Stochastic Processes and their Applications, Oaxaca, 19th-24th June 2011

The scaling limit of critical random graphs

Christina Goldschmidt
University of Warwick

Joint work with

Louigi Addario-Berry
McGill University

Nicolas Broutin
INRIA Rocquencourt

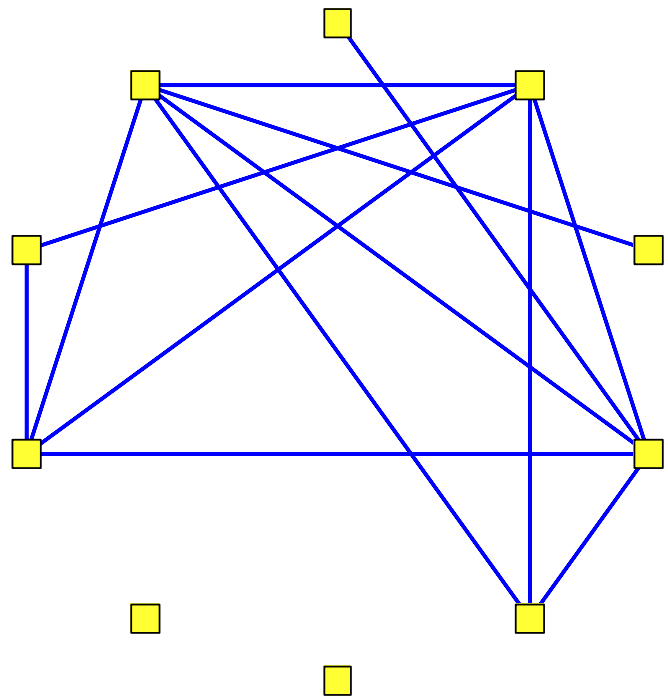
Grégory Miermont
Université Paris-Sud

Part I: the Erdős-Rényi random graph

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 $G(n, p)$.

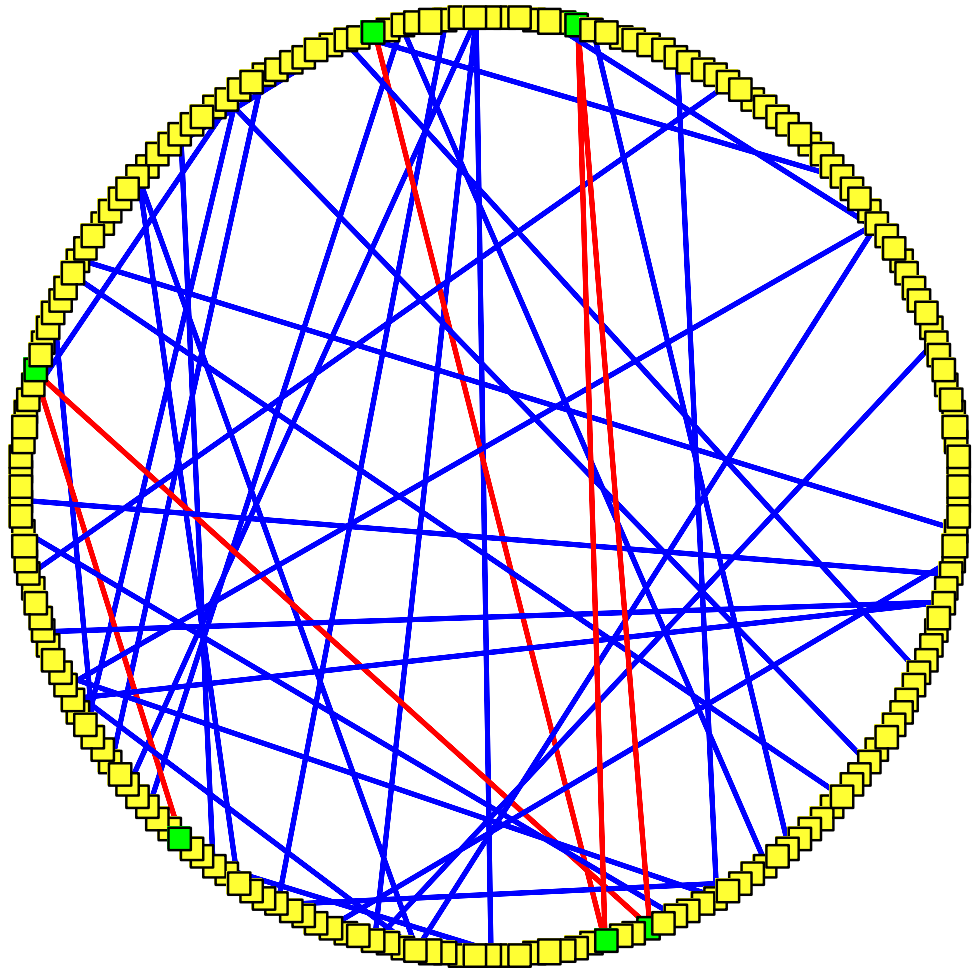
Example: $n = 10$, $p = 0.4$ (vertex labels omitted).



The phase transition

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

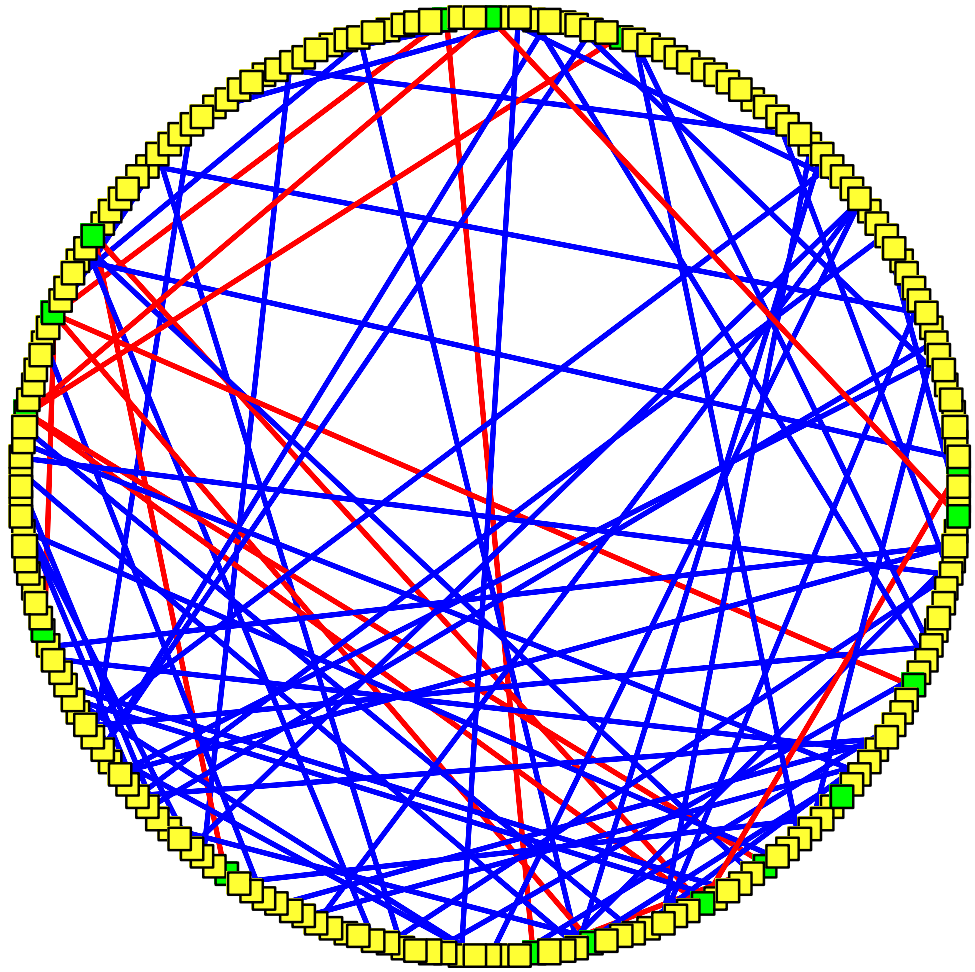
$n = 200,\ c = 0.4$



The phase transition

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

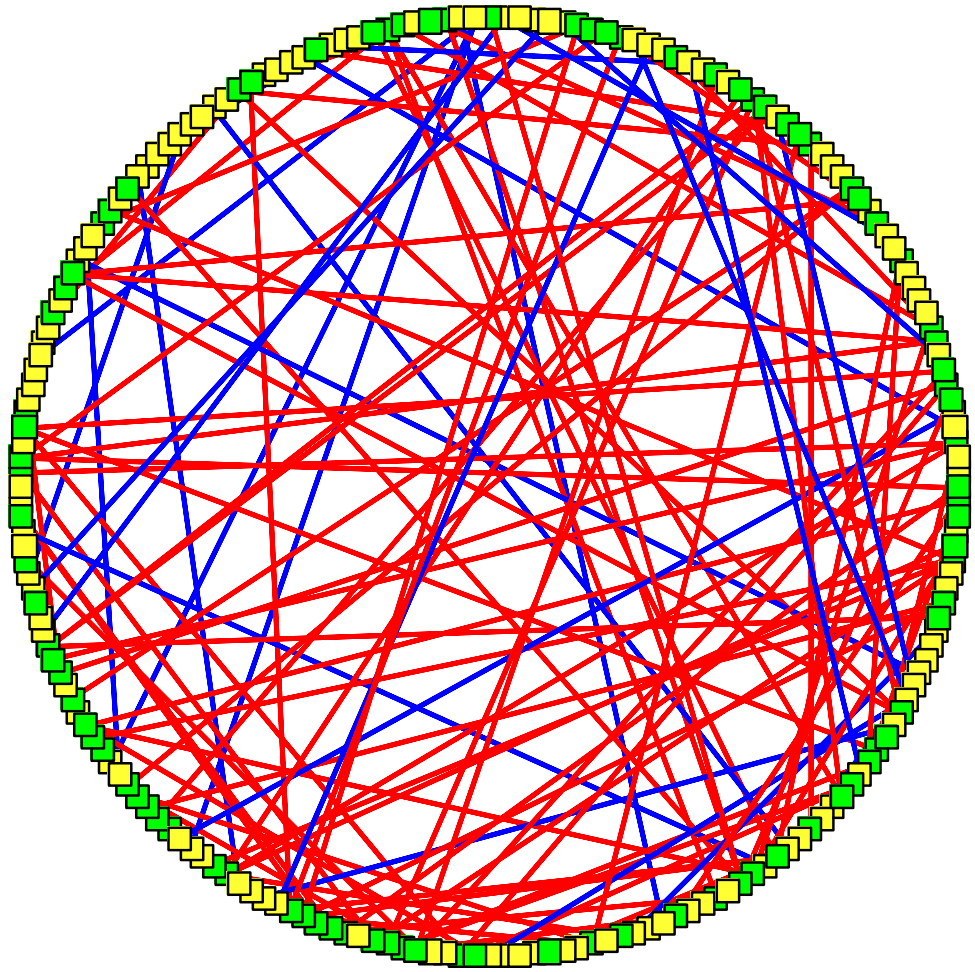
$n = 200,\ c = 0.8$



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

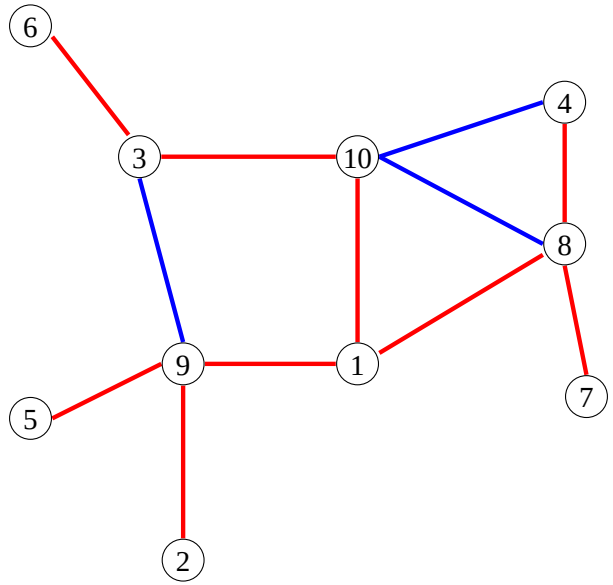
The critical window: $p = \frac{1}{n} + \frac{\lambda}{n^{4/3}}$, where $\lambda \in \mathbb{R}$. For such p, the largest components have size $\Theta(n^{2/3})$.

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 need to think about 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 $n \to \infty$,

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

Convergence for the first co-ordinate takes place in

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

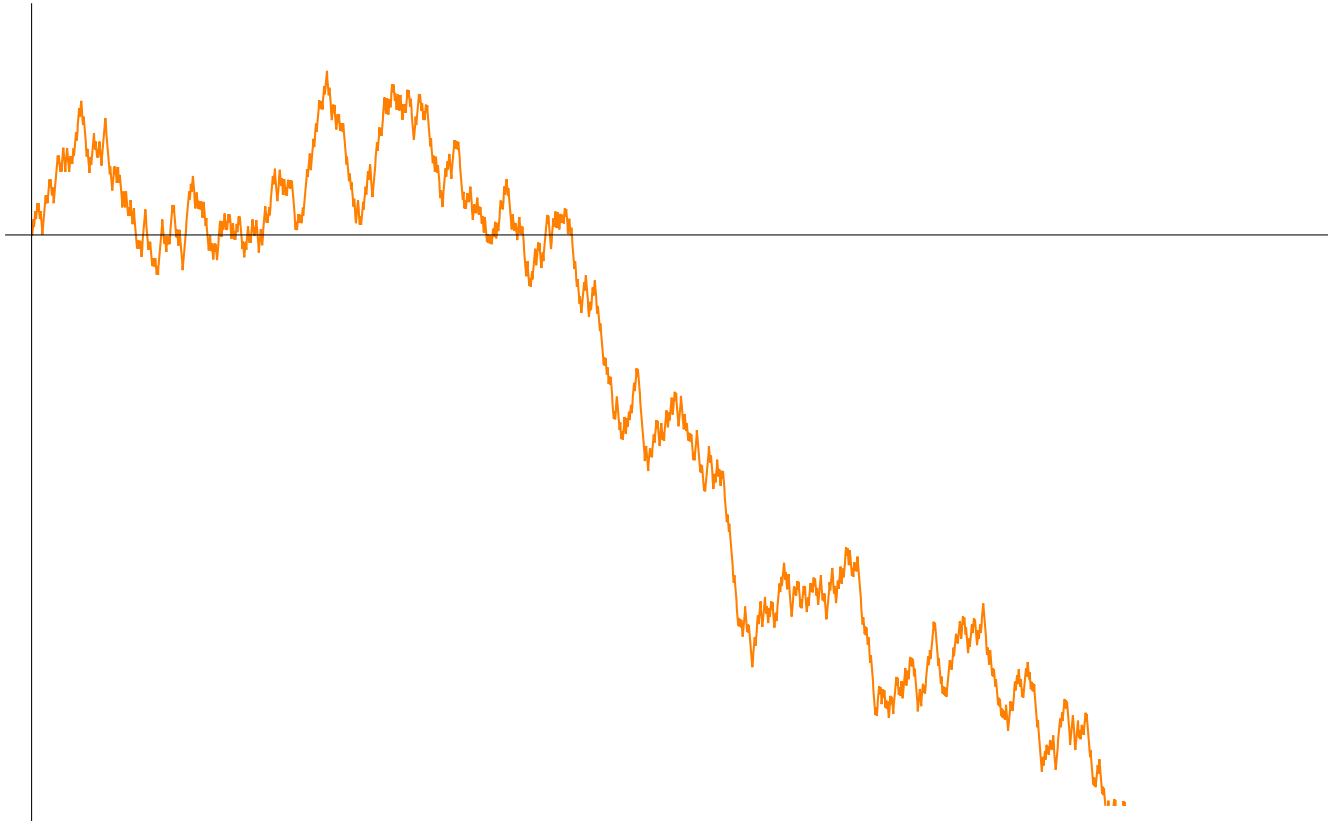
and for the second in the sense of finite-dimensional distributions.

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.

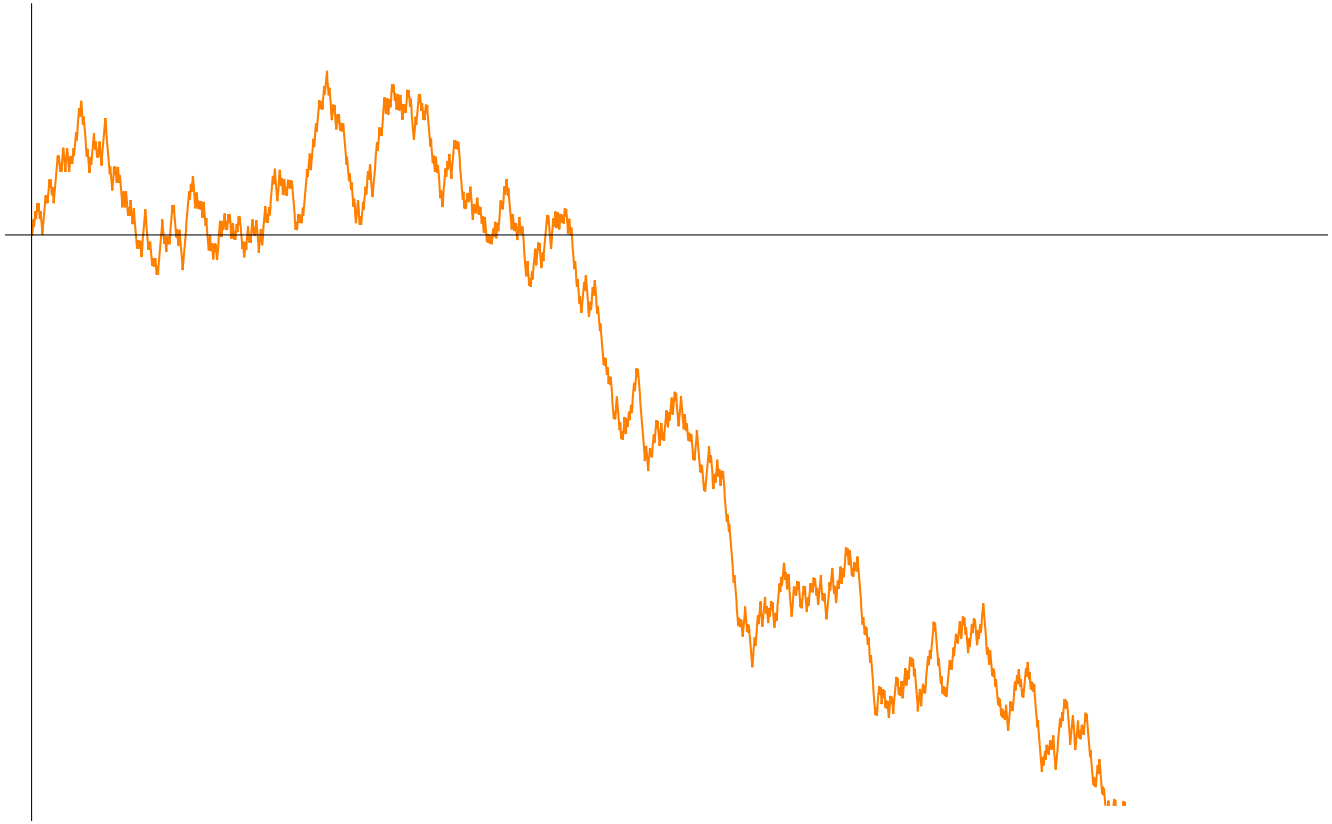


[Pictures by Louigi Addario-Berry]

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.

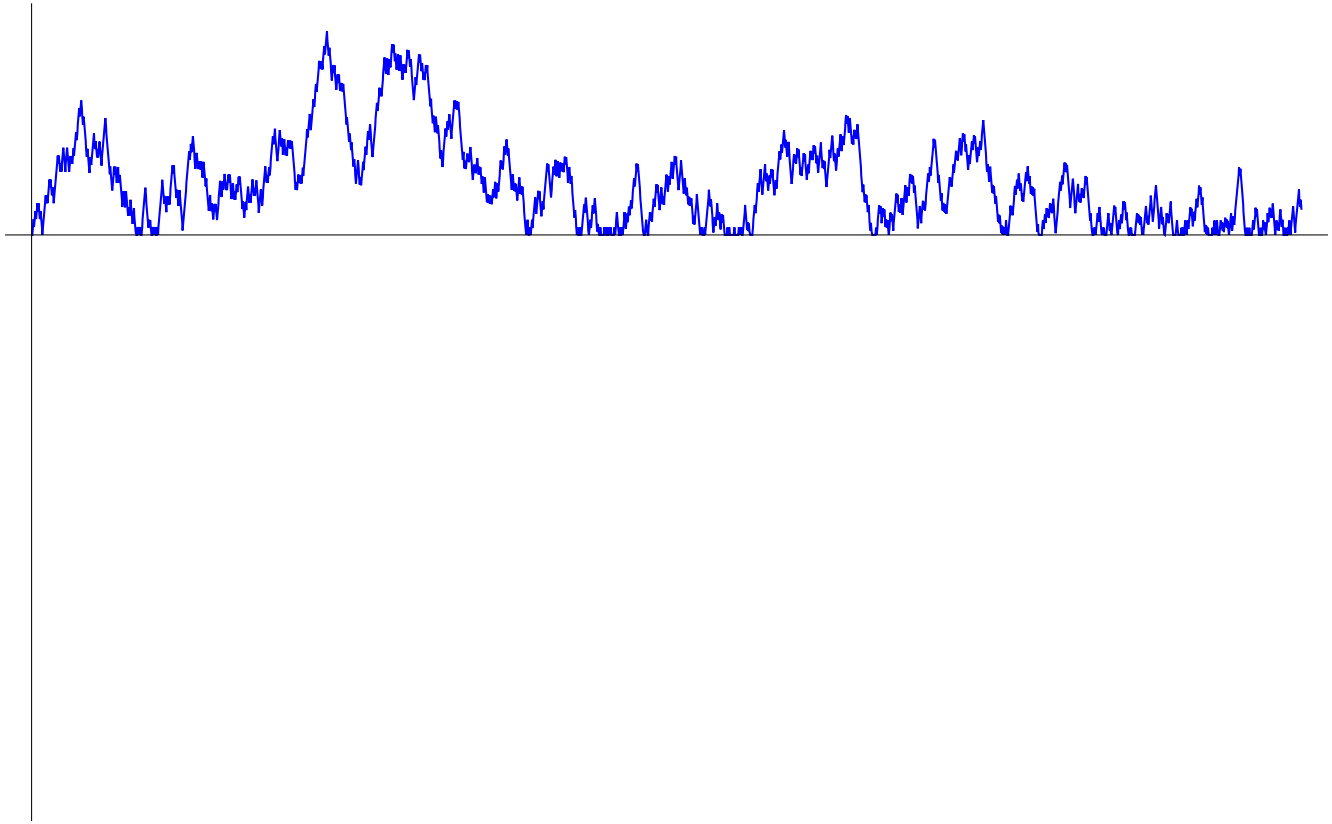


[Pictures by Louigi Addario-Berry]

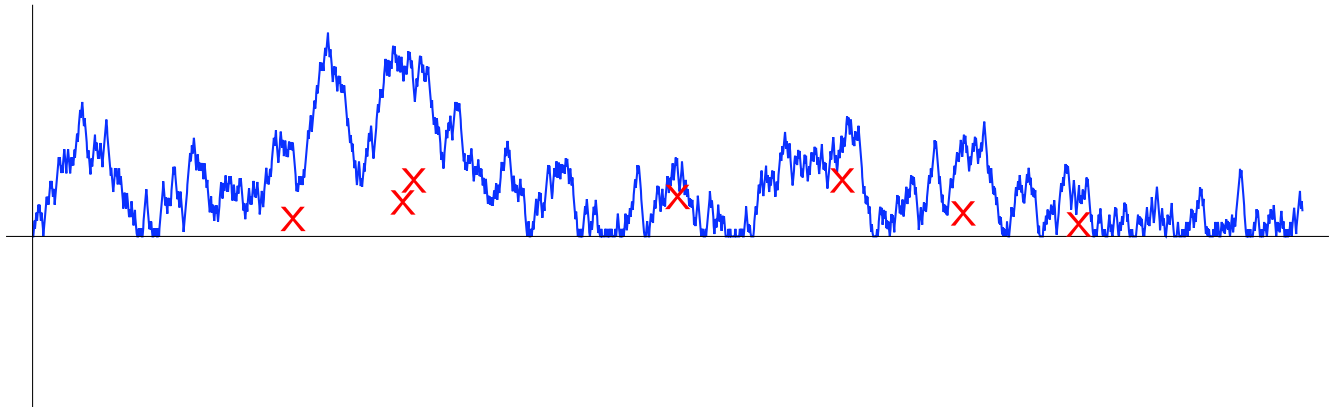
Limiting sizes and surpluses

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.



[Pictures by Louigi Addario-Berry]



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.

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.

Convergence will be in the sense of the Gromov-Hausdorff distance between compact metric spaces.

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.

General approach: 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.

General approach: 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.

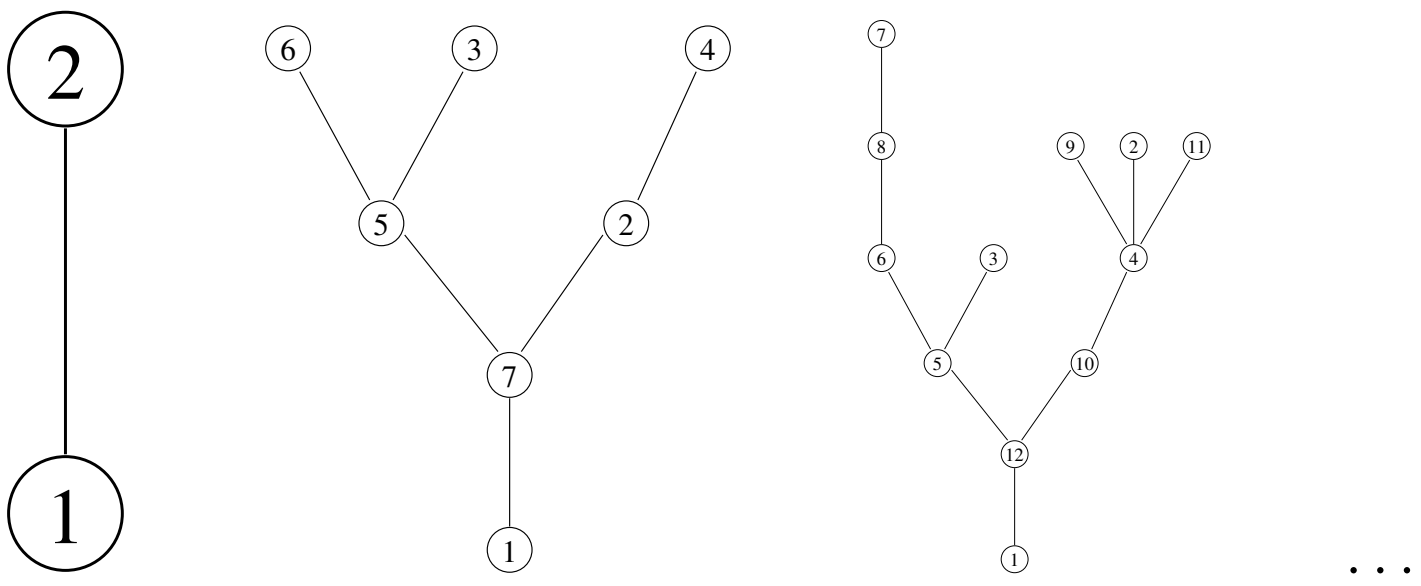
Warm-up: the tree case

Take a uniform random tree T_m on vertices labelled by $[m] = \{1, 2, \ldots, m\}$ and, using the ordinary graph distance, think of it as a metric space.

Warm-up: the tree case

Take a uniform random tree T_m on vertices labelled by $[m] = \{1, 2, \ldots, m\}$ and, using the ordinary graph distance, think of it as a metric space.

Now rescale the edge-lengths by $1/\sqrt{m}$:



Warm-up: the tree case

Theorem (Aldous (1993)) As $m \to \infty$,

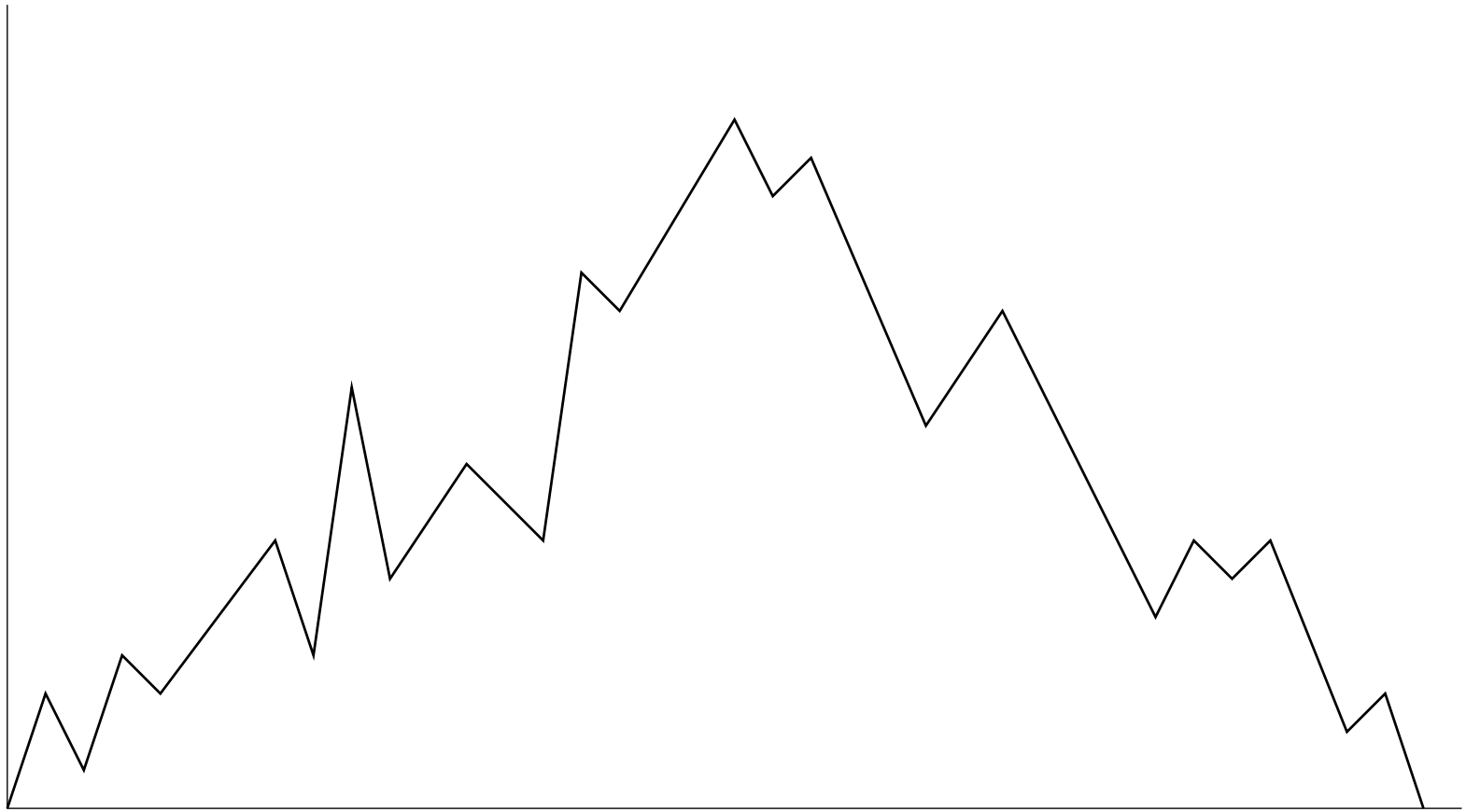
$$\frac{1}{\sqrt{m}} T_m \xrightarrow{d} \mathcal{T},$$

where the convergence is in the Gromov-Hausdorff distance.

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

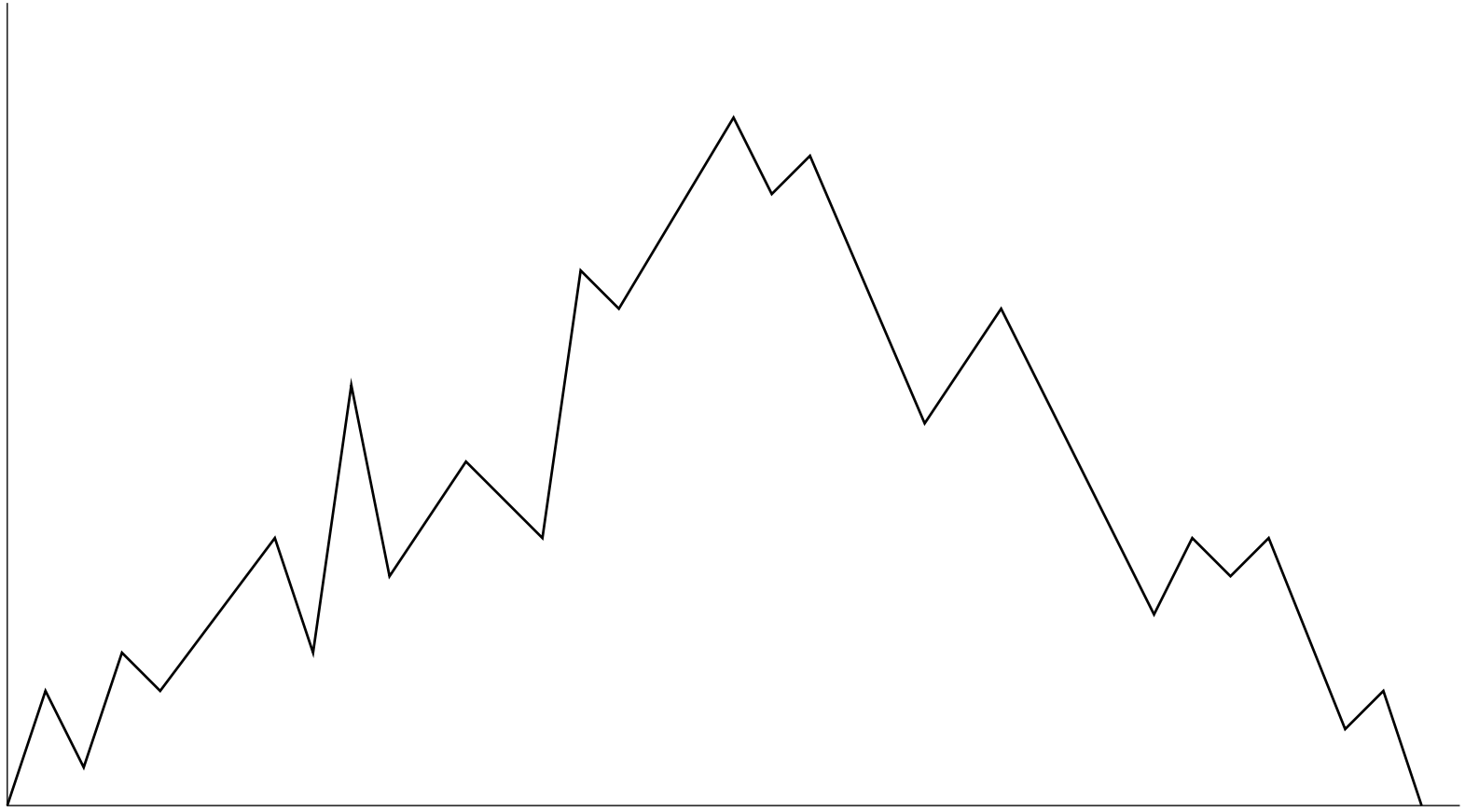
Trees from excursions

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

Now put glue on the underside of the excursion and push the two sides together...



Trees from excursions

Now put glue on the underside of the excursion and push the two sides together...

Trees from excursions

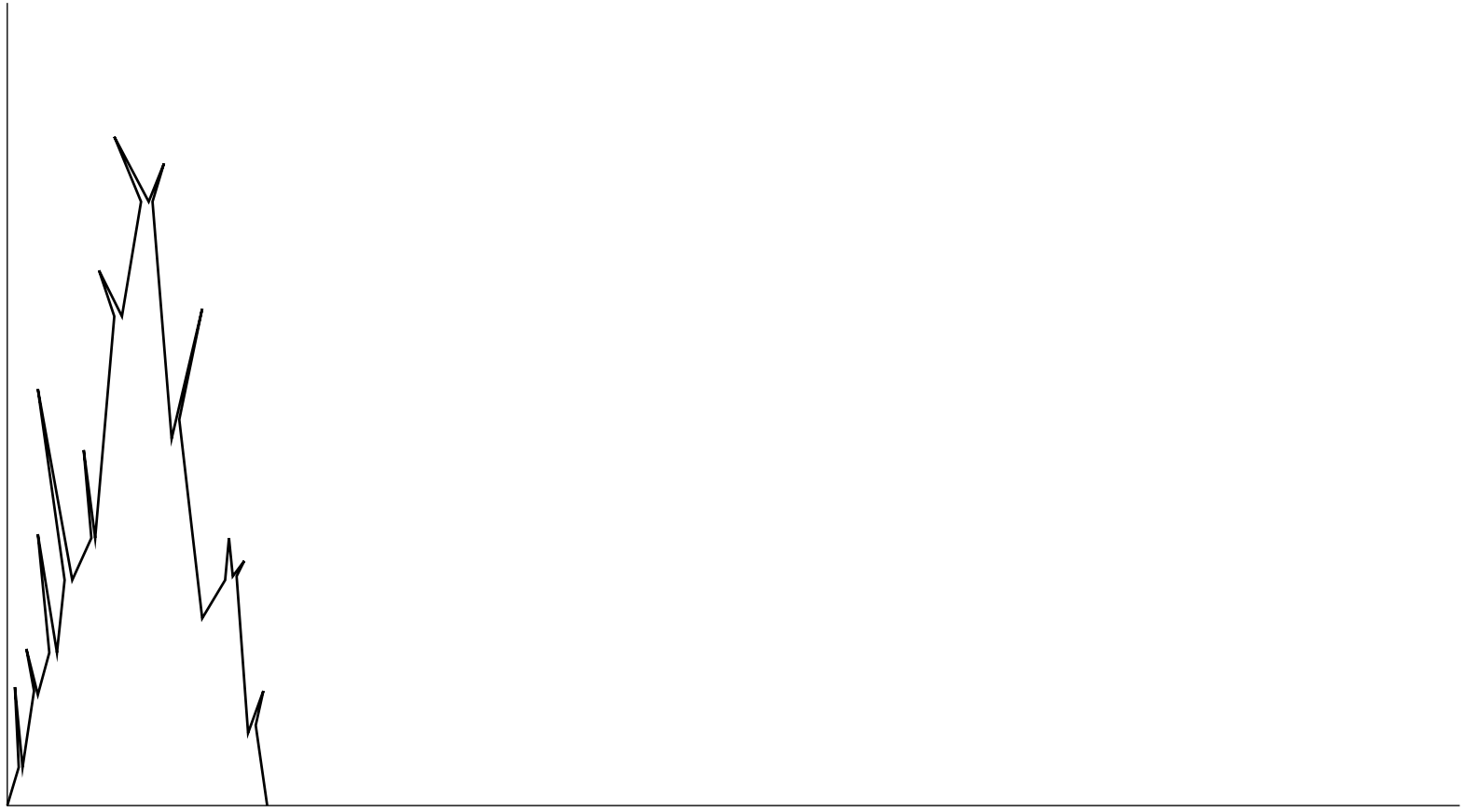
Now put glue on the underside of the excursion and push the two sides together...

Trees from excursions

Now put glue on the underside of the excursion and push the two sides together...

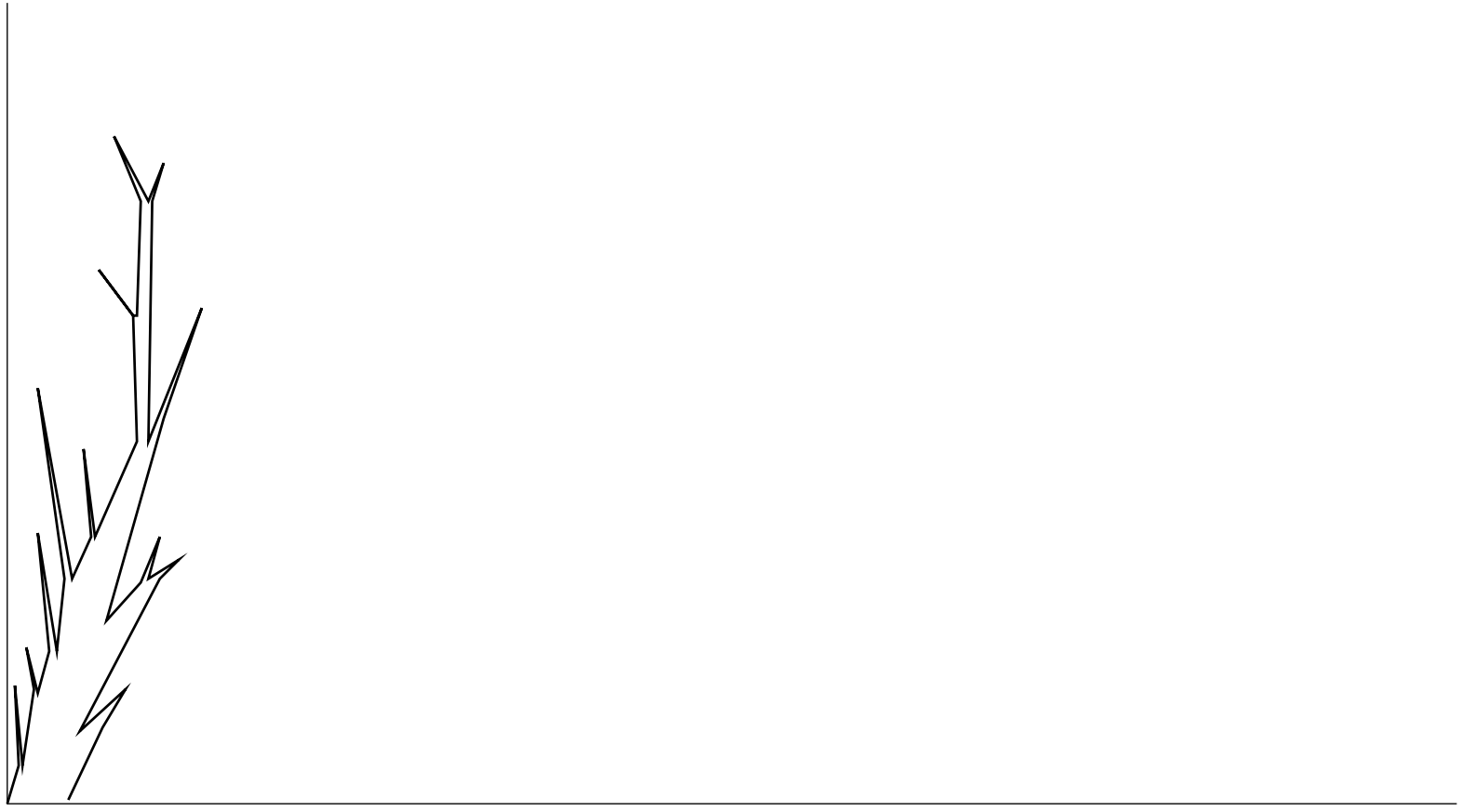
Trees from excursions

Now put glue on the underside of the excursion and push the two sides together...



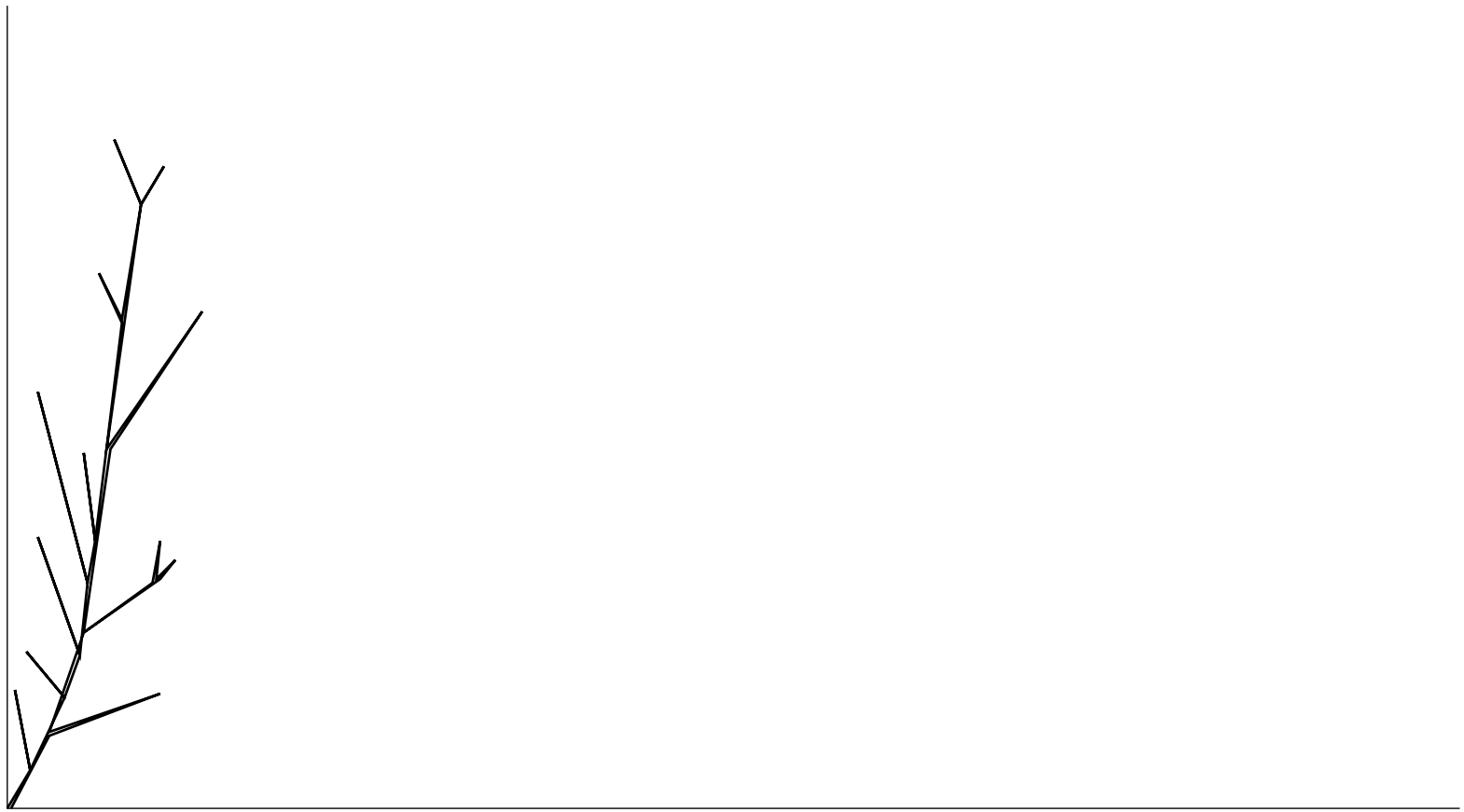
Trees from excursions

Now put glue on the underside of the excursion and push the two sides together...



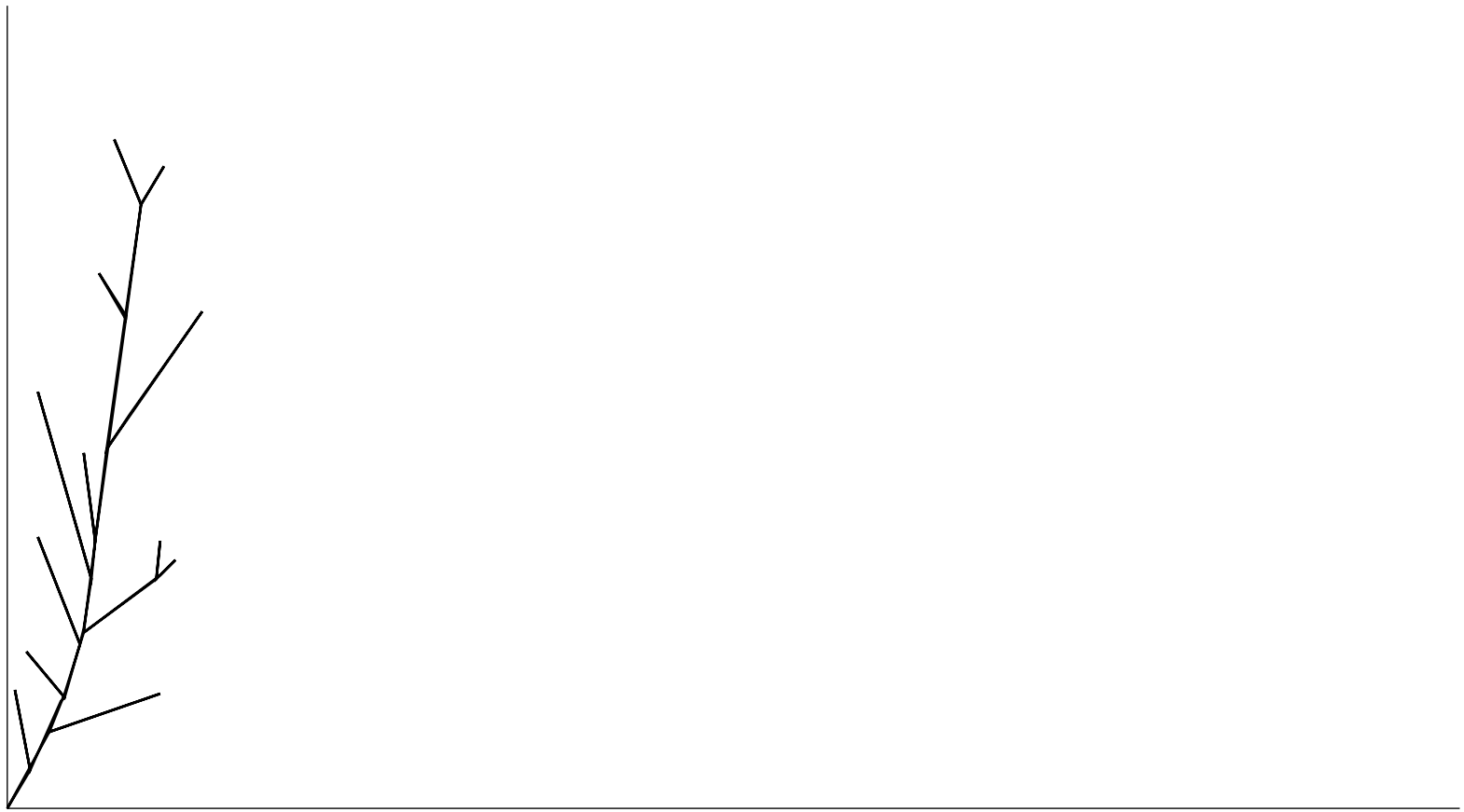
Trees from excursions

Now put glue on the underside of the excursion and push the two sides together...

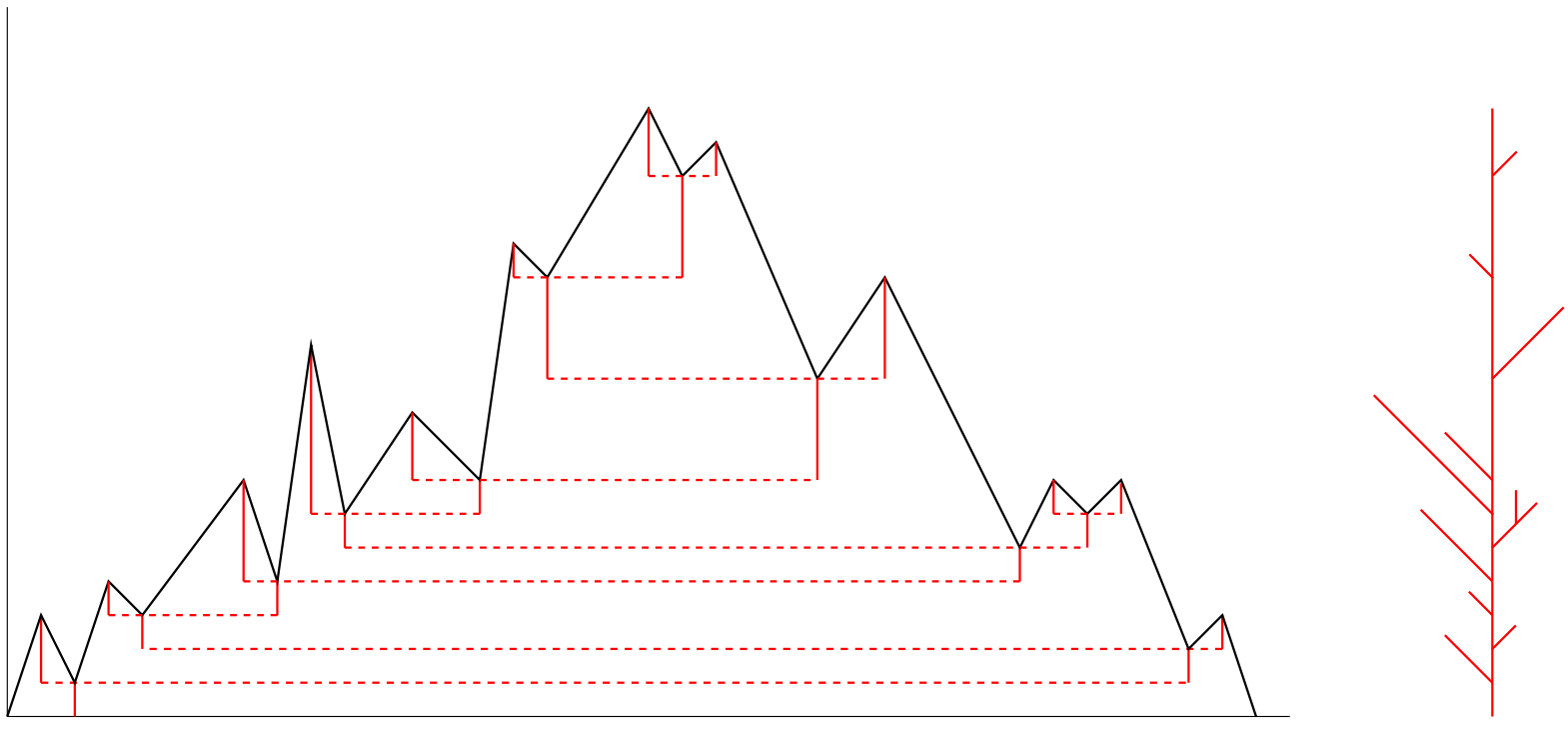


Trees from excursions

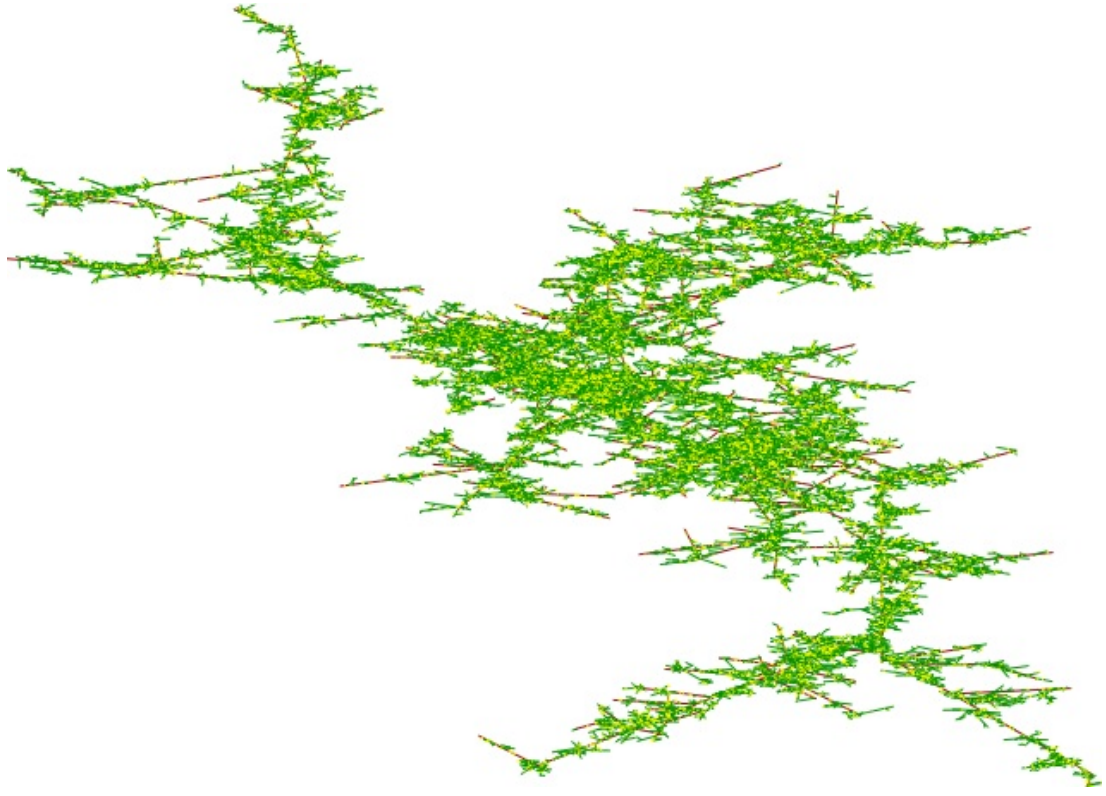
Now put glue on the underside of the excursion and push the two sides together to get a tree.



Trees from excursions



Trees from excursions



[Picture by Grégory Miermont]

The Brownian continuum random tree is the tree we obtain by doing this glueing procedure to the function $2e$, where $(e(x), 0 \leq x \leq 1)$ a standard Brownian excursion.

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

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

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* scaled 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

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* scaled 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

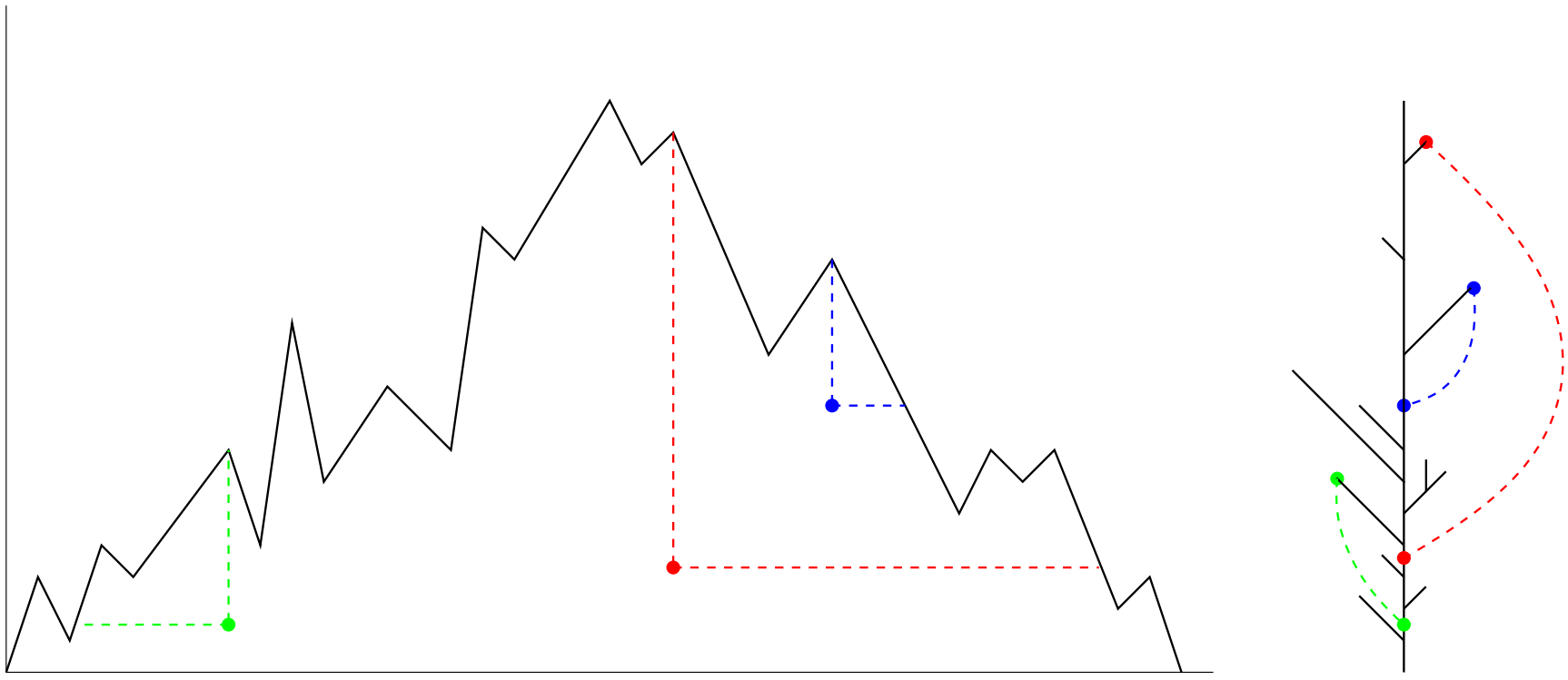
Consider the process $(B^\lambda(t), t \geq 0)$. An excursion $\tilde{e}^{(x)}$ of this process, conditioned to have length x, has a distribution specified by

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

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



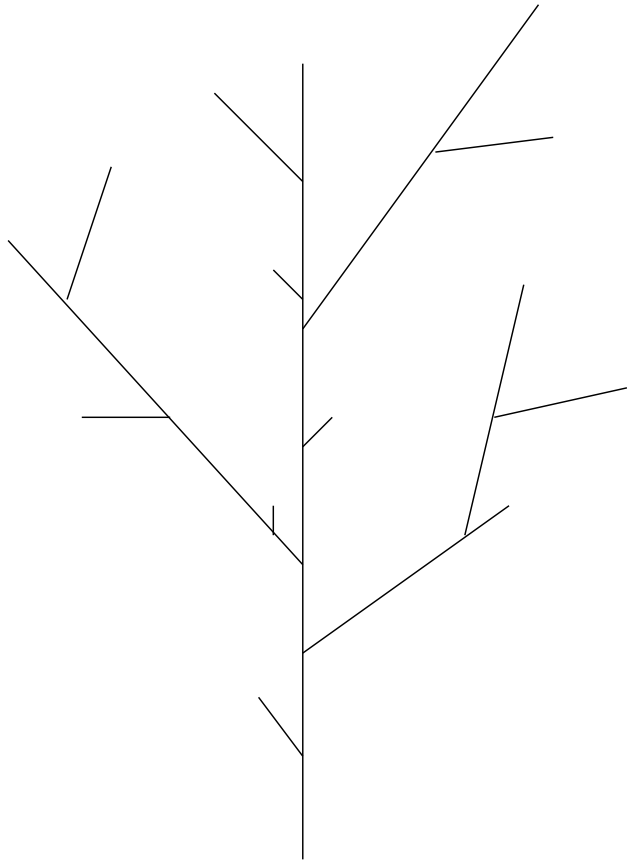
A point at (x, y) identifies the vertex v at height $h(x)$ with the vertex at distance y along the path from the root to v.

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.

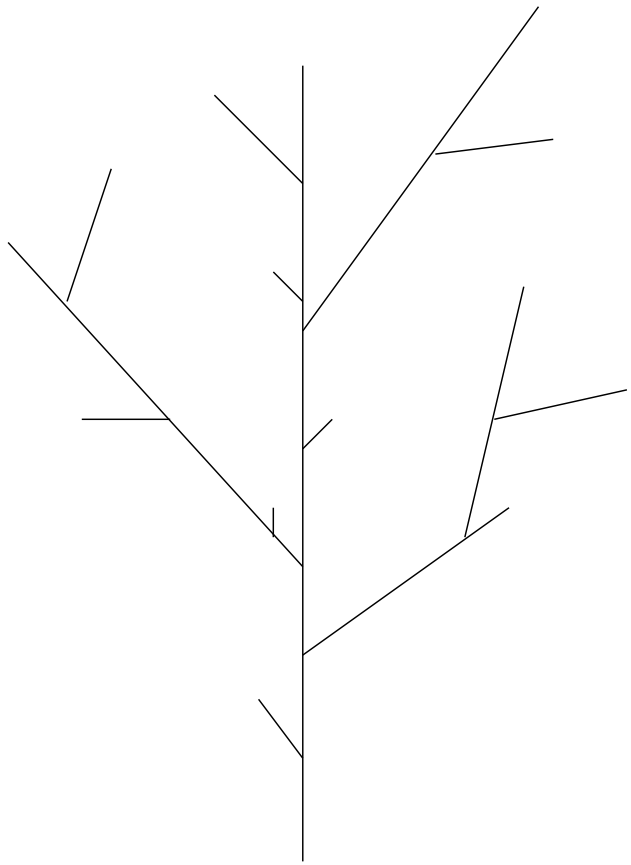
A limiting component

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



A limiting component

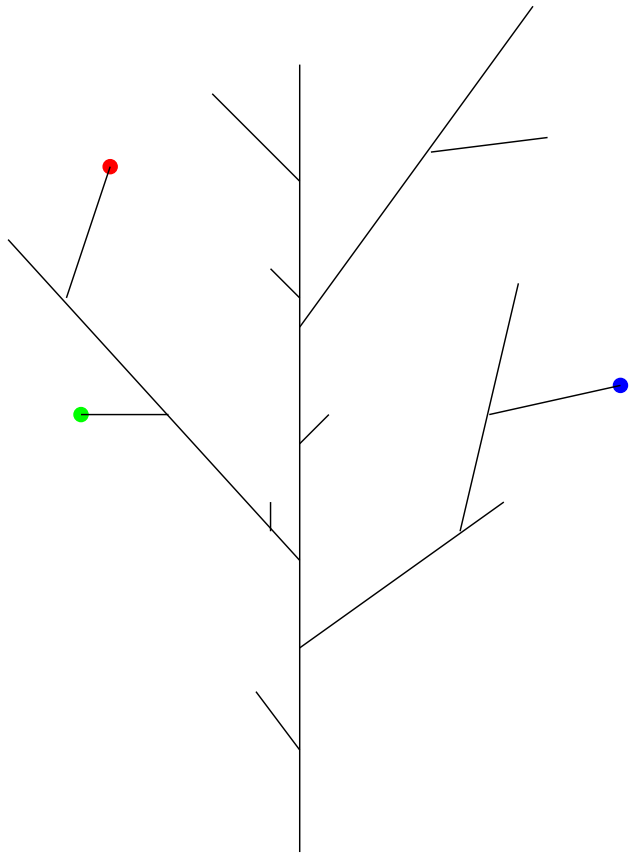
Sample a tilted excursion $\tilde{e}^{(x)}$ of length x and use it to create a CRT $\tilde{\mathcal{T}}$.



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

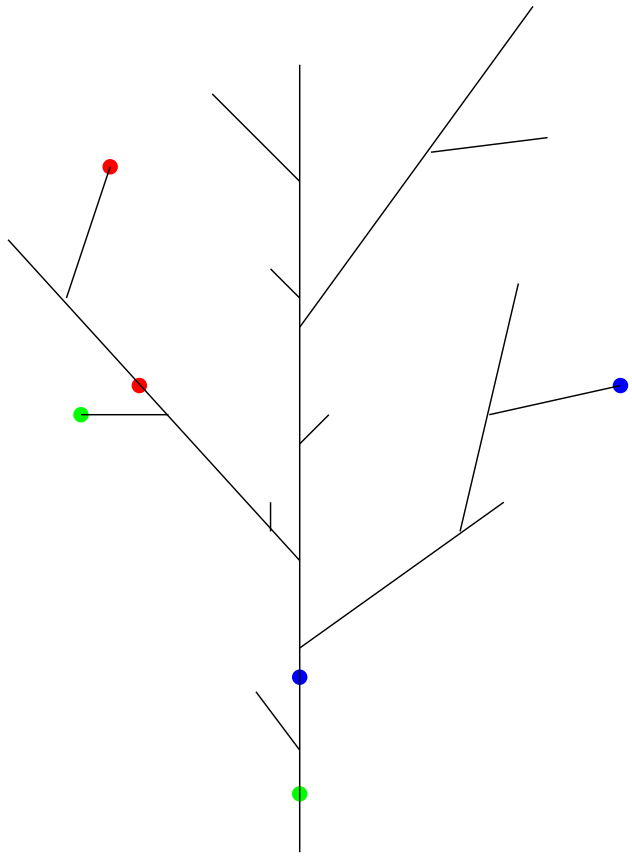
A limiting component

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.)



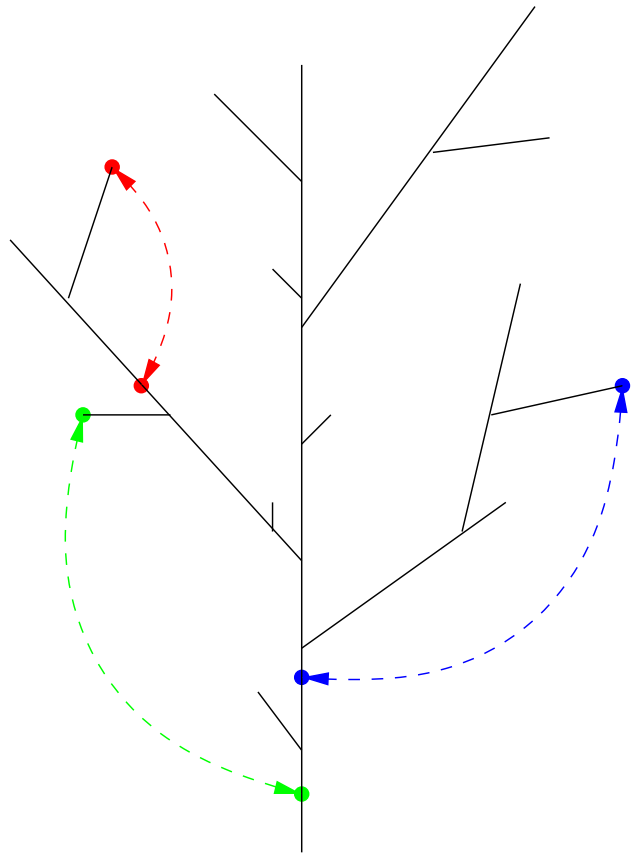
A limiting component

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 distance

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

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 four states: unexplored, 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).

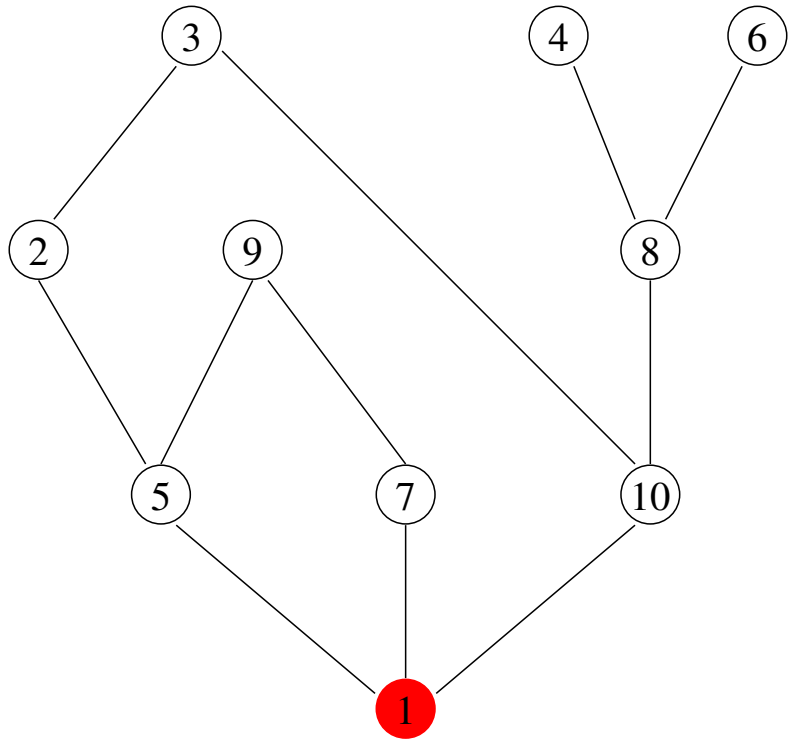
We will begin by taking an arbitrary connected graph G and picking out a particular spanning tree via a depth-first exploration.

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

Let $X(k)$ be the number of vertices alive at step k.

Depth-first exploration

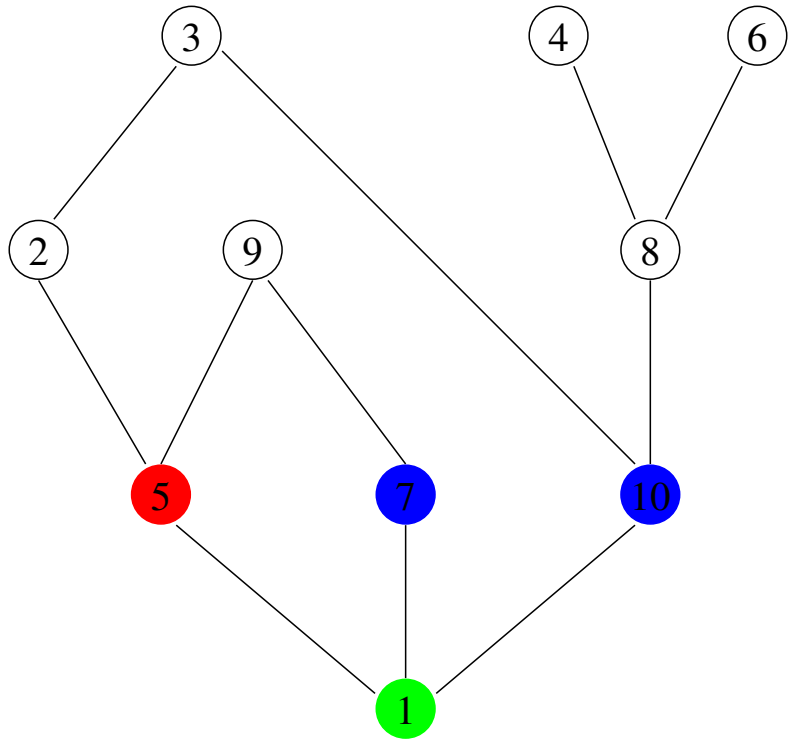
Step 0: initialization



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

Depth-first exploration

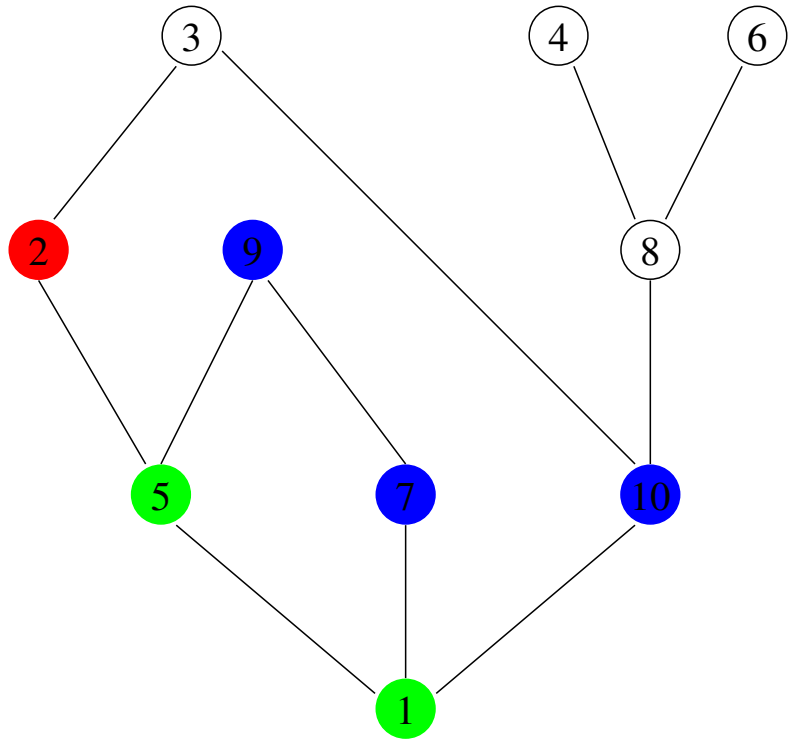
Step 1



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

Depth-first exploration

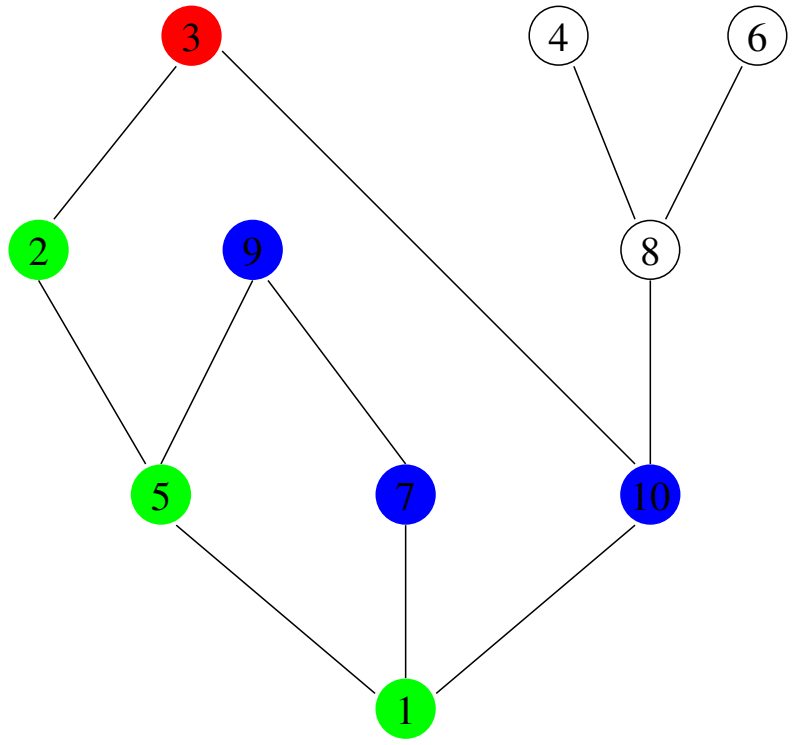
Step 2



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

Depth-first exploration

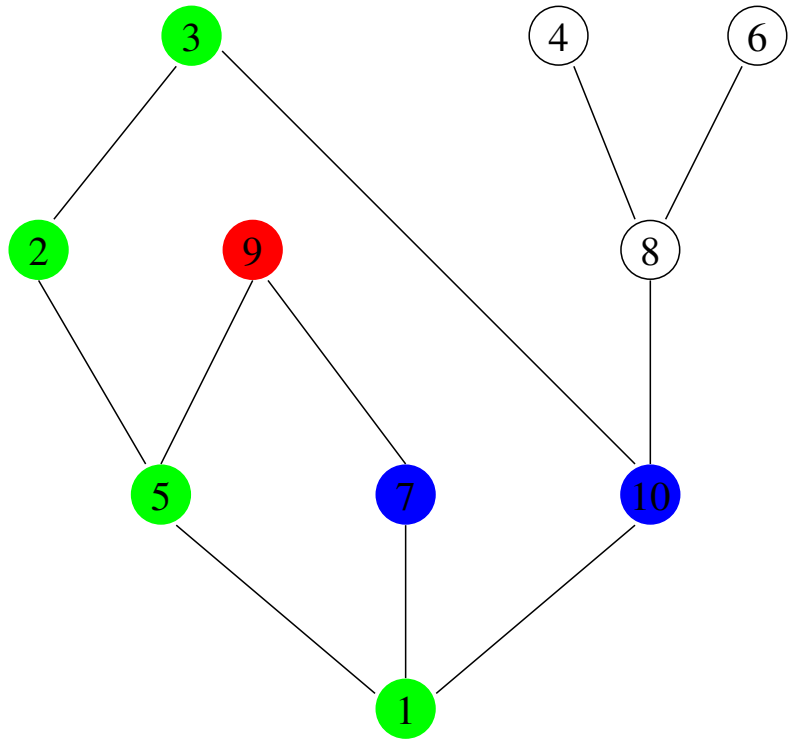
Step 3



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

Depth-first exploration

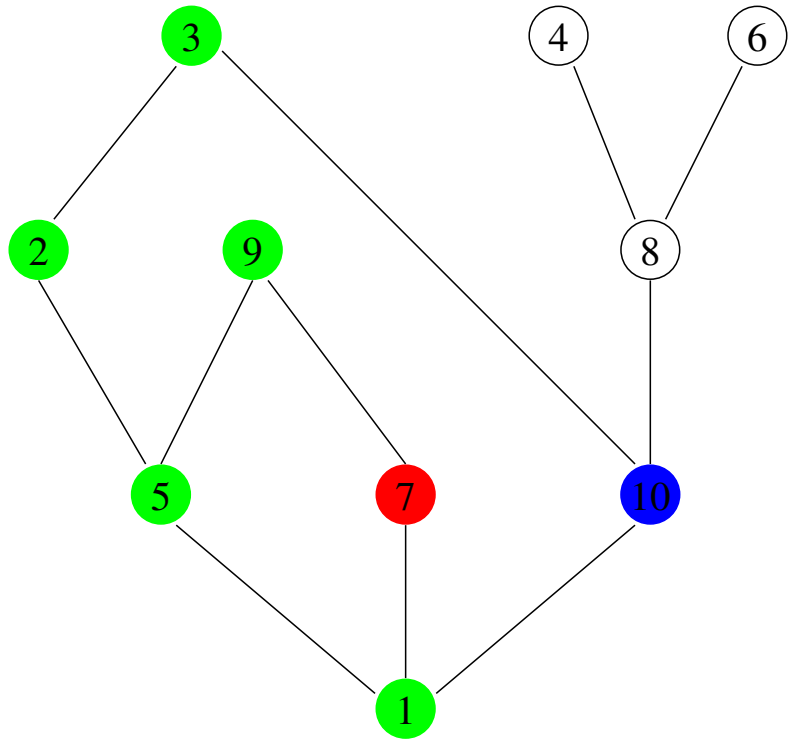
Step 4



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

Depth-first exploration

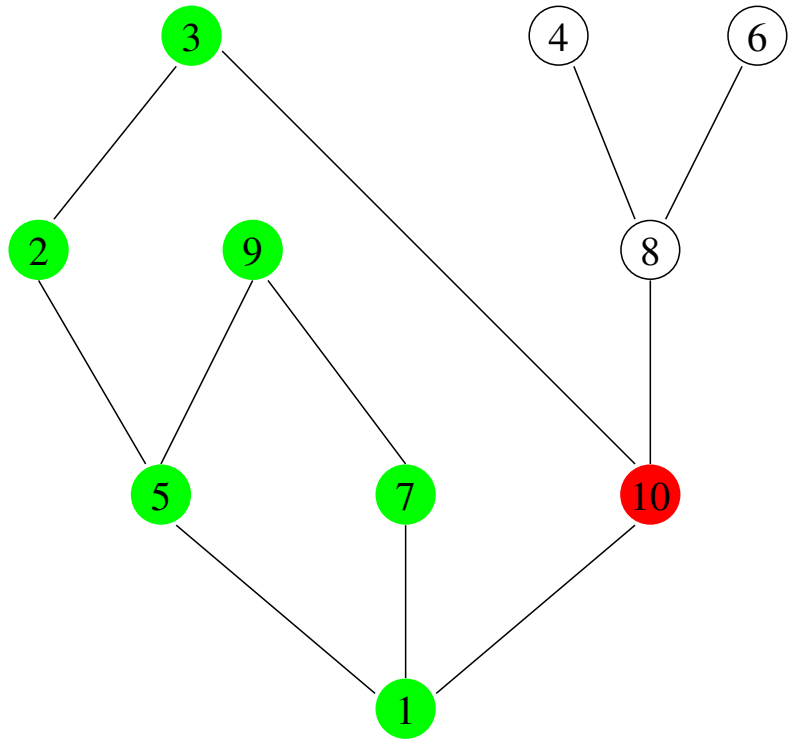
Step 5



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

Depth-first exploration

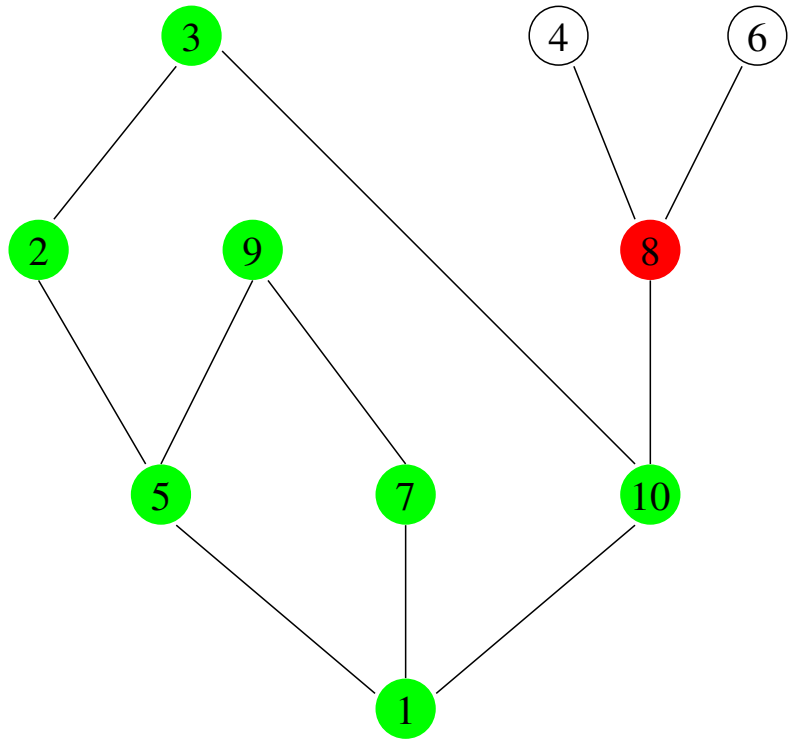
Step 6



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

Depth-first exploration

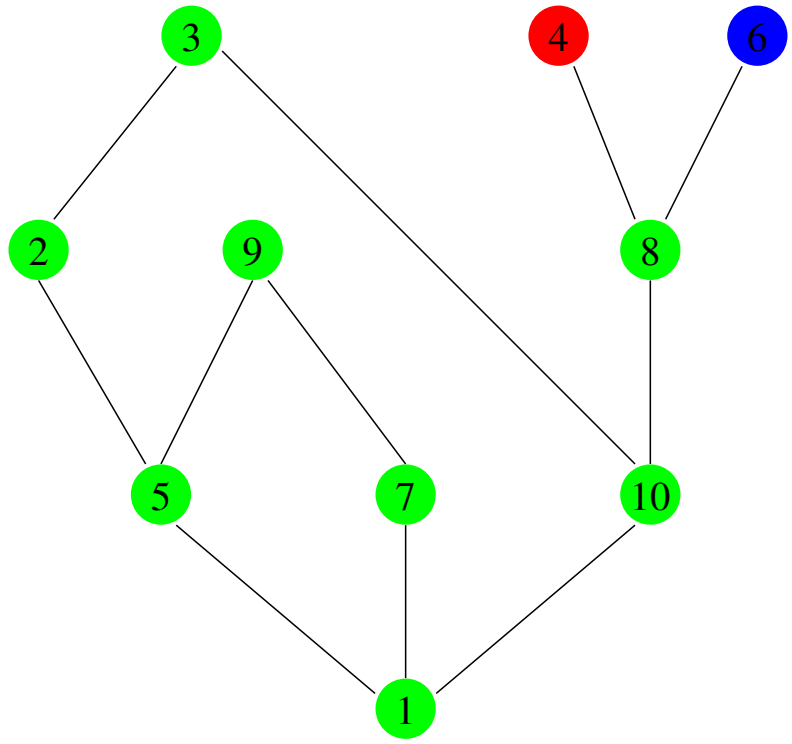
Step 7



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

Depth-first exploration

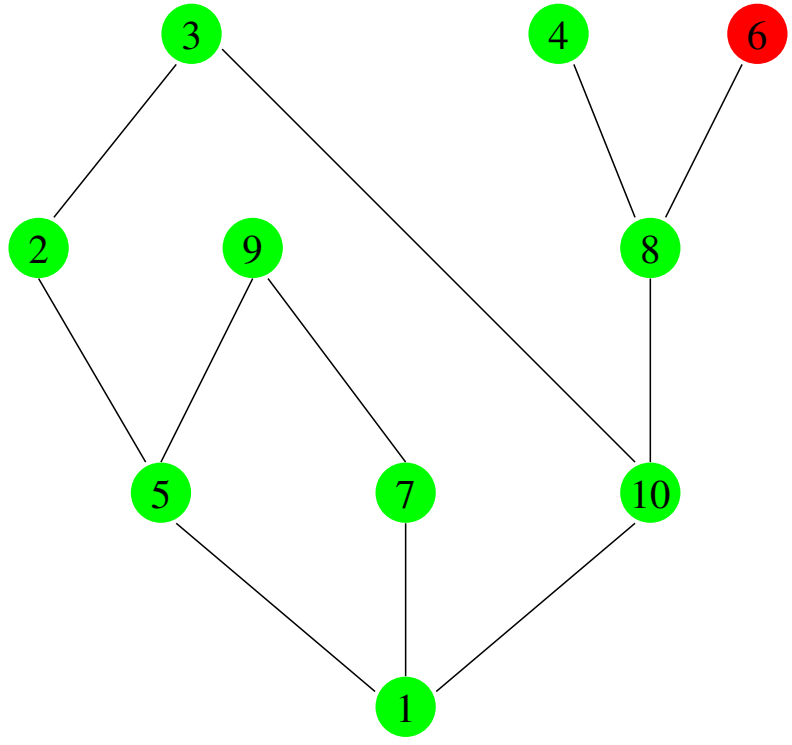
Step 8



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

Depth-first exploration

Step 9

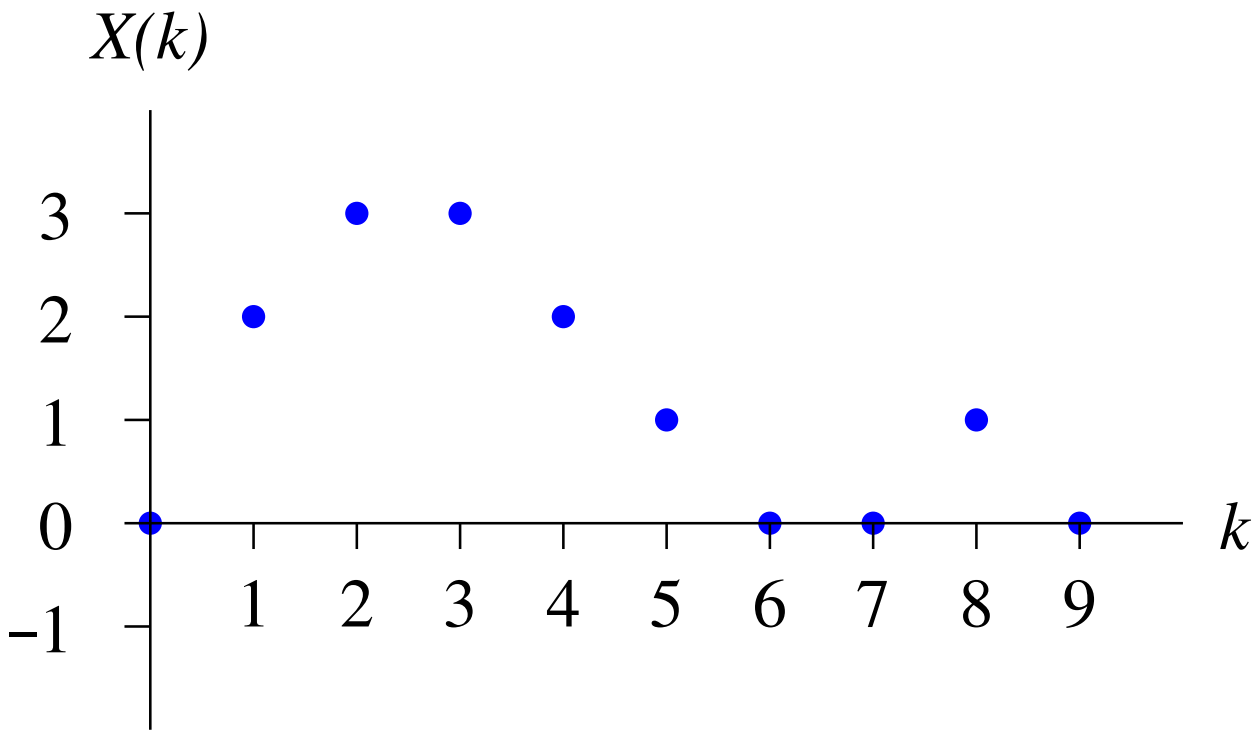


Current: 6 Alive: none Dead: 1, 5, 2, 3, 9, 7, 10, 8, 4

$X(9) = 0$.

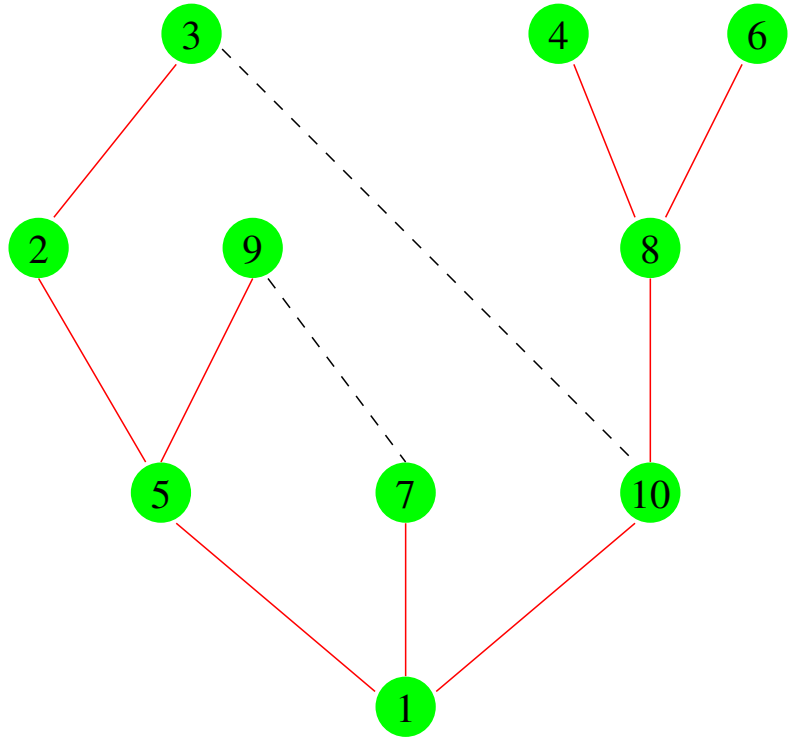
Depth-first walk

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



Depth-first tree

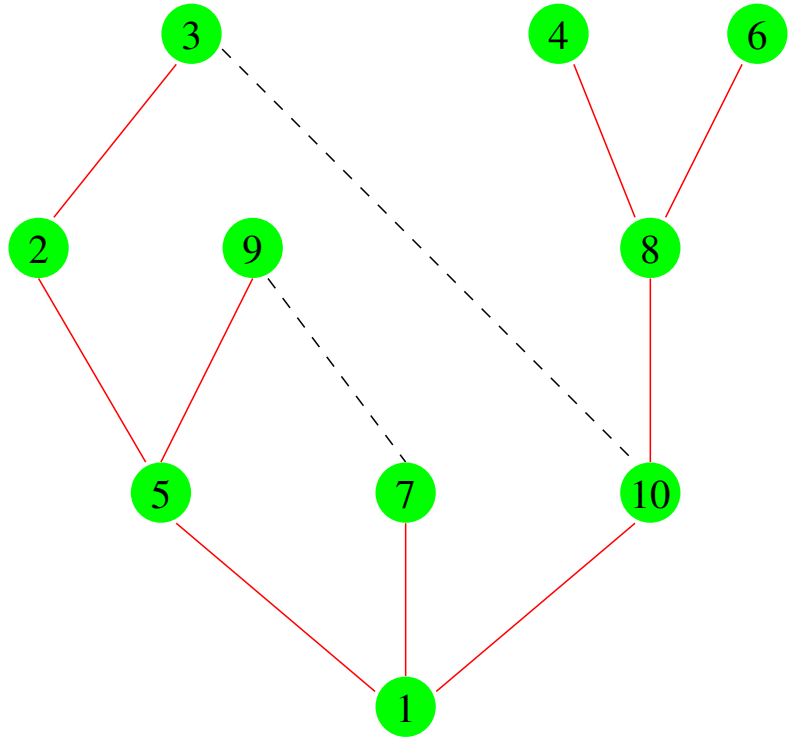
We essentially explored this tree; the dashed edges made no difference to the depth-first exploration.



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 $T(G)$. It turns out that we can recover this tree (although not its labels) from the process X.

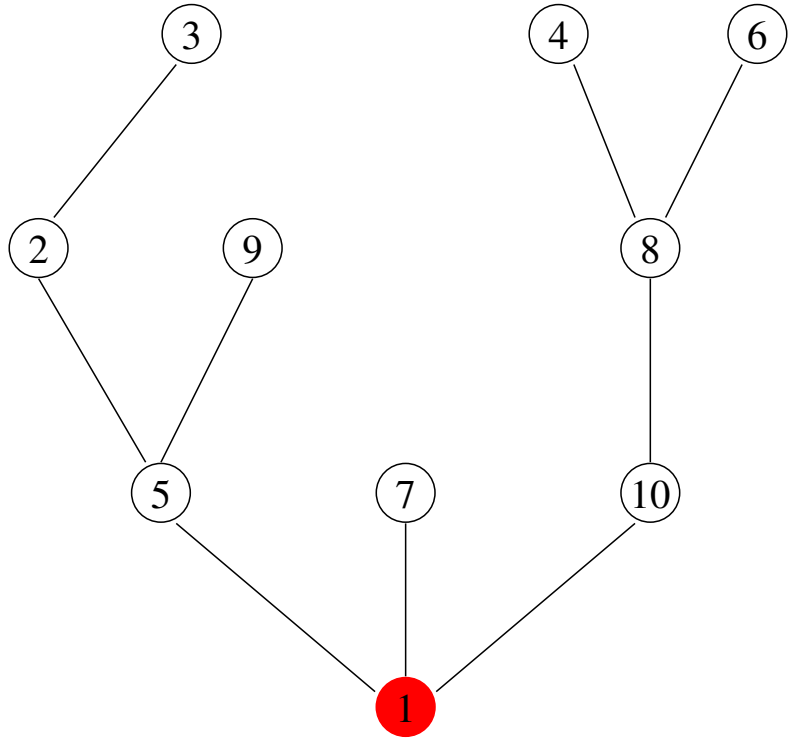
Permitted edges

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?

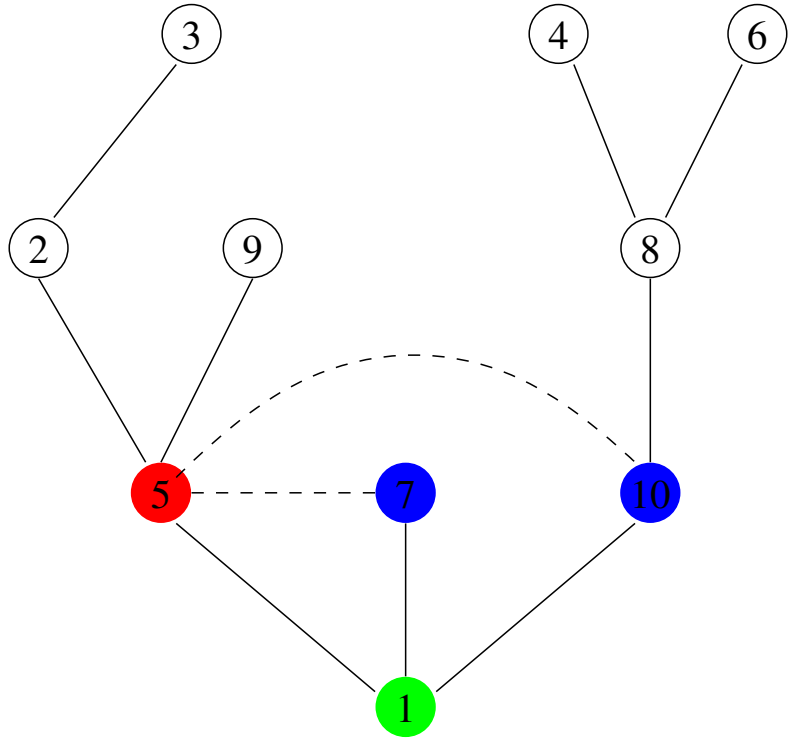
Call such edges permitted.

Depth-first walk and permitted edges



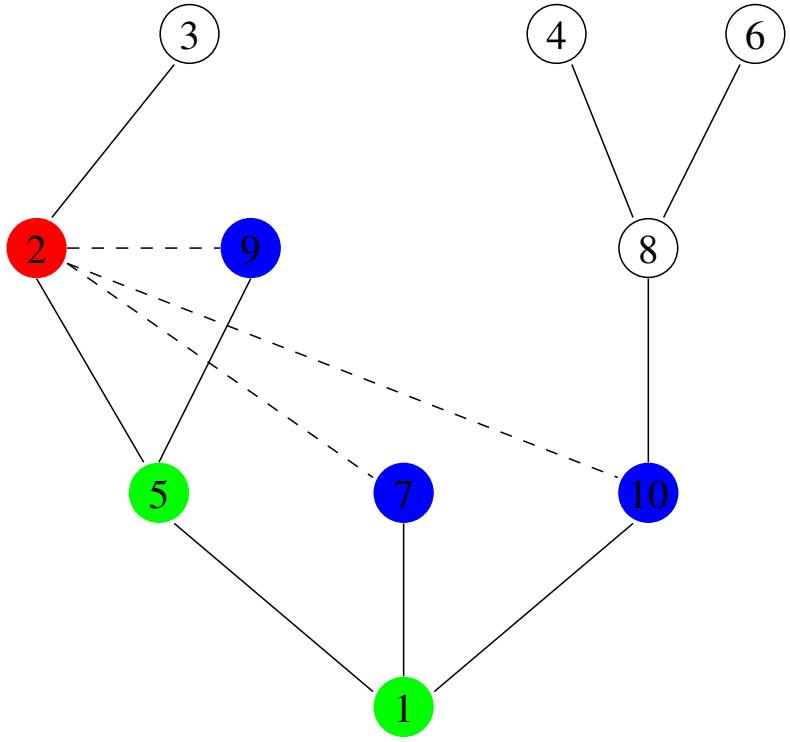
Step 0: $X(0) = 0$.

Depth-first walk and permitted edges



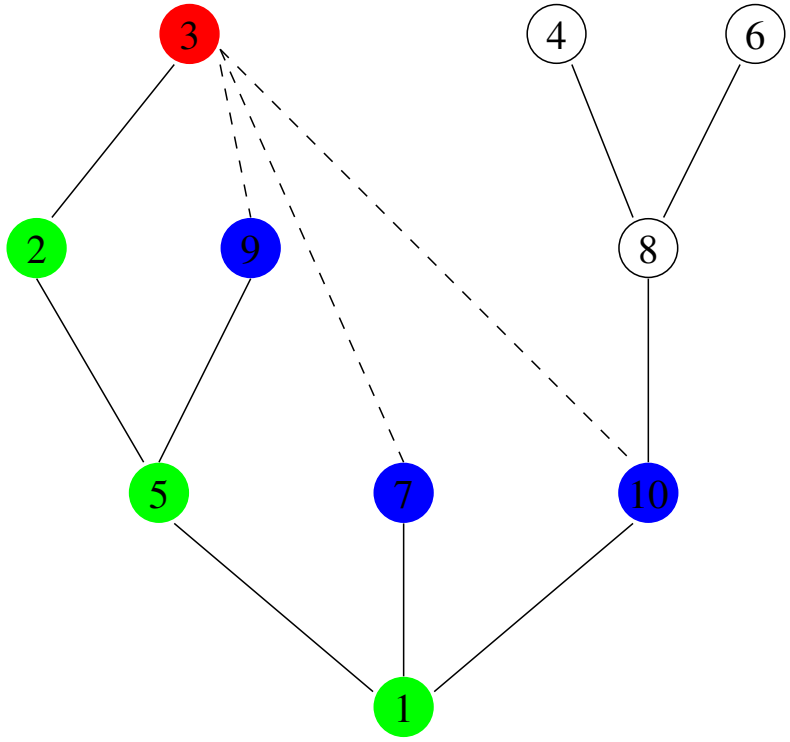
Step 1: $X(1) = 2$.

Depth-first walk and permitted edges



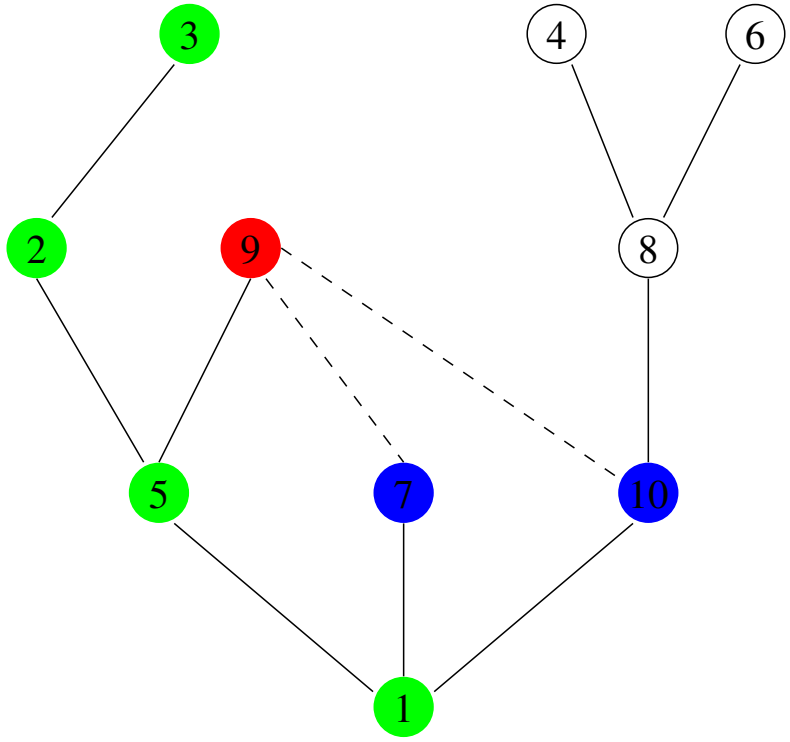
Step 2: $X(2) = 3$.

Depth-first walk and permitted edges



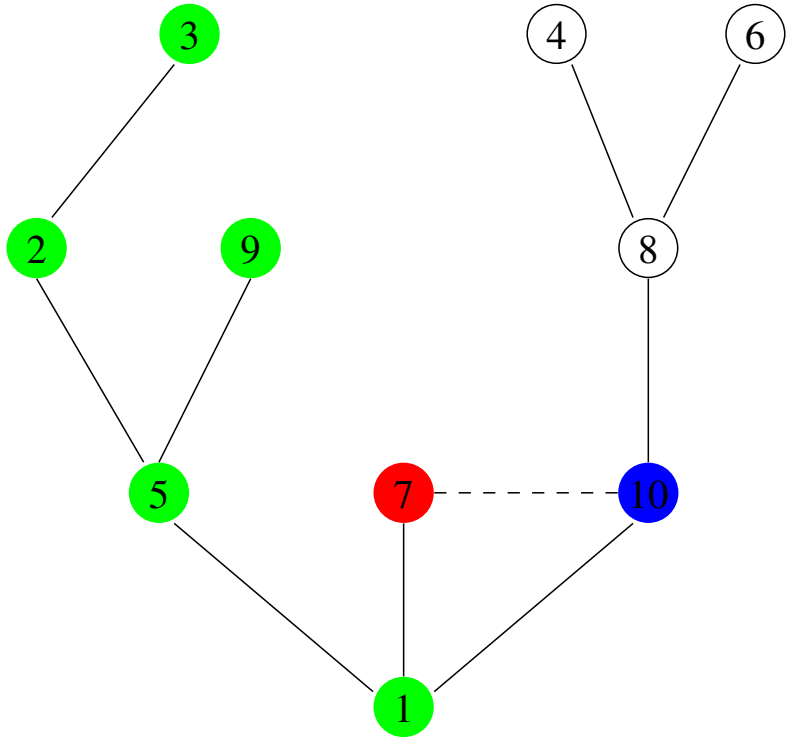
Step 3: $X(3) = 3$.

Depth-first walk and permitted edges



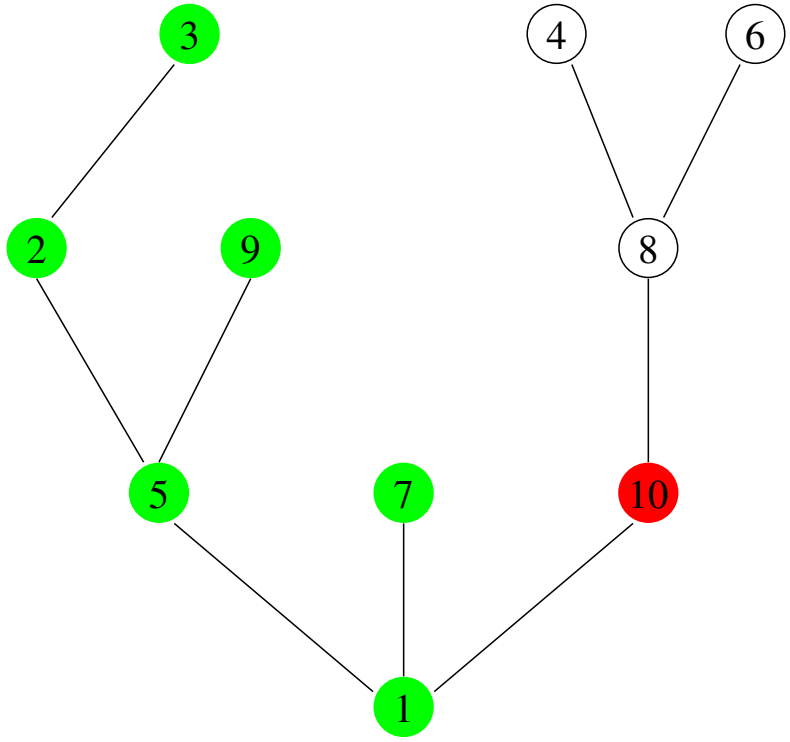
Step 4: $X(4) = 2$.

Depth-first walk and permitted edges



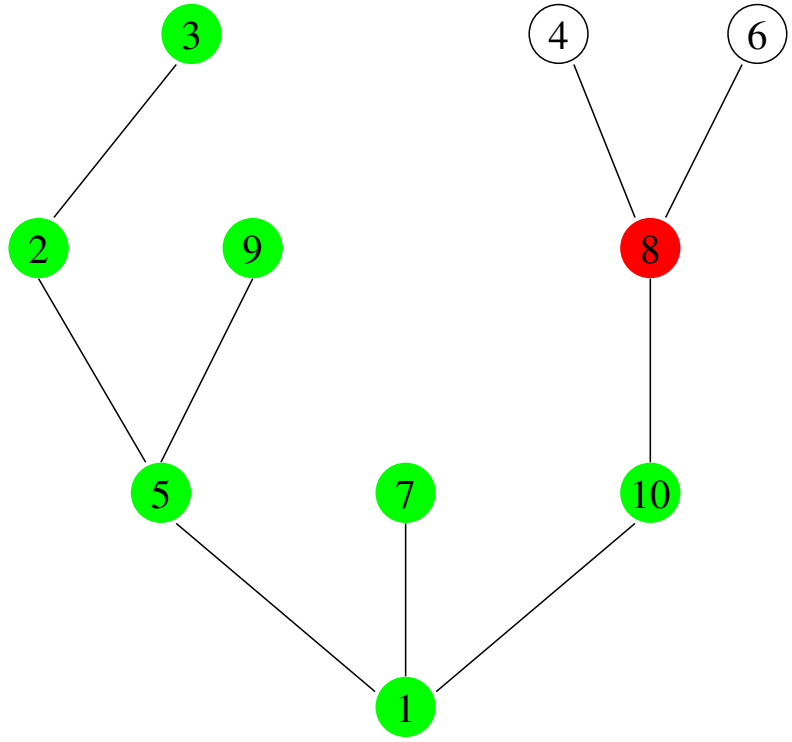
Step 5: $X(5) = 1$.

Depth-first walk and permitted edges



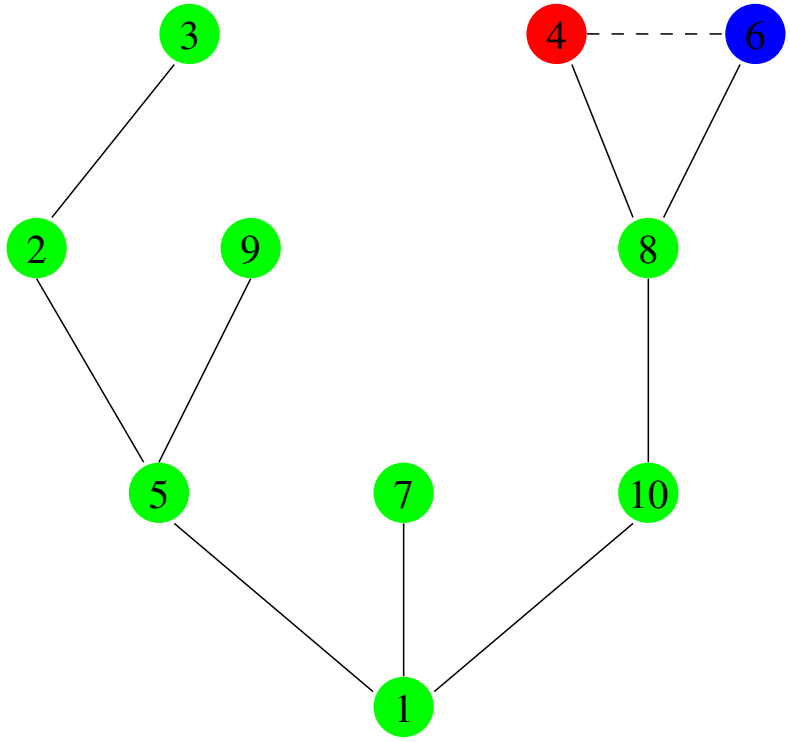
Step 6: $X(6) = 0$.

Depth-first walk and permitted edges



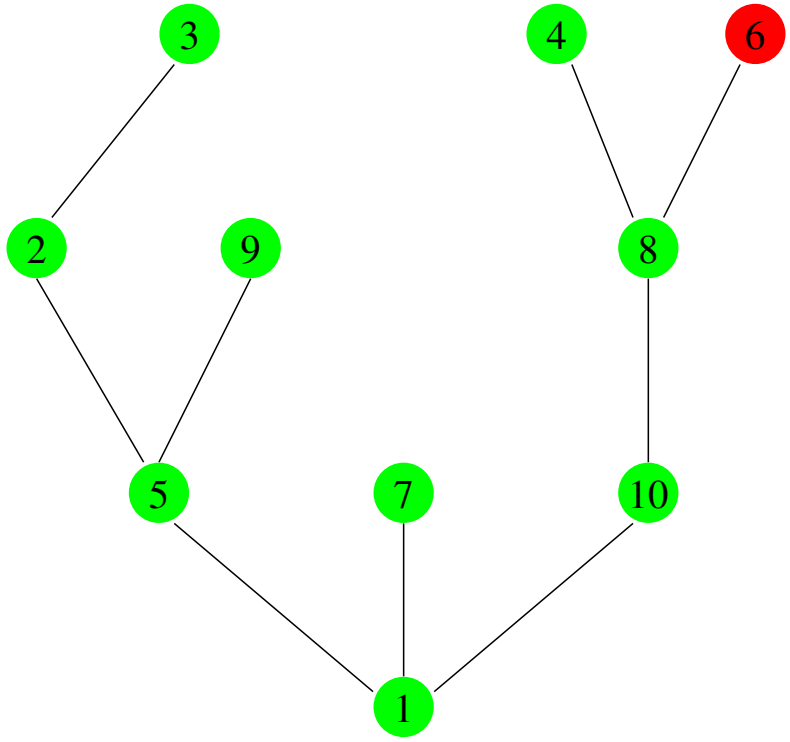
Step 7: $X(7) = 0$.

Depth-first walk and permitted edges



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

Depth-first walk and permitted edges



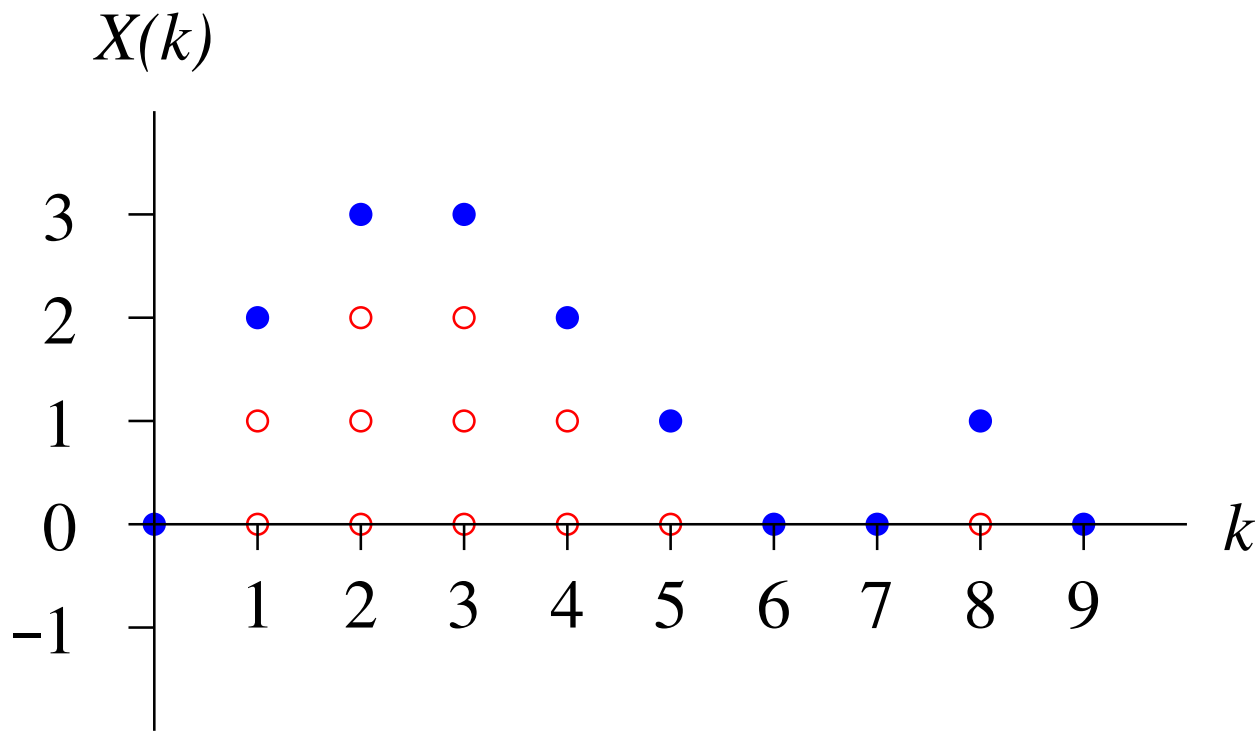
Step 10: $X(9) = 0$.

Area

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

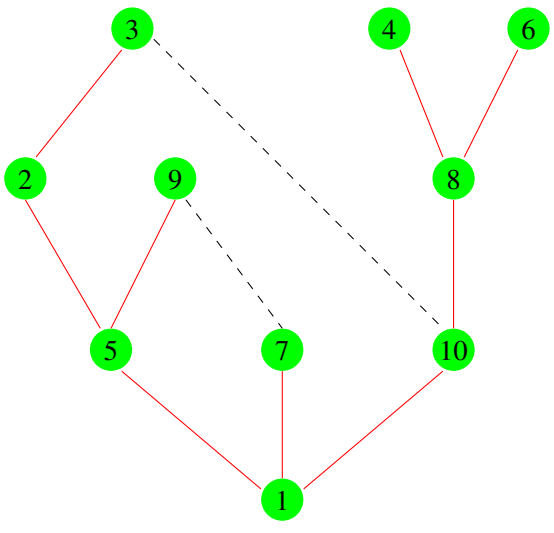
$$a(T) = \sum_{k=0}^{m-1} X(k).$$

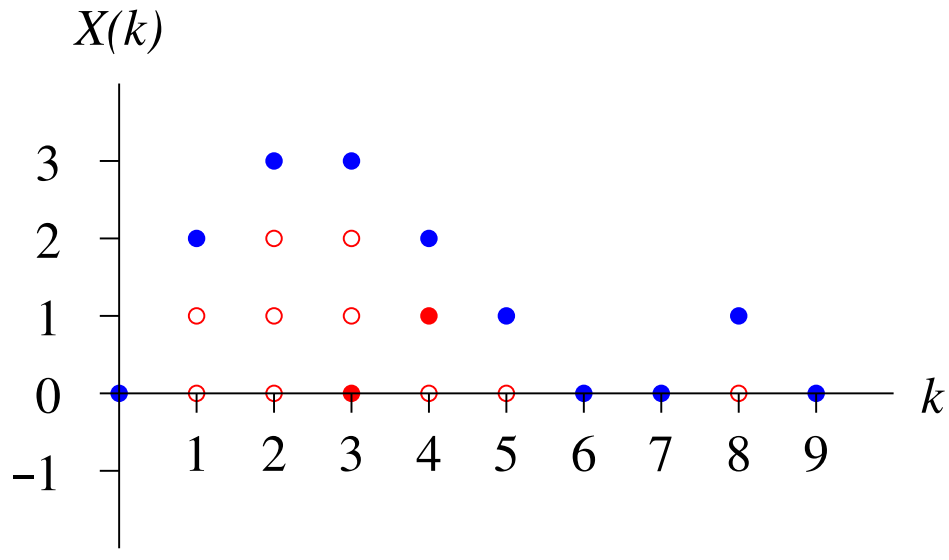
We call this the area of T.



Surplus edges

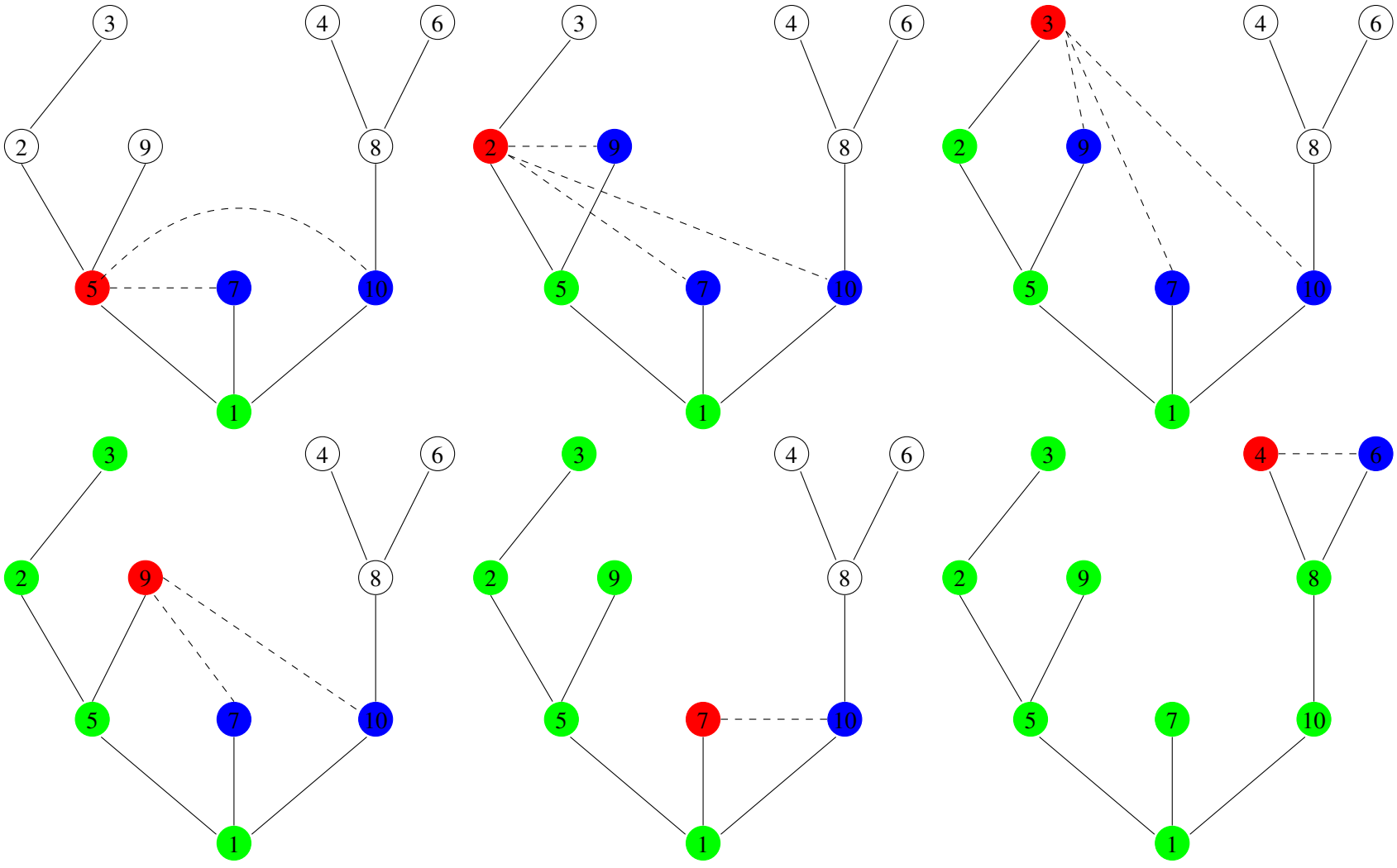
We can easily set up a bijection between the permitted edges and the integer points under the graph of the depth-first walk. For our original graph, the two surplus edges correspond to the filled red points below:





Surplus edges

Notice also that surplus edges almost go to ancestors: in fact, the permitted edges all go to younger children of ancestors of the current vertex.



Classifying graphs by depth-first tree

Let $\mathbb{T}_{[m]}$ be the set of trees with label-set $[m] = \{1, 2, \ldots, m\}$ and let $\mathbb{G}_T$ be the set of graphs G such that $T(G) = T$.

Classifying graphs by depth-first tree

Let $\mathbb{T}_{[m]}$ be the set of trees with label-set $[m] = \{1, 2, \ldots, m\}$ and let $\mathbb{G}_T$ be the set of graphs G such that $T(G) = T$. Then

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

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

Classifying graphs by depth-first tree

Let $\mathbb{T}_{[m]}$ be the set of trees with label-set $[m] = \{1, 2, \ldots, m\}$ and let $\mathbb{G}_T$ be the set of graphs G such that $T(G) = T$. Then

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

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

Moreover, $|\mathbb{G}_T| = 2^{a(T)}$.

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 $G(n,p)$,

$$\begin{aligned}\mathbb{P}\left(G_m^p = G\right) &\propto \mathbb{P}\left(G(m,p) = G\right) = p^{m+s-1}(1-p)^{\binom{m}{2}-s+1} \\ &\propto (p/(1-p))^s.\end{aligned}$$

□

Taking limits

When $m \sim xn^{2/3}$ and $p = n^{-1} + \lambda n^{-4/3}$, we need to show that

- the tree $\tilde{T}_m^p$ converges to a CRT coded by a tilted excursion;
- the locations of the surplus edges converge to the locations in our limiting picture.

Taking limits

When $m \sim xn^{2/3}$ and $p = n^{-1} + \lambda n^{-4/3}$, we need to show that

- the tree $\tilde{T}_m^p$ converges to a CRT coded by a tilted excursion;
- the locations of the surplus edges converge to the locations in our limiting picture.

For simplicity, take $p = m^{-3/2}$ (i.e. $x = 1$, $\lambda = 0$); the general case is similar.

Taking limits

Write $\tilde{X}^m$ for the depth-first walk associated with $\tilde{T}_m^p$. Then

$$a\left(\tilde{T}_m^p\right) = \int_0^m \tilde{X}^m(s)ds.$$

Taking limits

Write $\tilde{X}^m$ for the depth-first walk associated with $\tilde{T}_m^p$. Then

$$a\left(\tilde{T}_m^p\right) = \int_0^m \tilde{X}^m(s)ds.$$

If T_m is a uniform random tree on $[m]$ and X^m is its depth-first walk, then

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

Taking limits

Use the change of measure to get from $\tilde{X}^m$ to X^m: for any bounded continuous function f,

$$\mathbb{E}\left[f\left(m^{-1/2}\tilde{X}^m(mt), 0 \le t \le 1\right)\right]$$

$$= \frac{\mathbb{E}\left[f\left(m^{-1/2}X^m(mt), 0 \le t \le 1\right)(1-p)^{-\int_0^m X^m(u)du}\right]}{\mathbb{E}\left[(1-p)^{-\int_0^m X^m(u)du}\right]}$$

Taking limits

Use the change of measure to get from $\tilde{X}^m$ to X^m: for any bounded continuous function f,

$$\mathbb{E}\left[f\left(m^{-1/2}\tilde{X}^m(mt), 0 \le t \le 1\right)\right]$$
$$= \frac{\mathbb{E}\left[f\left(m^{-1/2}X^m(mt), 0 \le t \le 1\right)(1-p)^{-\int_0^m X^m(u)du}\right]}{\mathbb{E}\left[(1-p)^{-\int_0^m X^m(u)du}\right]}$$

Note that by a change of variable in the integral,

$$(1-p)^{-\int_0^m X^m(u)du} = (1-p)^{-m^{3/2}\int_0^1 m^{-1/2}X^m(ms)ds}.$$

Taking limits

Use the change of measure to get from $\tilde{X}^m$ to X^m: for any bounded continuous function f,

$$\mathbb{E}\left[f\left(m^{-1/2}\tilde{X}^m(mt), 0 \le t \le 1\right)\right]$$
$$= \frac{\mathbb{E}\left[f\left(m^{-1/2}X^m(mt), 0 \le t \le 1\right)(1-p)^{-\int_0^m X^m(u)du}\right]}{\mathbb{E}\left[(1-p)^{-\int_0^m X^m(u)du}\right]}$$

Note that by a change of variable in the integral,

$$(1-p)^{-\int_0^m X^m(u)du} = (1-p)^{-m^{3/2}\int_0^1 m^{-1/2}X^m(ms)ds}.$$

Since $p = \frac{1}{m^{3/2}}$,

$$(1-p)^{-m^{3/2}\int_0^1 m^{-1/2}X^m(ms)ds} \xrightarrow{d} \exp\left(\int_0^1 e(u)du\right).$$

Taking limits

Taking care with the limits, we obtain

$$\mathbb{E}\left[f\left(m^{-1/2}\tilde{X}^m(mt), 0 \leq t \leq 1\right)\right] \rightarrow \frac{\mathbb{E}\left[f(e)\exp\left(\int_0^1 e(u)du\right)\right]}{\mathbb{E}\left[\exp\left(\int_0^1 e(u)du\right)\right]} = \mathbb{E}\left[f(\tilde{e})\right].$$

Taking limits

Taking care with the limits, we obtain

$$\mathbb{E}\left[f\left(m^{-1/2}\tilde{X}^m(mt), 0 \le t \le 1\right)\right] \to \frac{\mathbb{E}\left[f(e)\exp\left(\int_0^1 e(u)du\right)\right]}{\mathbb{E}\left[\exp\left(\int_0^1 e(u)du\right)\right]} = \mathbb{E}\left[f(\tilde{e})\right].$$

This (after a bit more work) entails that

$$\frac{1}{\sqrt{m}}\tilde{T}_m^p \xrightarrow{d} \tilde{\mathcal{T}}.$$

Taking limits

Taking care with the limits, we obtain

$$\mathbb{E}\left[f\left(m^{-1/2}\tilde{X}^m(mt), 0 \le t \le 1\right)\right] \to \frac{\mathbb{E}\left[f(e)\exp\left(\int_0^1 e(u)du\right)\right]}{\mathbb{E}\left[\exp\left(\int_0^1 e(u)du\right)\right]} = \mathbb{E}\left[f(\tilde{e})\right].$$

This (after a bit more work) entails that

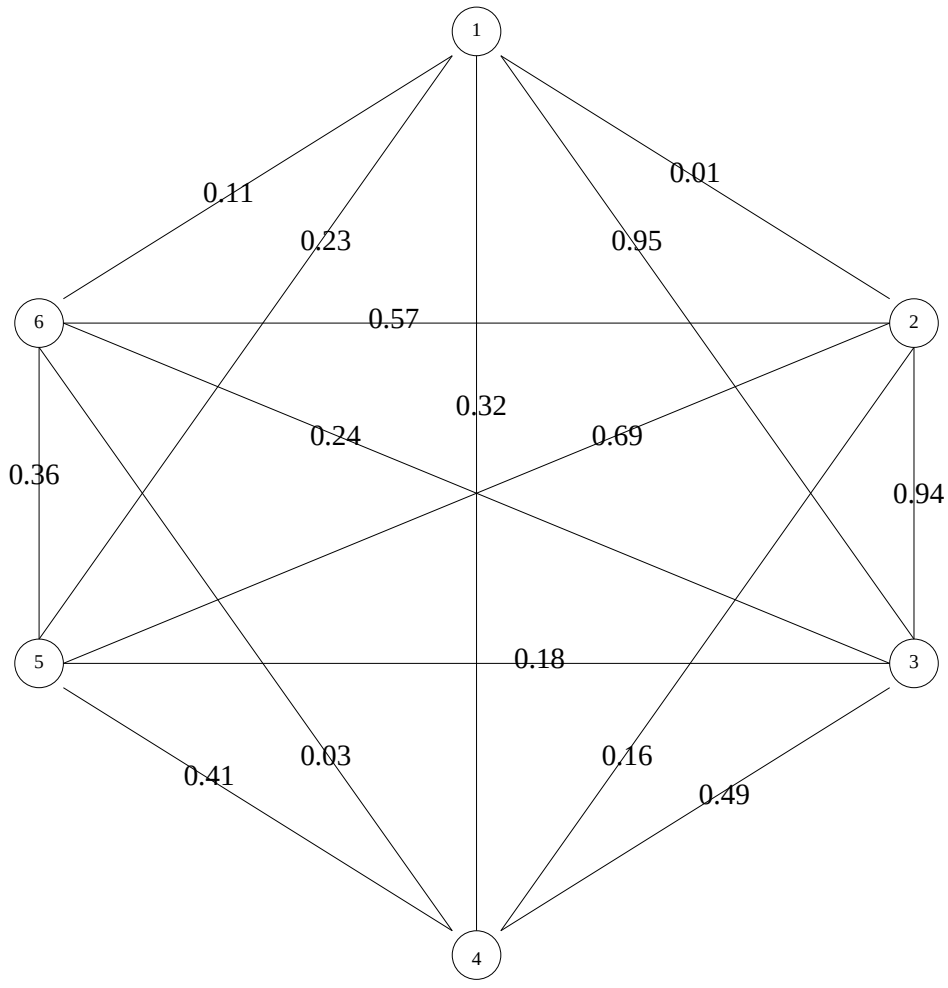
$$\frac{1}{\sqrt{m}}\tilde{T}_m^p \xrightarrow{d} \tilde{\mathcal{T}}.$$

The surplus edges form a Binomial point process under $\tilde{X}^m$ which, after rescaling, converges to a homogeneous Poisson point process under $\tilde{e}$. The rules for the locations of the surplus edges also pass to the limit to give the locations of the vertex identifications.

Part II: application to the minimum spanning tree of the complete graph

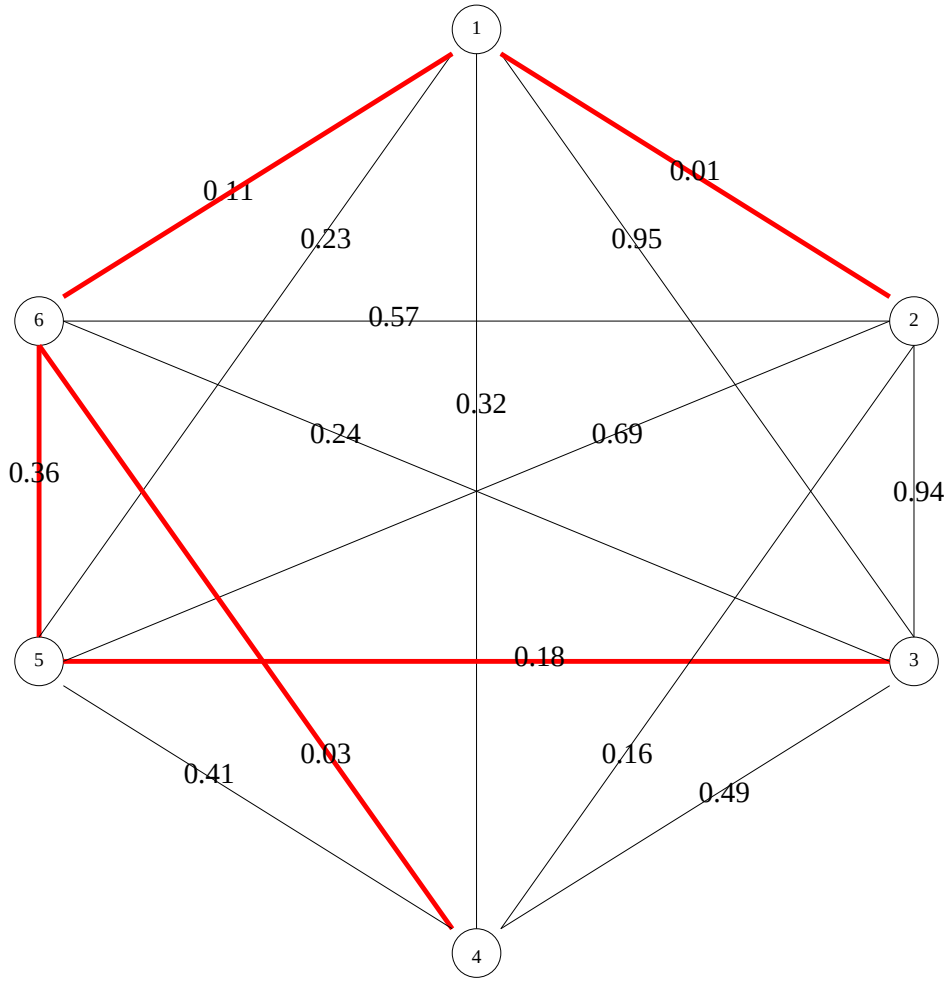
The minimum spanning tree of the complete graph

Consider the complete graph on n vertices with independent edge-weights having Uniform(0,1) distribution.



The minimum spanning tree of the complete graph

The minimum spanning tree (MST), M_n, is the spanning subtree of least weight.



Question (originally posed by David Aldous)

What does the MST of the complete graph on n vertices look like for large n?

Question (originally posed by David Aldous)

What does the MST of the complete graph on n vertices look like for large n?

In particular, does it have a scaling limit?

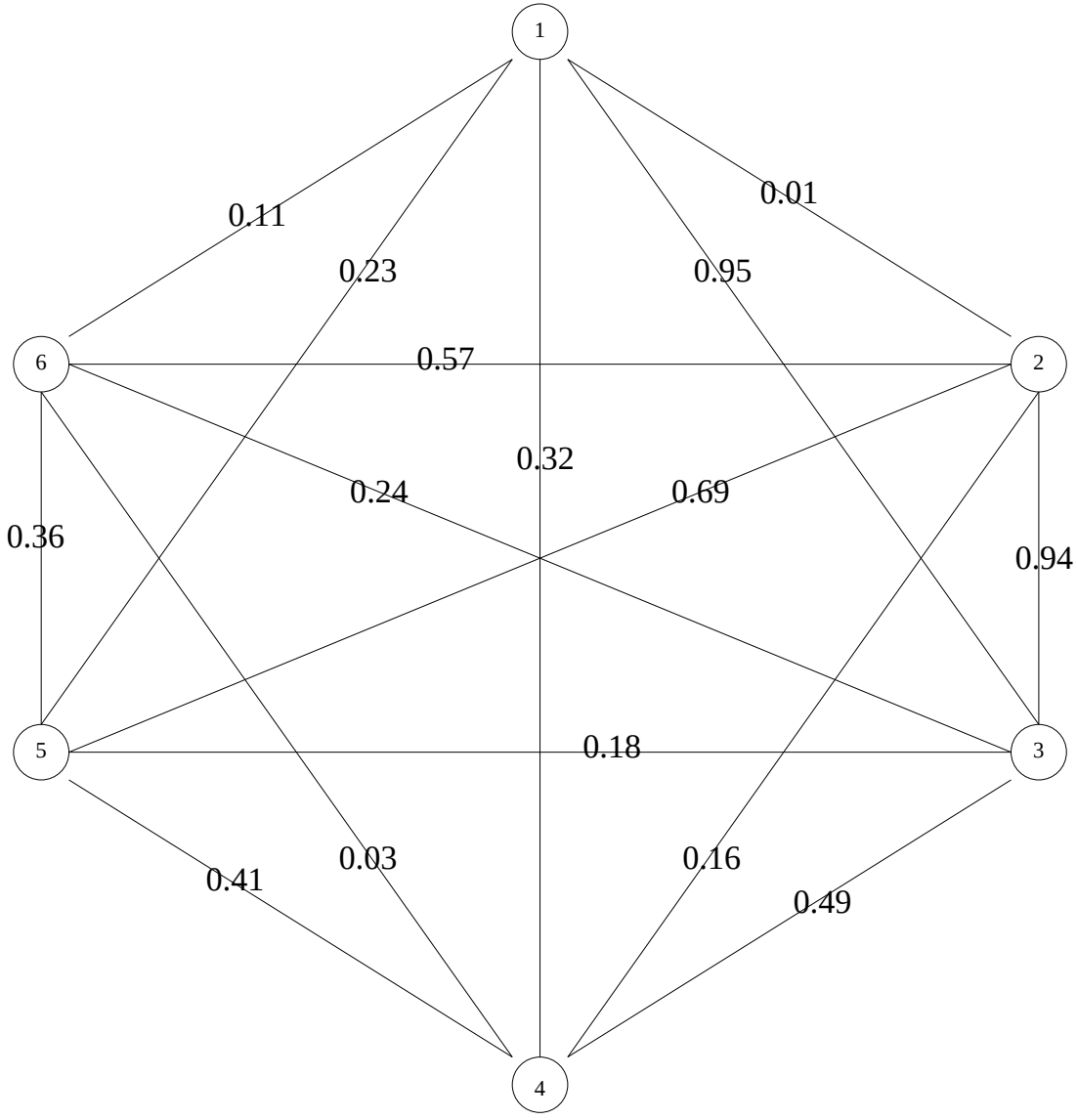
Question (originally posed by David Aldous)

What does the MST of the complete graph on n vertices look like for large n?

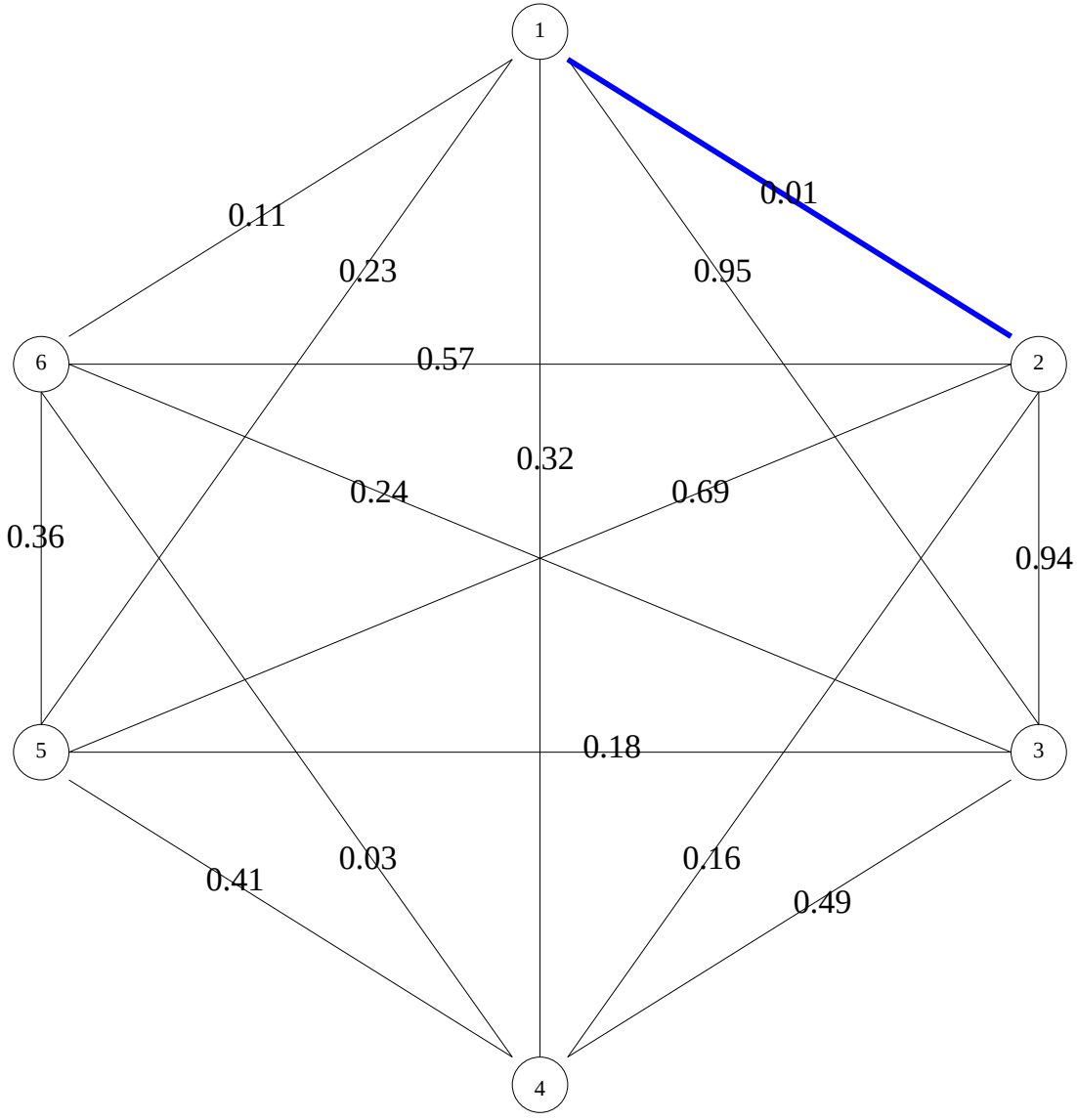
In particular, does it have a scaling limit?

We can construct M_n very easily using Kruskal's algorithm, which at each step simply adds the lowest weight edge possible, as long as it does not create a cycle.

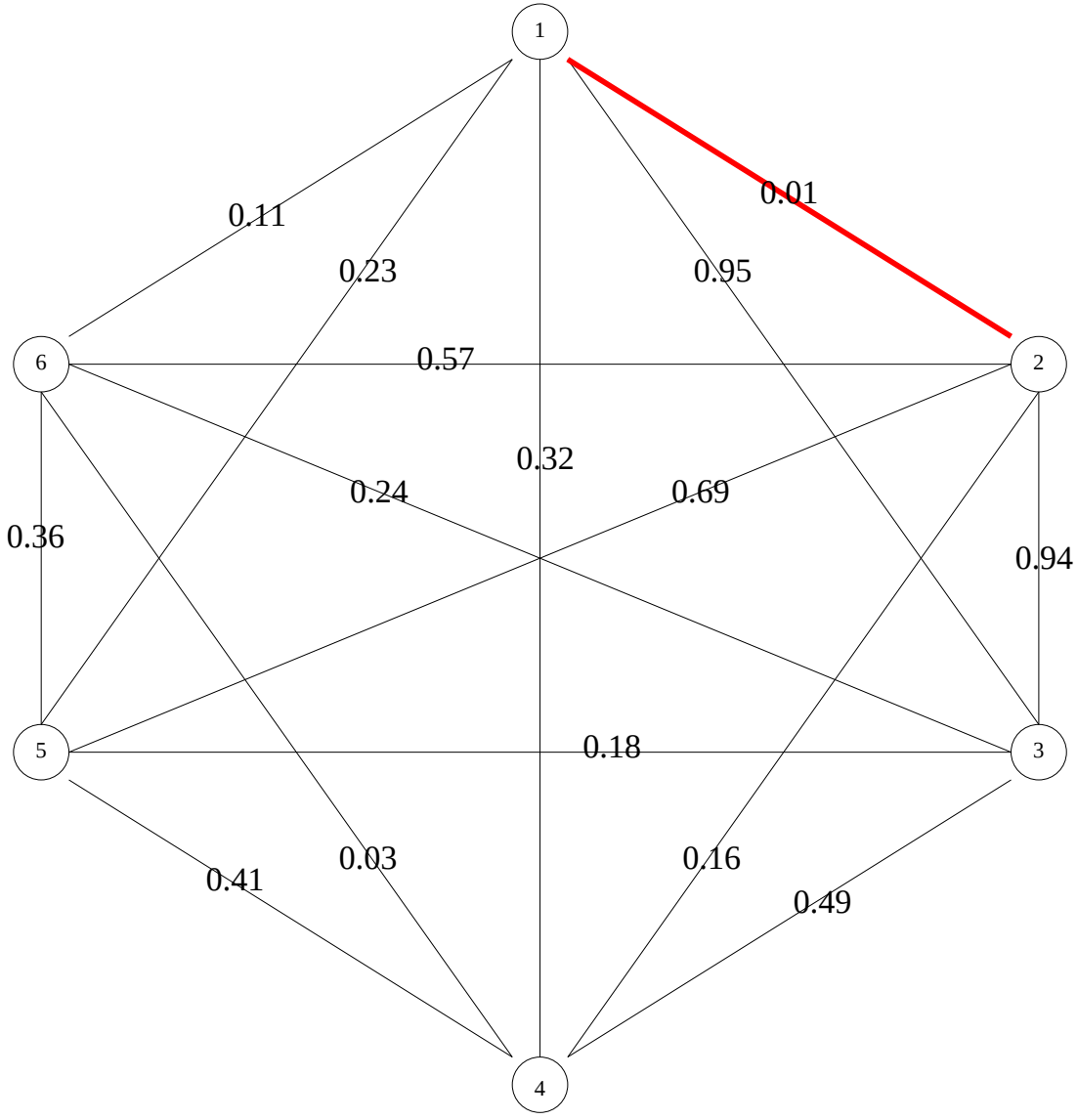
Kruskal's algorithm



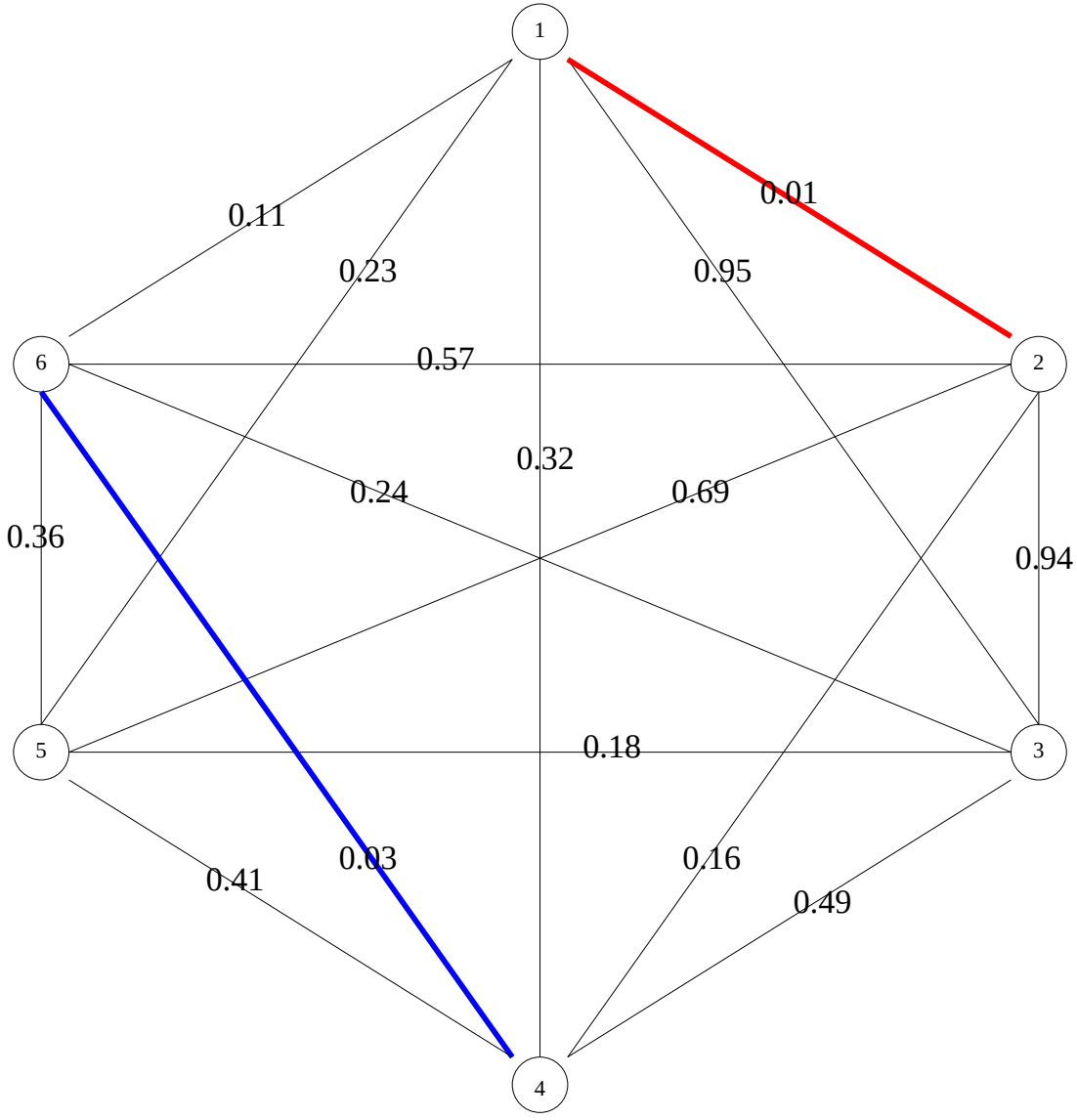
Kruskal's algorithm



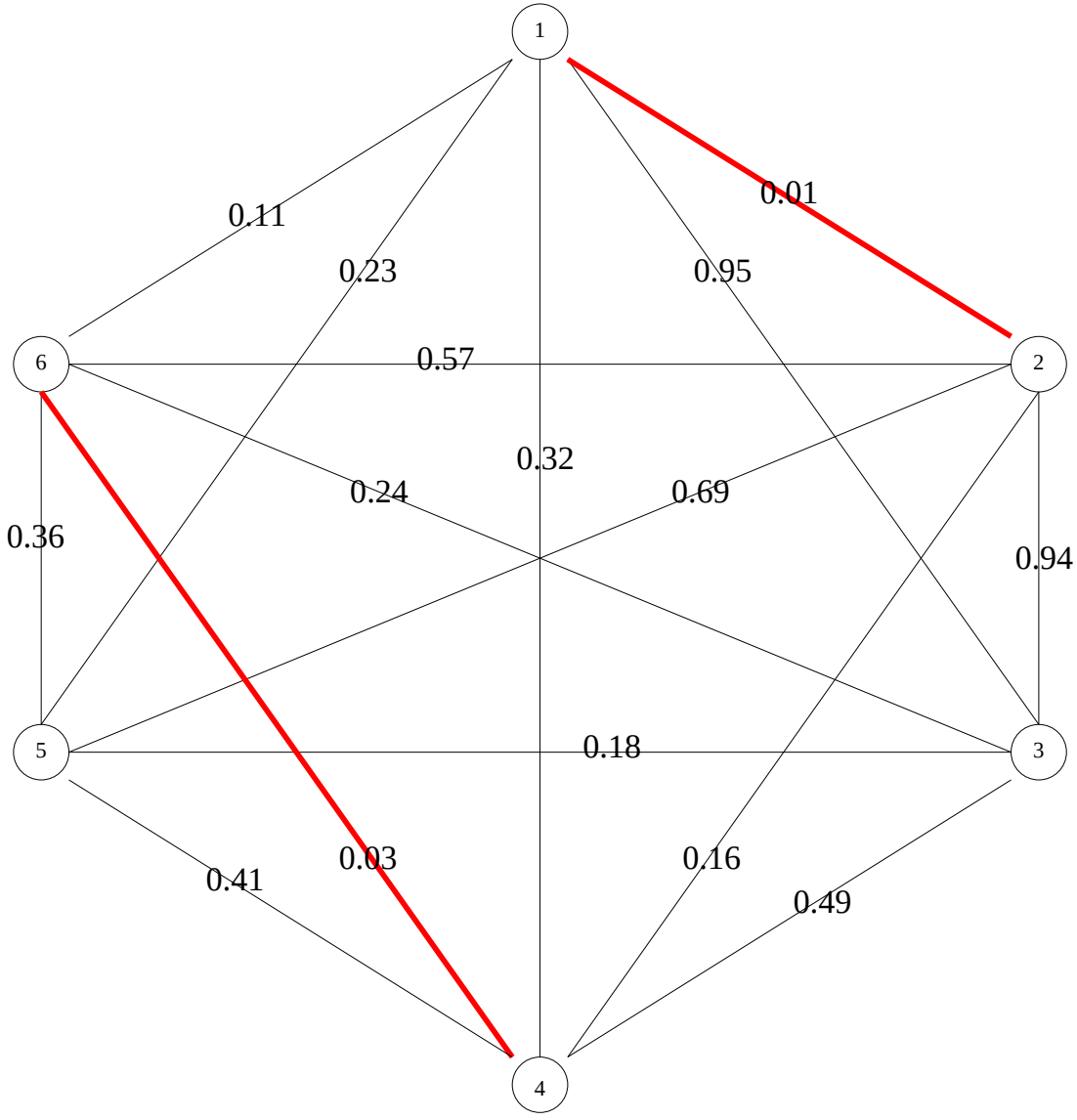
Kruskal's algorithm



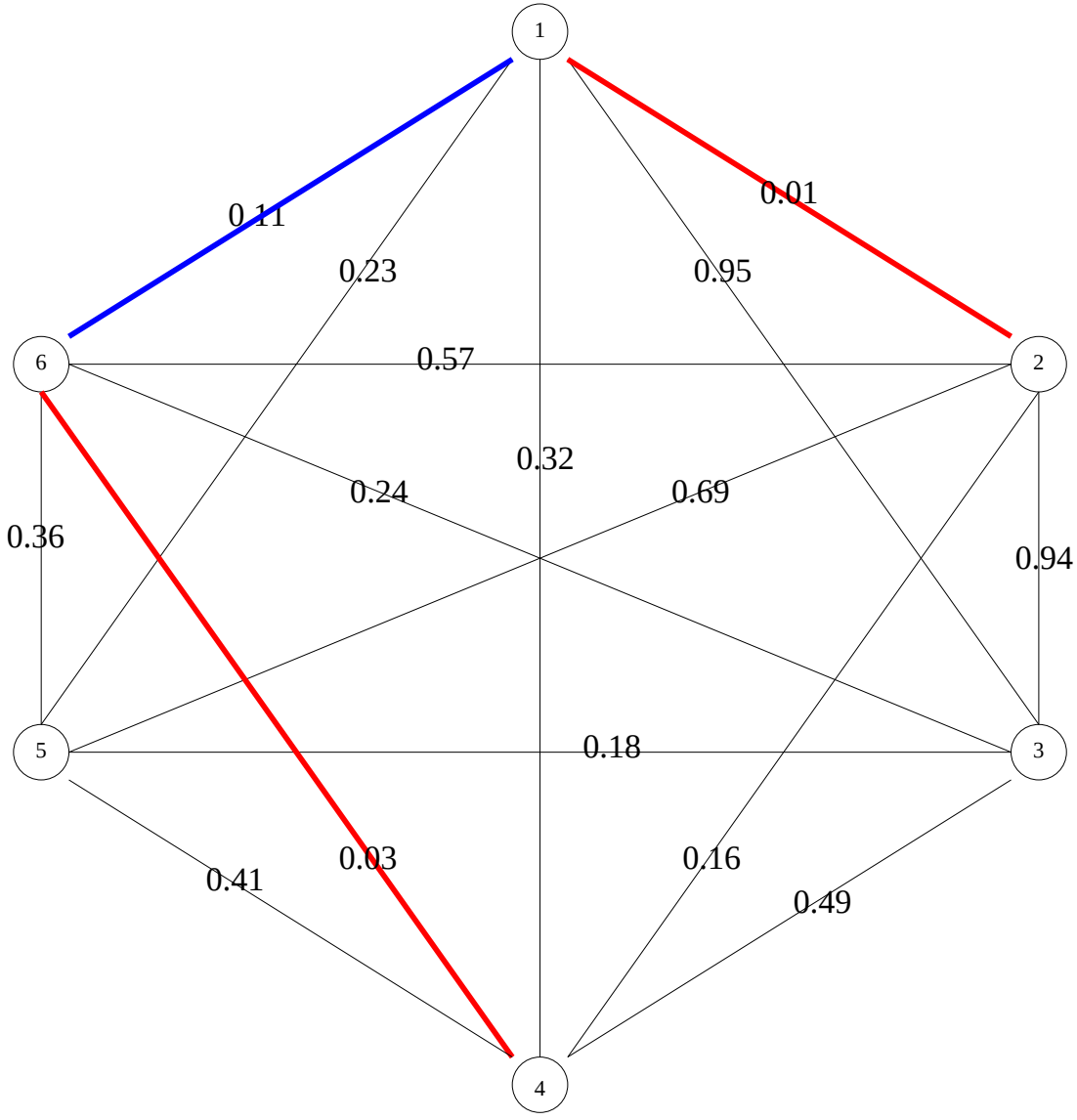
Kruskal's algorithm



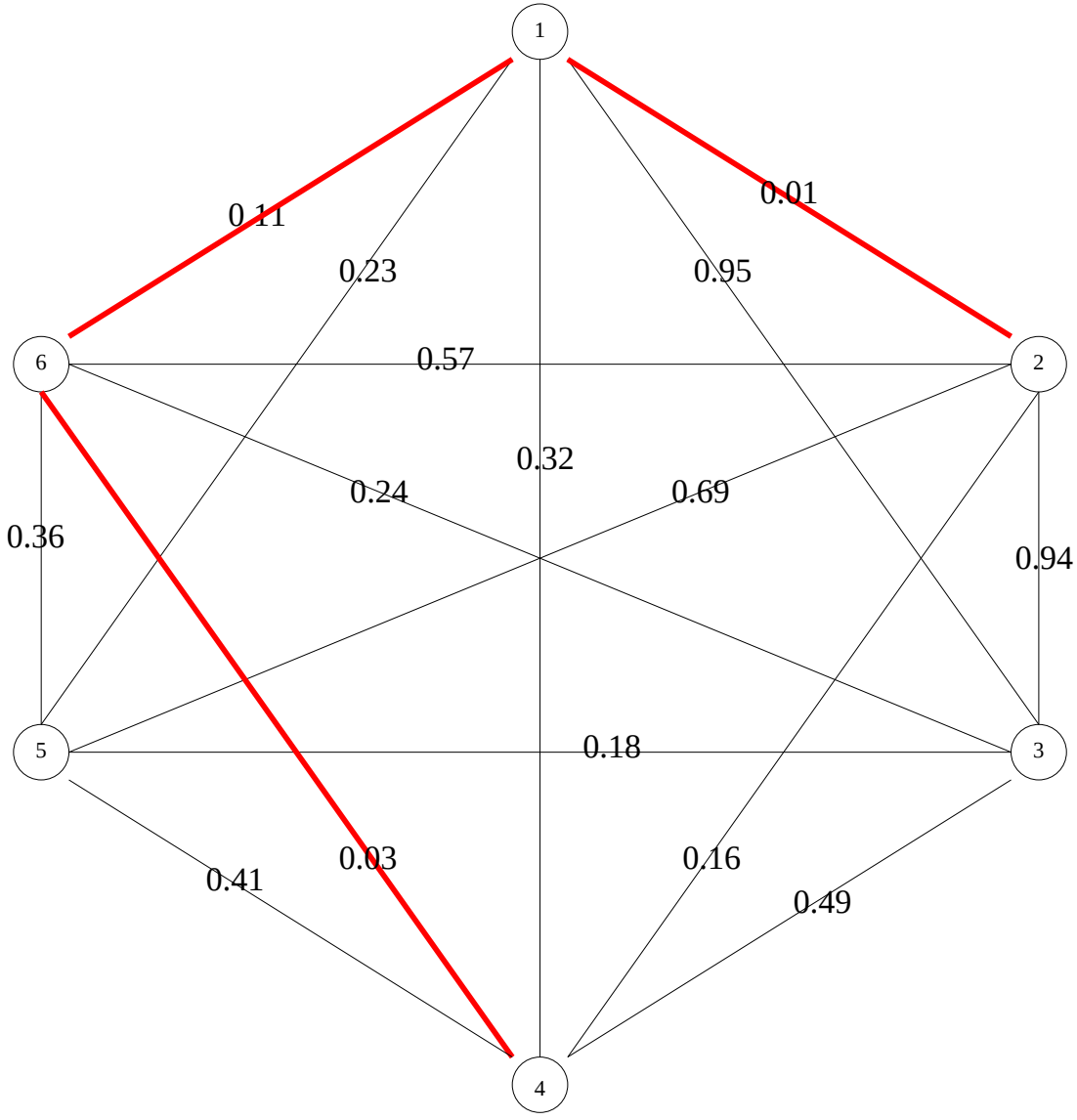
Kruskal's algorithm



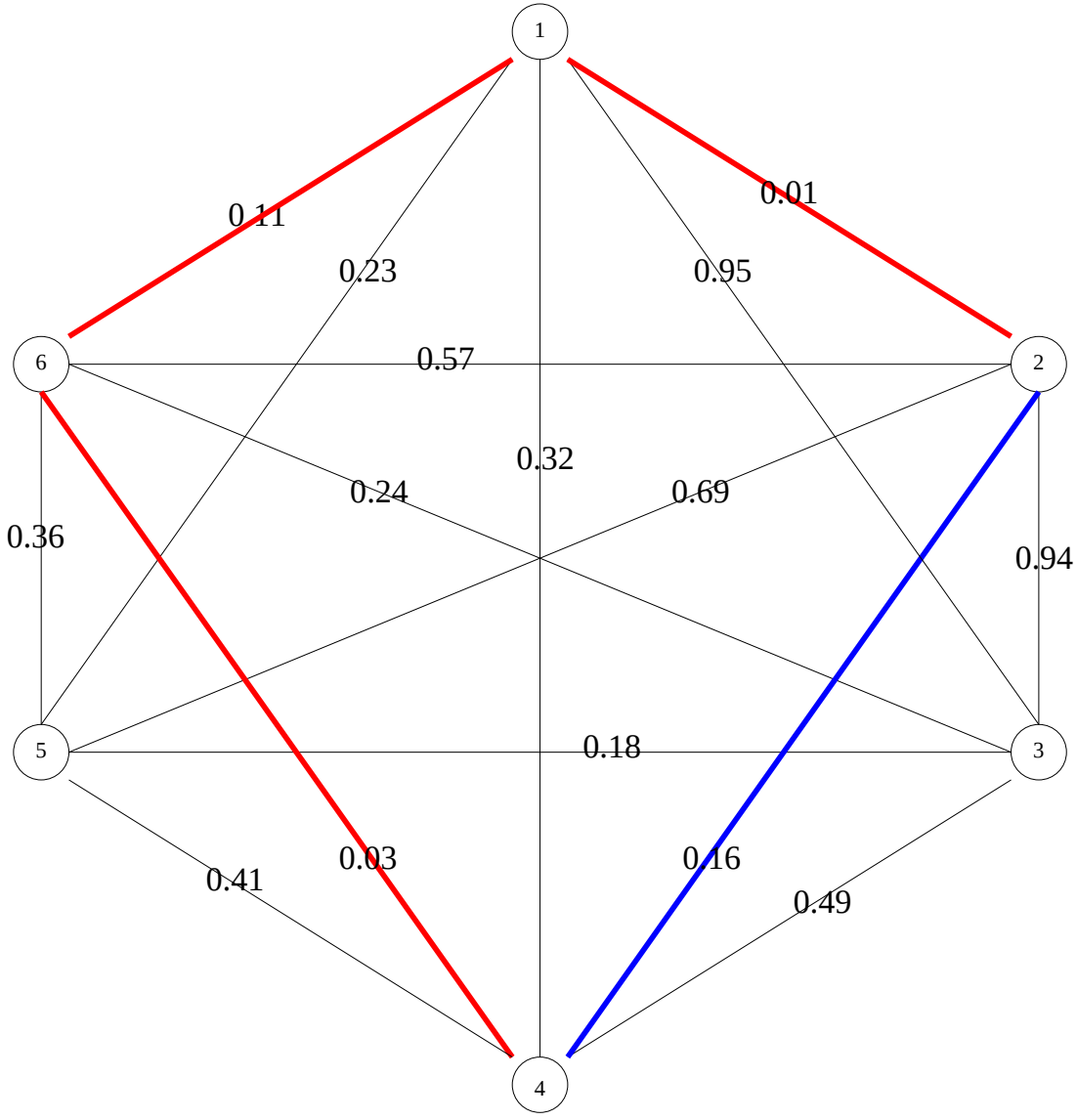
Kruskal's algorithm



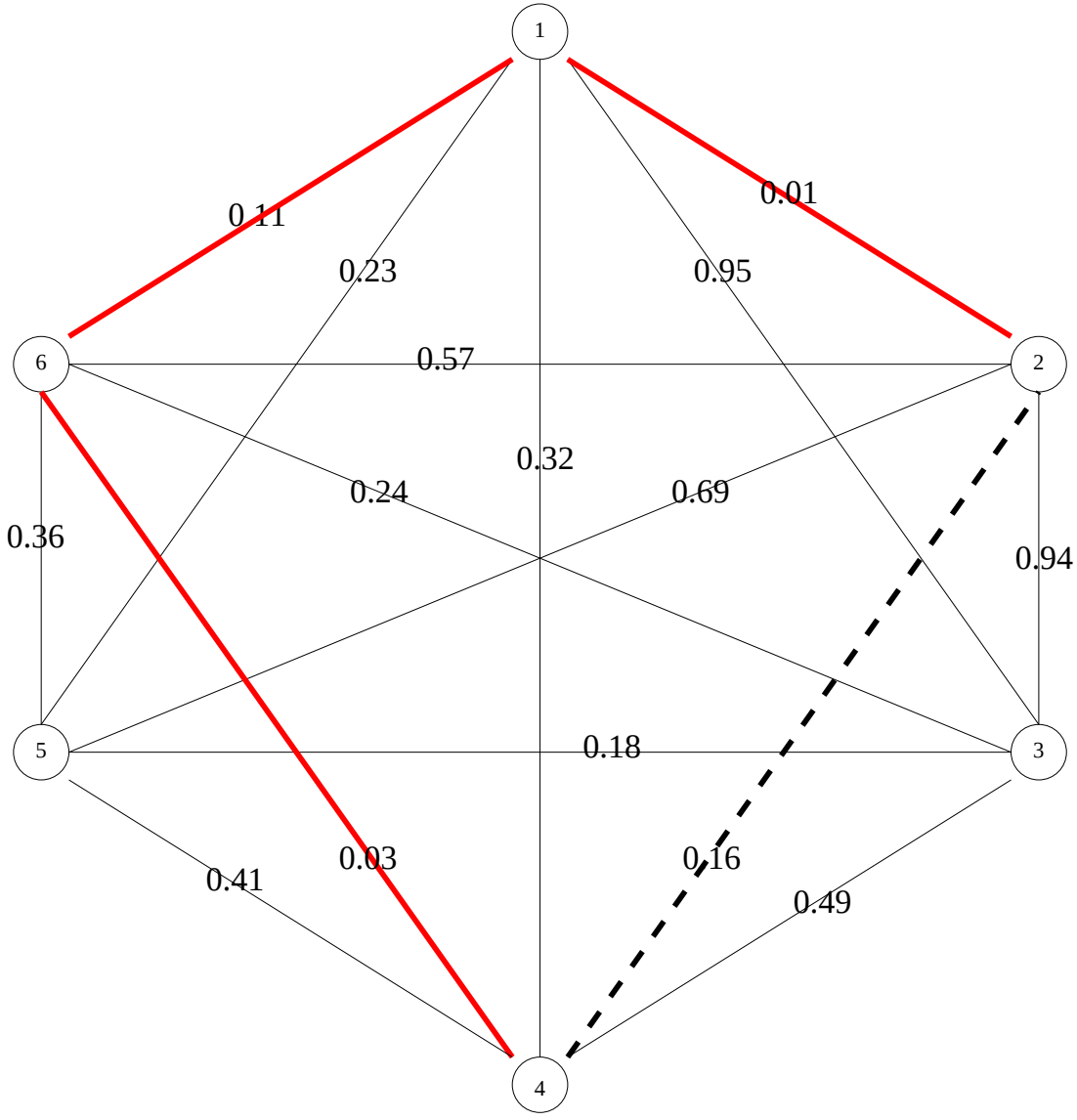
Kruskal's algorithm



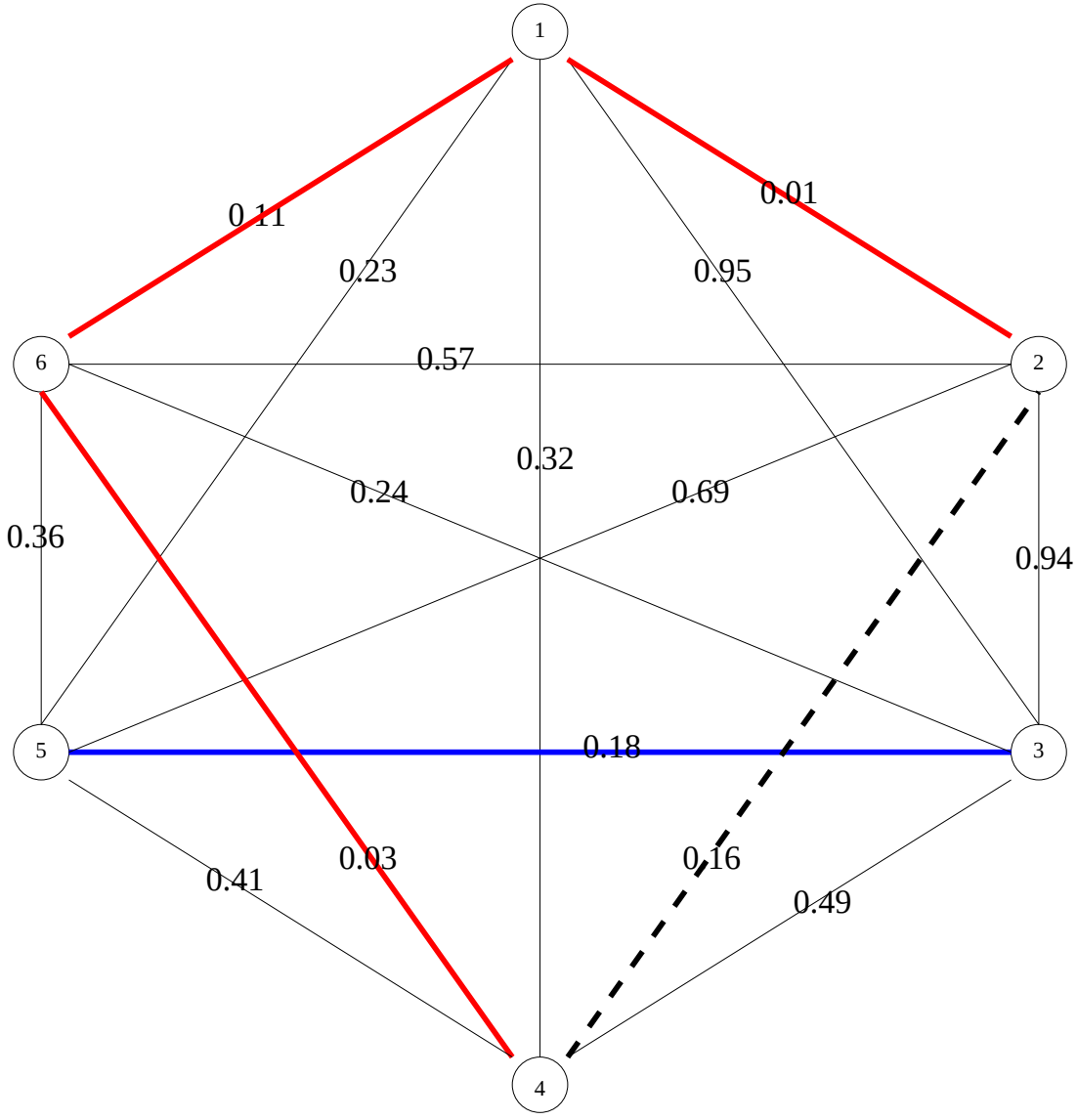
Kruskal's algorithm



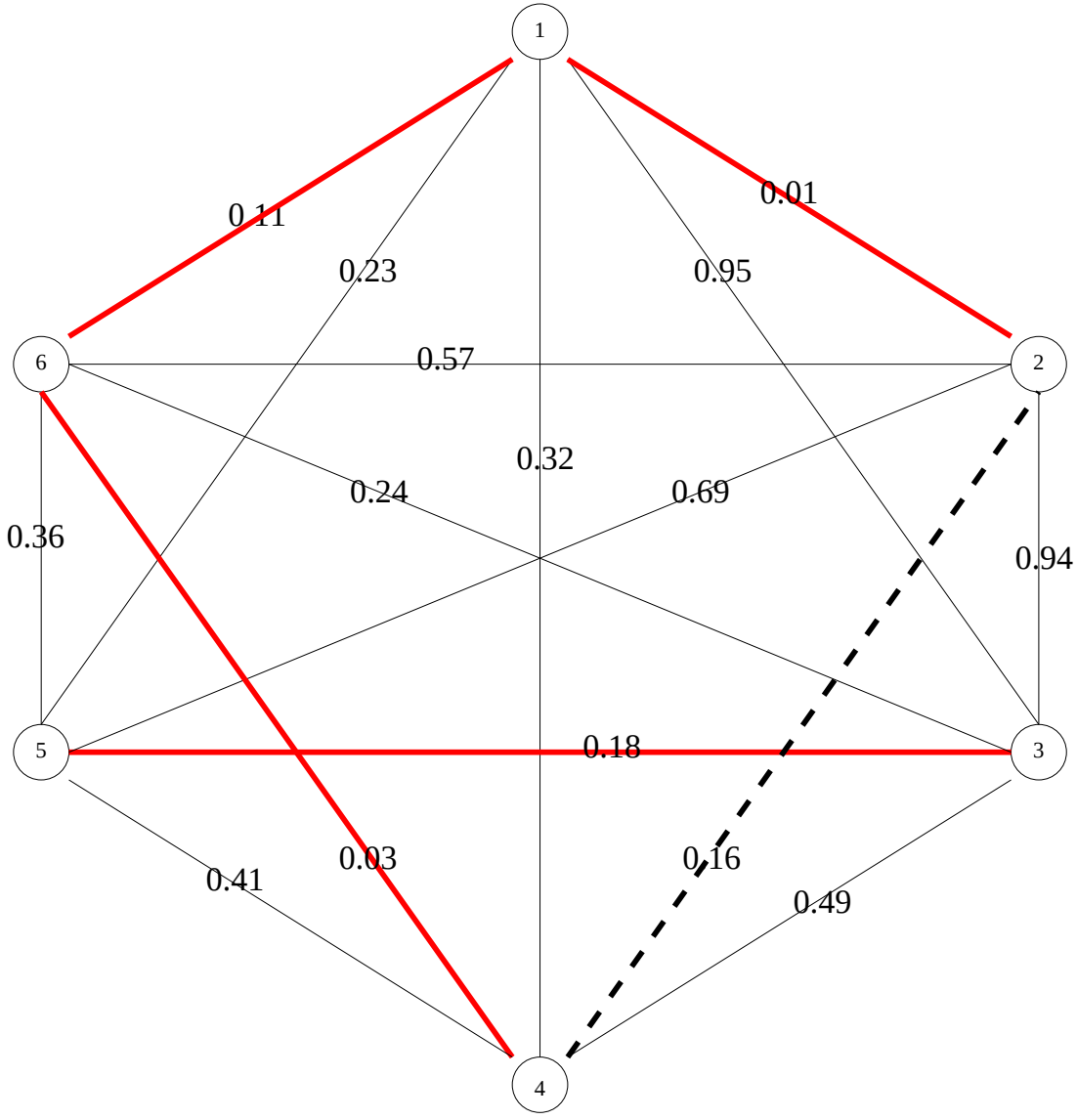
Kruskal's algorithm



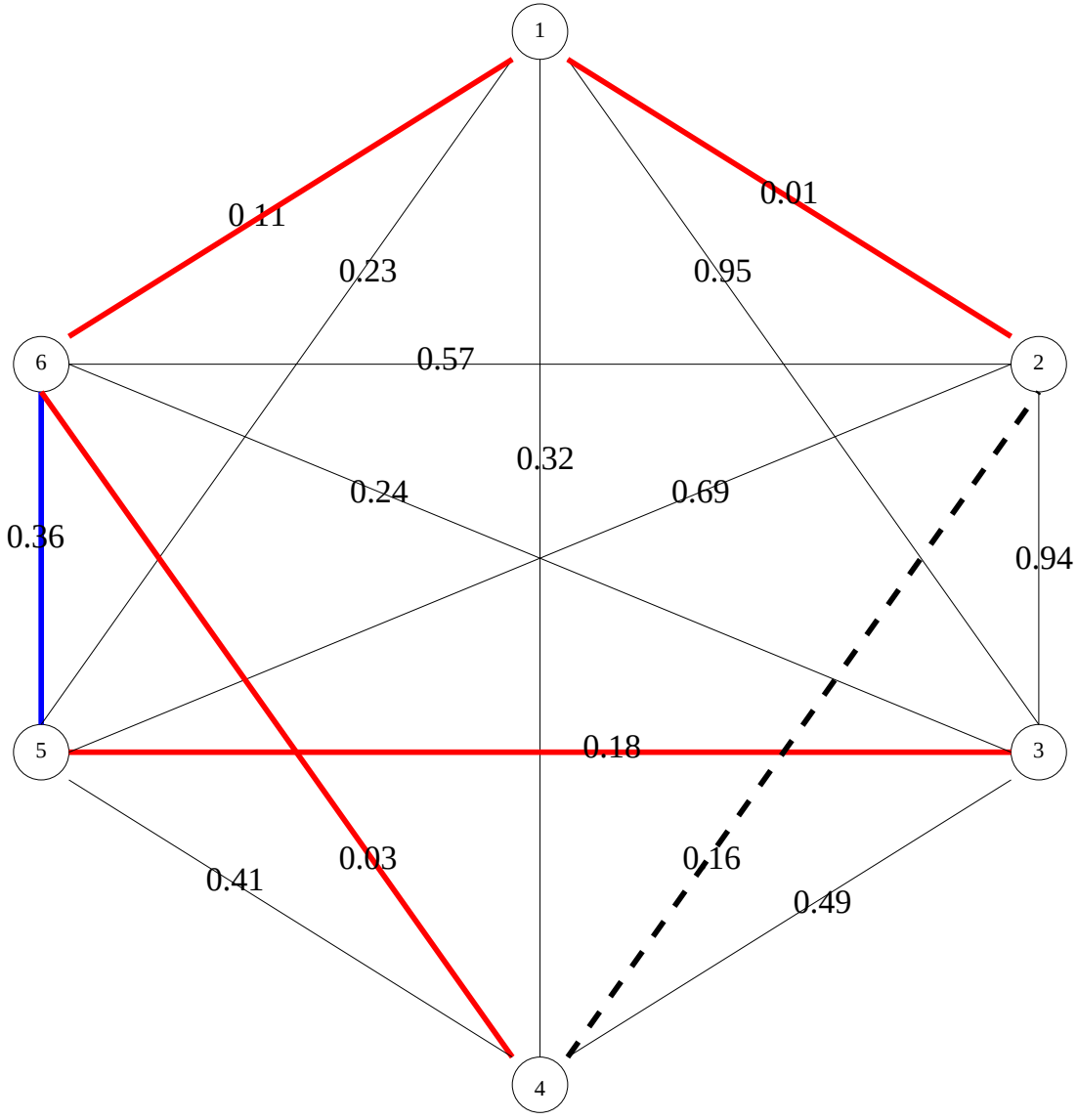
Kruskal's algorithm



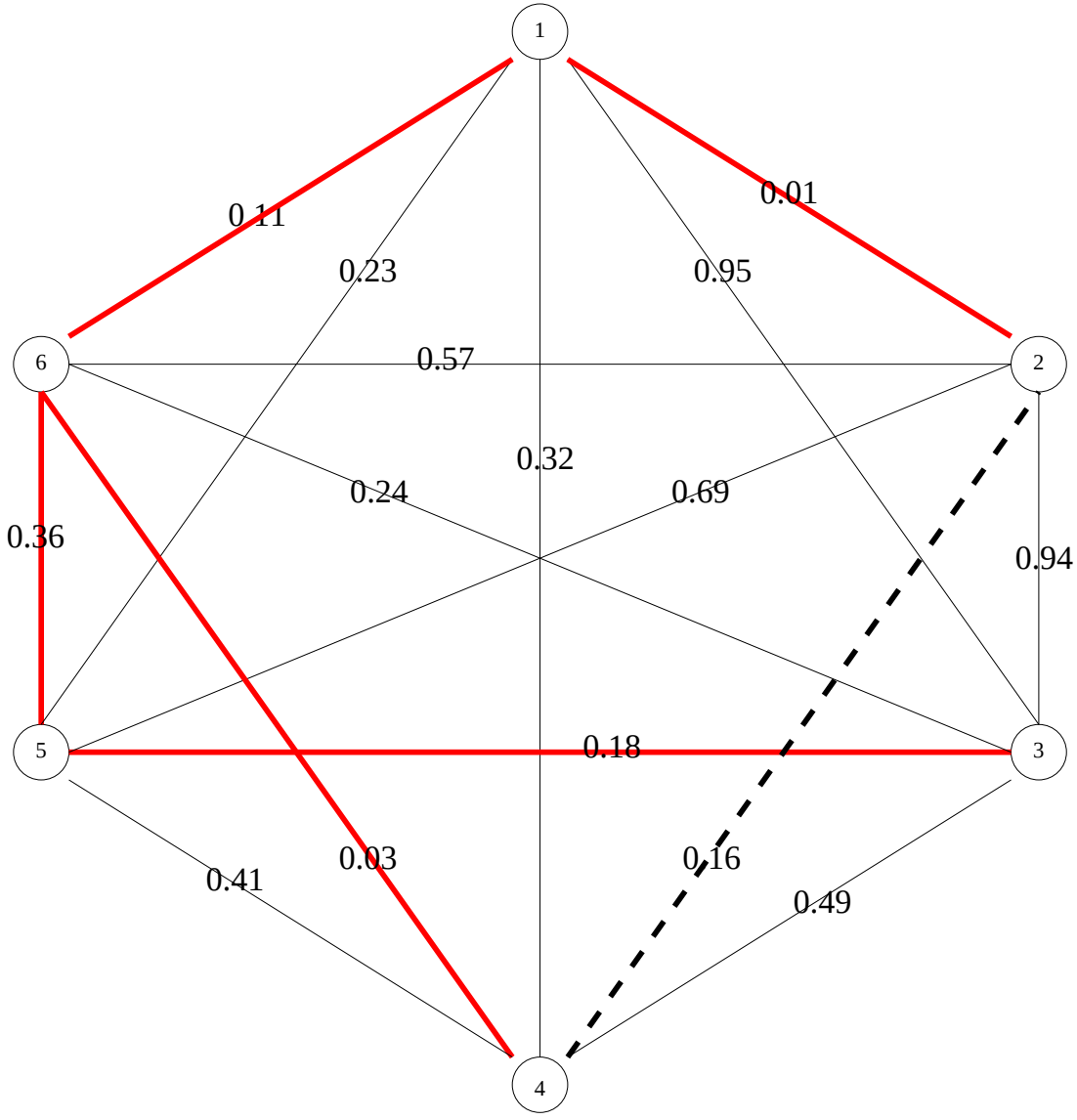
Kruskal's algorithm



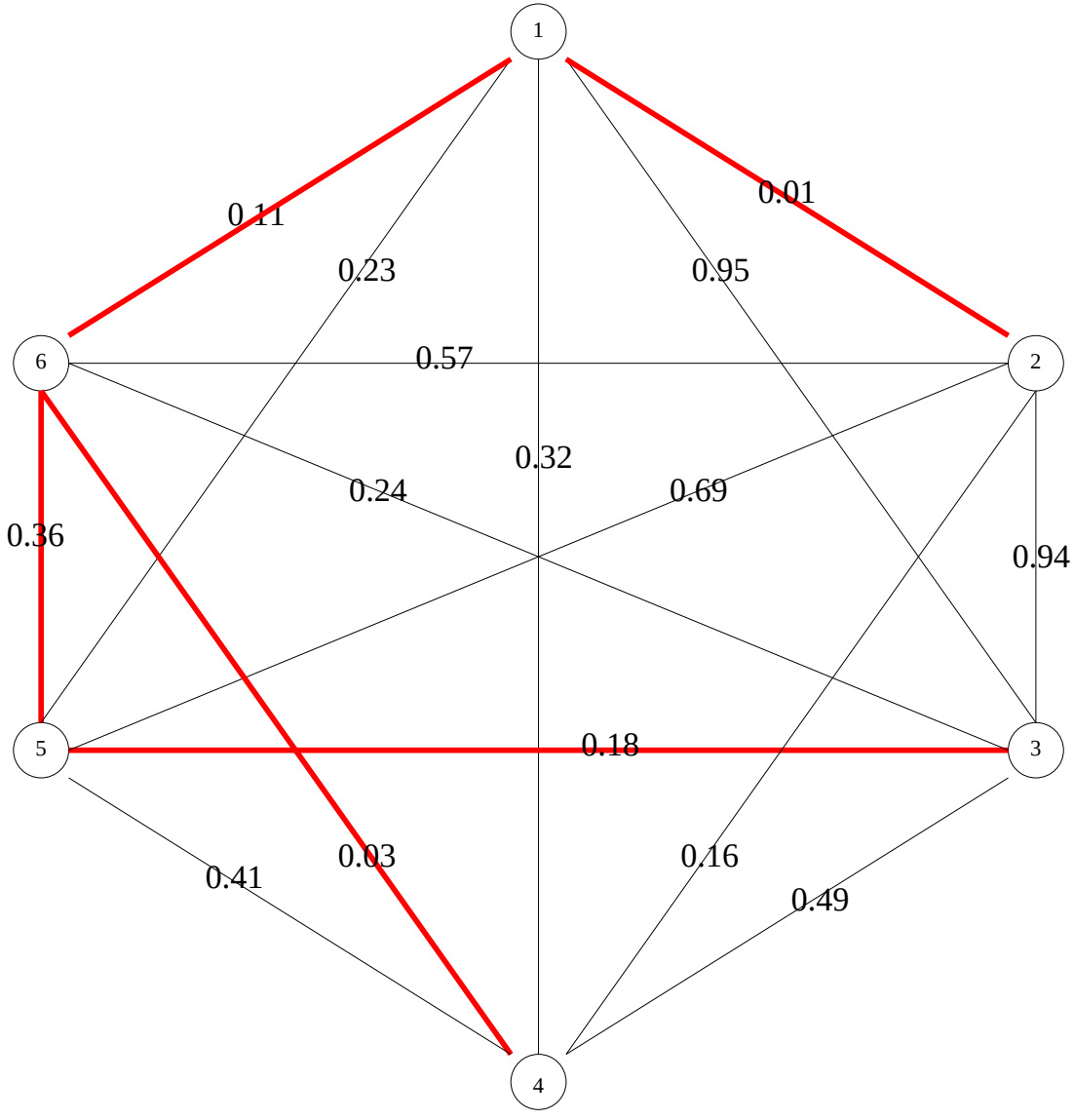
Kruskal's algorithm



Kruskal's algorithm



Kruskal's algorithm: the MST



Erdős-Rényi process

Consider a dynamic version of the Erdős-Rényi random graph, constructed as follows.

Each edge of the complete graph is assigned a Uniform(0,1) random variable, which tells us at what time that edge is added to the random graph. So at a fixed time t, the graph consists of all the edges with weight at most t. Since weights are uniform, each edge is thus present with probability t and we have a realisation of $G(n, t)$.

Call the resulting graph-valued process the Erdős-Rényi process.

Kruskal and Erdős-Rényi processes

As first observed by Addario-Berry, Broutin and Reed (2009), by using the same weights, we can think of (a continuous-time version of) Kruskal's algorithm as "sitting inside" the Erdős-Rényi process.

Call this the Kruskal process.

Kruskal and Erdős-Rényi processes

As first observed by Addario-Berry, Broutin and Reed (2009), by using the same weights, we can think of (a continuous-time version of) Kruskal's algorithm as "sitting inside" the Erdős-Rényi process.

Call this the Kruskal process.

In particular, the components in both processes (which are all trees in the case of Kruskal) have the same vertex-sets.

Kruskal and Erdős-Rényi processes

As first observed by Addario-Berry, Broutin and Reed (2009), by using the same weights, we can think of (a continuous-time version of) Kruskal's algorithm as "sitting inside" the Erdős-Rényi process.

Call this the Kruskal process.

In particular, the components in both processes (which are all trees in the case of Kruskal) have the same vertex-sets.

To get from the state of the Erdős-Rényi process at a fixed time to the state of the Kruskal process, we simply need to break the cycles in the components of the Erdős-Rényi graph at high-weight edges. (Note that the edges we remove in this cycle-breaking are *not* the same as the surplus edges!)

Kruskal and Erdős-Rényi processes

Since the component sizes are the same, the critical window is the same (i.e. times $t = n^{-1} + \lambda n^{-4/3}$, $\lambda \in \mathbb{R}$). It is, of course, natural to want to look inside the critical window, rescale the edge-lengths in both graph processes by $n^{-1/3}$ and take the limit as $n \to \infty$.

Kruskal and Erdős-Rényi processes: marginal distributions

For a fixed time $t = n^{-1} + \lambda n^{-4/3}$, we already know the limit of the Erdős-Rényi process; call it

$$(\mathcal{C}_1^\lambda, \mathcal{C}_2^\lambda, \ldots).$$

Kruskal and Erdős-Rényi processes: marginal distributions

For a fixed time $t = n^{-1} + \lambda n^{-4/3}$, we already know the limit of the Erdős-Rényi process; call it

$$(\mathcal{C}_1^\lambda, \mathcal{C}_2^\lambda, \ldots).$$

In order to obtain the limit for the Kruskal process at the same fixed time, we again just need to break the cycles. This gives us a sequence

$$(\mathcal{K}_1^\lambda, \mathcal{K}_2^\lambda, \ldots)$$

of trees.

Kruskal and Erdős-Rényi processes: marginal distributions

For a fixed time $t = n^{-1} + \lambda n^{-4/3}$, we already know the limit of the Erdős-Rényi process; call it

$$(\mathcal{C}_1^\lambda, \mathcal{C}_2^\lambda, \ldots).$$

In order to obtain the limit for the Kruskal process at the same fixed time, we again just need to break the cycles. This gives us a sequence

$$(\mathcal{K}_1^\lambda, \mathcal{K}_2^\lambda, \ldots)$$

of trees.

Each component ($\mathcal{C}_i^\lambda$ or $\mathcal{K}_i^\lambda$) has a mass (the length of the coding excursion) and comes naturally endowed with a measure which allows us to pick uniform points.

Kruskal and Erdős-Rényi processes: dynamics

The dynamics of the two finite-n processes may be (approximately) described as follows:

Kruskal and Erdős-Rényi processes: dynamics

The dynamics of the two finite-n processes may be (approximately) described as follows:

In both, for each pair of distinct components, at a rate given by the product of their masses, we put an edge between a uniform vertex in one and a uniform vertex in the other.

Kruskal and Erdős-Rényi processes: dynamics

The dynamics of the two finite-n processes may be (approximately) described as follows:

In both, for each pair of distinct components, at a rate given by the product of their masses, we put an edge between a uniform vertex in one and a uniform vertex in the other.

In addition, in the Erdős-Rényi process, at a rate given by the square of a component's mass, we pick two points uniformly at random from that component and join them by an edge.

Kruskal and Erdős-Rényi processes: dynamics

The dynamics of the two finite-n processes may be (approximately) described as follows:

In both, for each pair of distinct components, at a rate given by the product of their masses, we put an edge between a uniform vertex in one and a uniform vertex in the other.

In addition, in the Erdős-Rényi process, at a rate given by the square of a component's mass, we pick two points uniformly at random from that component and join them by an edge.

These dynamics should carry over to the limit, with vertex identifications replacing edge insertions.

Kruskal and Erdős-Rényi processes

Of course, we have to prove that there exist properly-defined Markov processes with the given limiting marginal distributions at "time" $\lambda \in \mathbb{R}$ and the right dynamics. Since our processes take values in the space of sequences of measured metric spaces, considerable care is required to do this.

Kruskal and Erdős-Rényi processes

Of course, we have to prove that there exist properly-defined Markov processes with the given limiting marginal distributions at "time" $\lambda \in \mathbb{R}$ and the right dynamics. Since our processes take values in the space of sequences of measured metric spaces, considerable care is required to do this.

Write

$$(\mathcal{C}_1^\lambda, \mathcal{C}_2^\lambda, \ldots)_{\lambda \in \mathbb{R}}$$

for the limit of the Erdős-Rényi process and

$$(\mathcal{K}_1^\lambda, \mathcal{K}_2^\lambda, \ldots)_{\lambda \in \mathbb{R}}$$

for the limit of the Kruskal process.

Back to the MST

The MST of the complete graph is obtained by running the Kruskal process to the end. But most of the "action" occurs in the critical window. In particular, we strongly believe that the metric structure of the MST has already been entirely built by the top of the critical window. So the scaling limit of the MST should essentially be the same as $\lim_{\lambda \to \infty} \mathcal{K}_1^\lambda$.

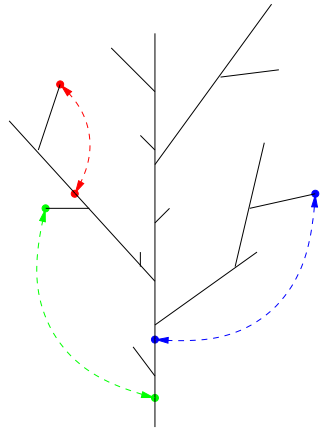
The scaling limit of the MST

Conjecture (on its way to being proved!) There exists a random compact metric space $\mathcal{M}$ such that

$$\frac{1}{n^{1/3}} M_n \xrightarrow{d} \mathcal{M}$$

as $n \to \infty$, in the sense of the Gromov-Hausdorff distance.

References



The continuum limit of critical random graphs
L. Addario-Berry, N. Broutin and C. Goldschmidt
Probability Theory and Related Fields, to appear.

Critical random graphs: limiting constructions and distributional properties
L. Addario-Berry, N. Broutin and C. Goldschmidt
Electronic Journal of Probability 15 (2010), pp.741-775.

Continuum Erdős-Rényi and Kruskal dynamics
L. Addario-Berry, N. Broutin, C. Goldschmidt and G. Miermont
in preparation.