted, ordered) Galton-Watson tree, with Poisson(1) offspring distribution and total progeny N. Assign the vertices labels uniformly at random from $\{1, 2, \ldots, N\}$ and then forget the ordering and the root. Let $\tilde{P}$ be the labelled tree obtained.

Uniform random trees

Let $\mathbb{T}_{[m]}$ be the set of labelled unordered trees on labels $[m] := \{1, 2, \ldots, m\}$. Cayley's formula tells us that $|\mathbb{T}_{[m]}| = m^{m-2}$. Write T_m for a tree chosen uniformly from $\mathbb{T}_{[m]}$; this is what we mean by a uniform random tree.

Proposition. Let P be a (rooted, ordered) Galton-Watson tree, with Poisson(1) offspring distribution and total progeny N. Assign the vertices labels uniformly at random from $\{1, 2, \ldots, N\}$ and then forget the ordering and the root. Let $\tilde{P}$ be the labelled tree obtained. Then, conditional on $N = m$, $\tilde{P}$ has the same distribution as T_m.

Uniform random trees

Now, using the ordinary graph distance, think of T_m as a metric space (rooted at 1).

Uniform random trees

Now, using the ordinary graph distance, think of T_m as a metric space (rooted at 1).

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

(Note that the Poisson(1) distribution has variance $\sigma^2 = 1$.)

Overview: the limit of the random graph

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

Overview: the limit of the random graph

In the tree case, we rescale distances by $1/\sqrt{m}$, where m is 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 $n^{-1/3}$.

Overview: the limit of the random graph

In the tree case, we rescale distances by $1/\sqrt{m}$, where m is 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 $(B^\lambda(t), t \geq 0)$ of length x corresponds to the limit of a component on $\sim xn^{2/3}$ vertices. Such an excursion codes a continuum random tree, which is a "spanning tree" for that limit component.

Overview: the limit of the random graph

In the tree case, we rescale distances by $1/\sqrt{m}$, where m is 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 $(B^\lambda(t), t \geq 0)$ of length x corresponds to the limit of a component on $\sim xn^{2/3}$ vertices. Such an excursion codes a continuum random tree, which is a "spanning tree" for that limit component. These CRT's 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.

Overview: the limit of the random graph

In the tree case, we rescale distances by $1/\sqrt{m}$, where m is 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 $(B^\lambda(t), t \geq 0)$ of length x corresponds to the limit of a component on $\sim xn^{2/3}$ vertices. Such an excursion codes a continuum random tree, which is a "spanning tree" for that limit component. These CRT's 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

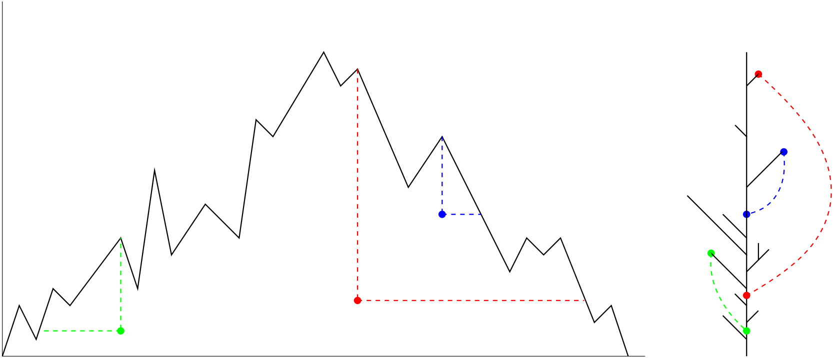
Consider the process $(B^{\lambda}(t), t \geq 0)$. An excursion $\tilde{e}^{(x)}$ of this process, conditioned to have length x, has a distribution specified by

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

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



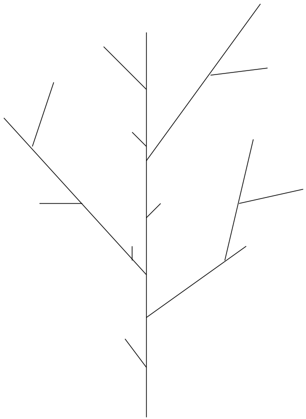
A point at (x, y) identifies the vertex v at height $h(x)$ with the vertex at distance y along the path from the root to v.

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.

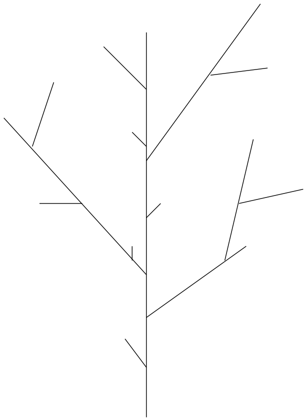
A limiting component

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



A limiting component

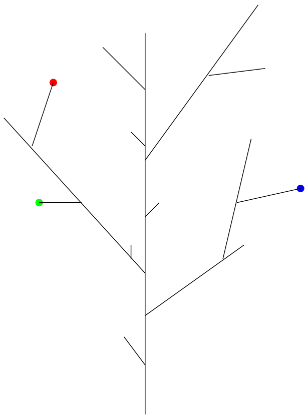
Sample a tilted excursion $\tilde{e}^{(x)}$ of length x and use it to create a CRT $\tilde{\mathcal{T}}$.



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

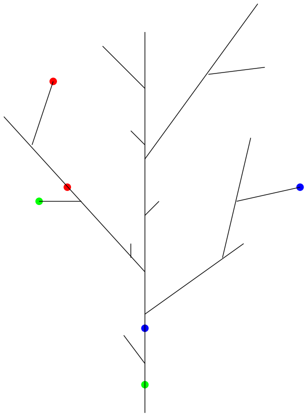
A limiting component

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.)



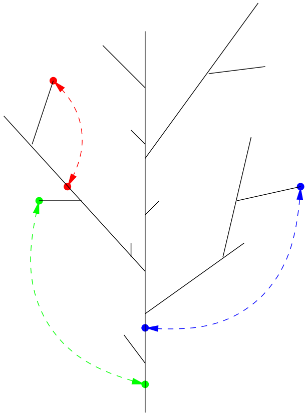
A limiting component

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 $D_n = \Theta(n^{1/3})$.

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.

Idea of proof

The key idea turns out to be study a component of $G(n, p)$ conditioned on its size but *not* on its surplus.

Idea of proof

The key idea turns out to be study a component of $G(n, p)$ conditioned on its size but *not* on its surplus.

We will begin by taking an arbitrary connected graph G and picking out a particular spanning tree via a depth-first exploration.

We explore the graph step-by-step. At each step, the vertices may be in one of three states: current, alive or dead.

Idea of proof

The key idea turns out to be study a component of $G(n, p)$ conditioned on its size but *not* on its surplus.

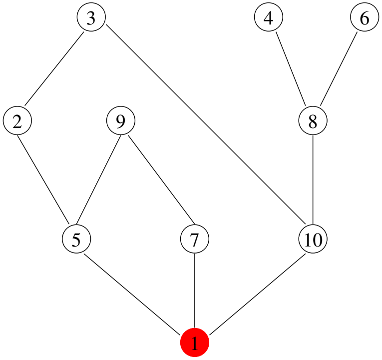
We will begin by taking an arbitrary connected graph G and picking out a particular spanning tree via a depth-first exploration.

We explore the graph step-by-step. At each step, the vertices may be in one of three states: current, alive or dead.

Let $X(k)$ be the number of vertices alive at step k. It turns out that this is the same as the depth-first walk of an underlying tree.

Depth-first exploration

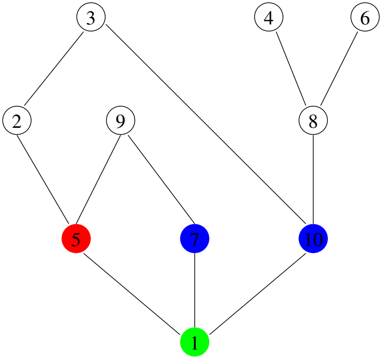
Step 0: initialization



Current: 1 Alive: none Dead: none $X(0) = 0$.

Depth-first exploration

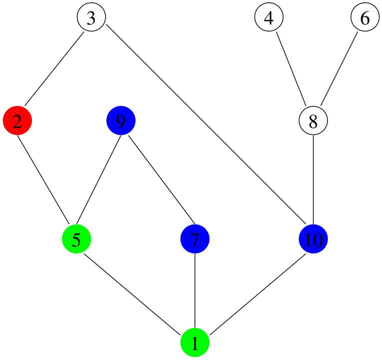
Step 1



Current: 5   Alive: 7, 10   Dead: 1 $X(1) = 2.$

Depth-first exploration

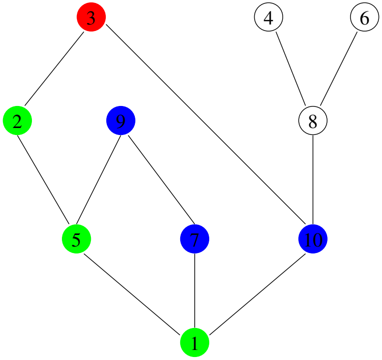
Step 2



Current: 2   Alive: 9, 7, 10   Dead: 1, 5 $X(2) = 3$.

Depth-first exploration

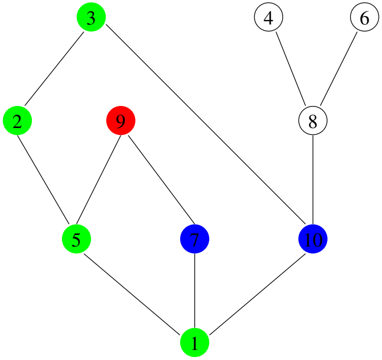
Step 3



Current: 3   Alive: 9, 7, 10   Dead: 1, 5, 2 $X(3) = 3$.

Depth-first exploration

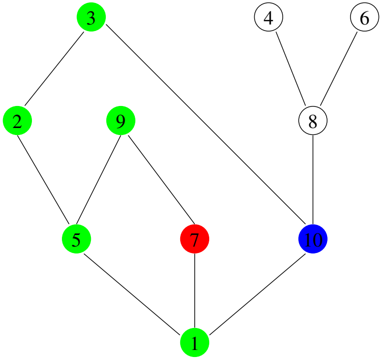
Step 4



Current: 9 Alive: 7, 10 Dead: 1, 5, 2, 3 $X(4) = 2$.

Depth-first exploration

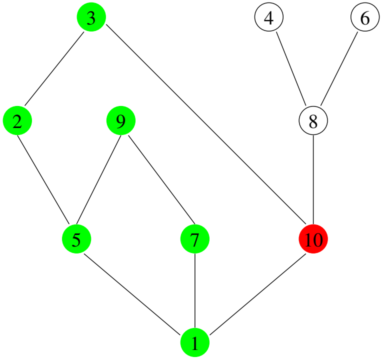
Step 5



Current: 7 Alive: 10 Dead: 1, 5, 2, 3, 9 $X(5) = 1.$

Depth-first exploration

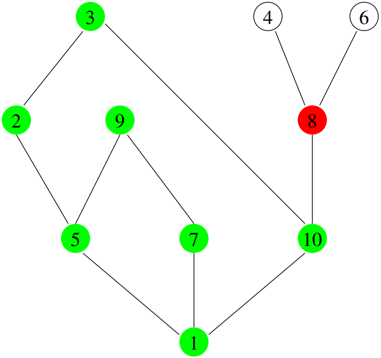
Step 6



Current: 10 Alive: none Dead: 1, 5, 2, 3, 9, 7 $X(6) = 0$.

Depth-first exploration

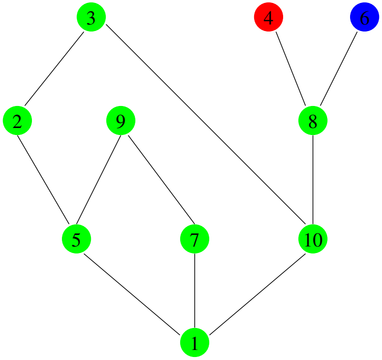
Step 7



Current: 8   Alive: none   Dead: 1, 5, 2, 3, 9, 7, 10 $X(7) = 0.$

Depth-first exploration

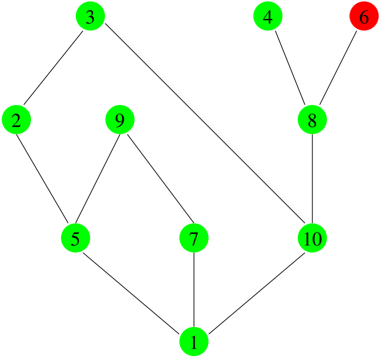
Step 8



Current: 4   Alive: 6   Dead: 1, 5, 2, 3, 9, 7, 10, 8 $X(8) = 1.$

Depth-first exploration

Step 9

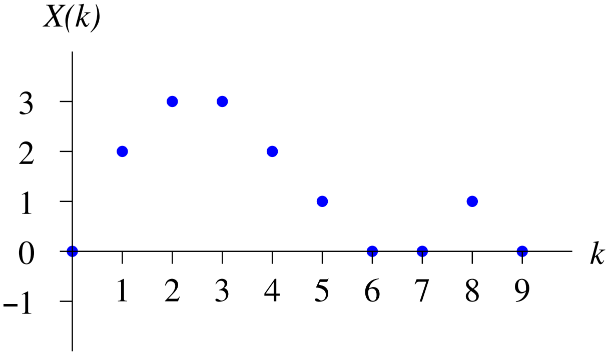


Current: 6 Alive: none Dead: 1, 5, 2, 3, 9, 7, 10, 8, 4

$X(9) = 0$.

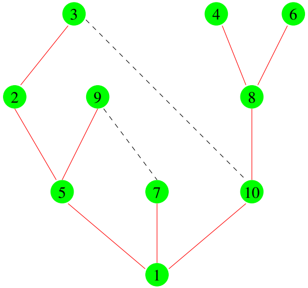
Depth-first walk

$X(k)$ = the number of vertices alive at the kth step of the depth-first exploration.



Depth-first tree

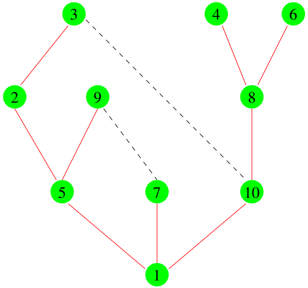
We essentially explored this tree; the dashed edges made no difference to the depth-first exploration.



Call it the depth-first tree associated with the graph G, and write $T(G)$.

Depth-first tree

We essentially explored this tree; the dashed edges made no difference to the depth-first exploration.



Call it the depth-first tree associated with the graph G, and write $T(G)$.

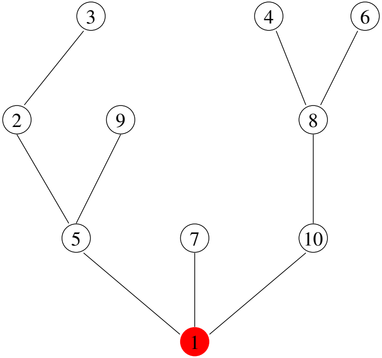
$(X(k), 0 \leq k \leq n-1)$ is the depth-first walk of this tree.

Permitted edges

For a given tree T, which connected graphs have depth-first tree T? In other words, where can we put surplus edges so that they don't change T?

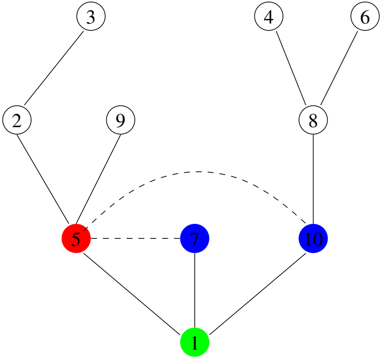
Call such edges permitted.

Depth-first walk and permitted edges



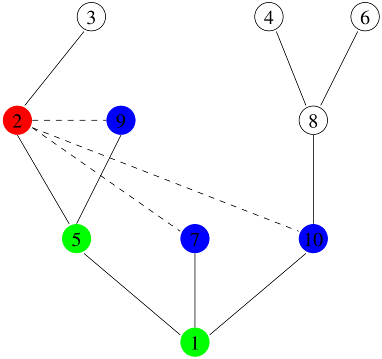
Step 0: $X(0) = 0$.

Depth-first walk and permitted edges



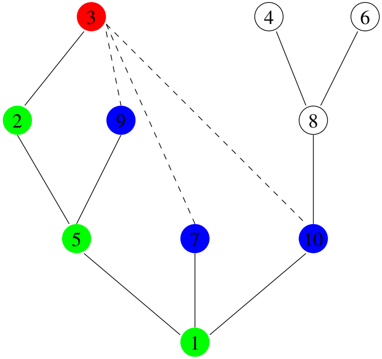
Step 1: $X(1) = 2$.

Depth-first walk and permitted edges



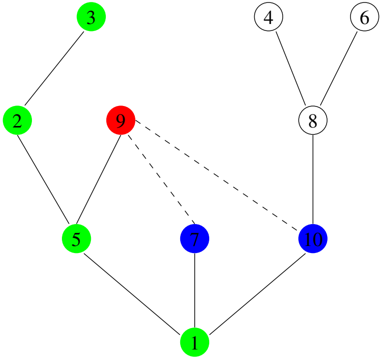
Step 2: $X(2) = 3$.

Depth-first walk and permitted edges



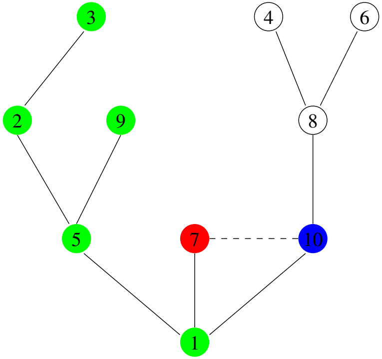
Step 3: $X(3) = 3$.

Depth-first walk and permitted edges



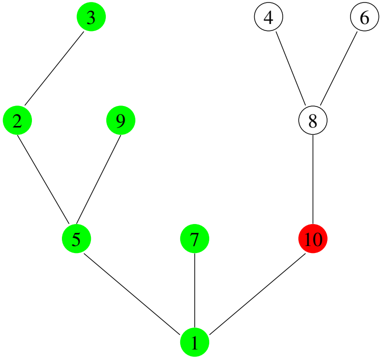
Step 4: $X(4) = 2$.

Depth-first walk and permitted edges



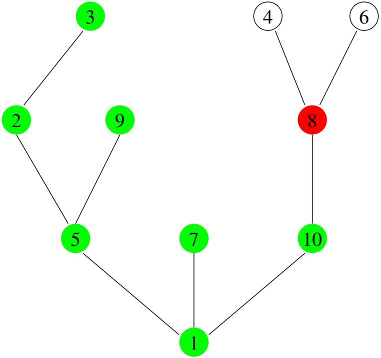
Step 5: $X(5) = 1$.

Depth-first walk and permitted edges



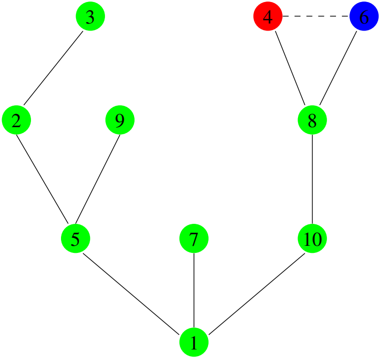
Step 6: $X(6) = 0$.

Depth-first walk and permitted edges



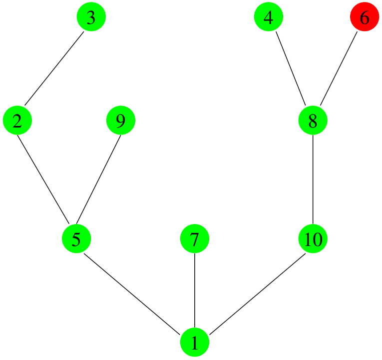
Step 7: $X(7) = 0$.

Depth-first walk and permitted edges



Step 8: $X(8) = 1$.

Depth-first walk and permitted edges



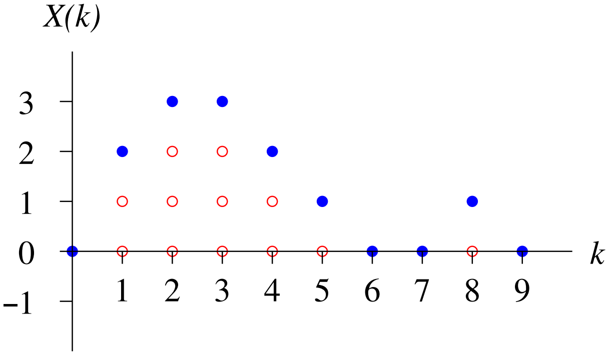
Step 10: $X(9) = 0$.

Area

At step $k \geq 0$ there are $X(k)$ permitted edges. So the total number is

$$a(T) = \sum_{k=0}^{m-1} X(k).$$

We call this the area of T.



Classifying graphs by depth-first tree

Let $\mathbb{G}_T$ be the set of graphs G such that $T(G) = T$. It follows that $|\mathbb{G}_T| = 2^{a(T)}$.

Classifying graphs by depth-first tree

Let $\mathbb{G}_T$ be the set of graphs G such that $T(G) = T$. It follows that $|\mathbb{G}_T| = 2^{a(T)}$.

Let $\mathbb{T}_{[m]}$ be the set of trees with label-set $[m] = \{1, 2, \ldots, m\}$. Then

$$\{\mathbb{G}_T : T \in \mathbb{T}_{[m]}\}$$

is a partition of the set of connected graphs on $[m]$.

Recipe for creating a connected graph on $[m]$

Create a connected graph $\tilde{G}_m^p$ as follows.

Recipe for creating a connected graph on $[m]$

Create a connected graph $\tilde{G}_m^p$ as follows.

- Pick a random labelled tree $\tilde{T}_m^p$ such that

$$\mathbb{P}\left(\tilde{T}_m^p = T\right) \propto (1-p)^{-a(T)}, \quad T \in \mathbb{T}_{[m]}.$$

Recipe for creating a connected graph on $[m]$

Create a connected graph $\tilde{G}_m^p$ as follows.

- Pick a random labelled tree $\tilde{T}_m^p$ such that

$$\mathbb{P}\left(\tilde{T}_m^p = T\right) \propto (1-p)^{-a(T)}, \quad T \in \mathbb{T}_{[m]}.$$

- Add each of the $a(\tilde{T}_m^p)$ permitted edges to $\tilde{T}_m^p$ independently with probability p.

Recipe for creating a connected graph on $[m]$

Lemma. $\tilde{G}_m^p$ has the same distribution as G_m^p, a component of $G(n,p)$ conditioned to have vertex-set $[m]$.

Recipe for creating a connected graph on $[m]$

Lemma. $\tilde{G}_m^p$ has the same distribution as G_m^p, a component of $G(n,p)$ conditioned to have vertex-set $[m]$.

Proof. For a connected graph G on $[m]$ which has $T(G) = T$ and surplus s,

$$\mathbb{P}\left(\tilde{G}_m^p = G\right) \propto (1-p)^{-a(T)} p^s (1-p)^{a(T)-s} = (p/(1-p))^s.$$

Recipe for creating a connected graph on $[m]$

Lemma. $\tilde{G}_m^p$ has the same distribution as G_m^p, a component of $G(n,p)$ conditioned to have vertex-set $[m]$.

Proof. For a connected graph G on $[m]$ which has $T(G) = T$ and surplus s,

$$\mathbb{P}\left(\tilde{G}_m^p = G\right) \propto (1-p)^{-a(T)} p^s (1-p)^{a(T)-s} = (p/(1-p))^s.$$

Likewise, by the definition of $G(n,p)$,

$$\mathbb{P}\left(G_m^p = G\right) \propto (p/(1-p))^s.$$

$\square$