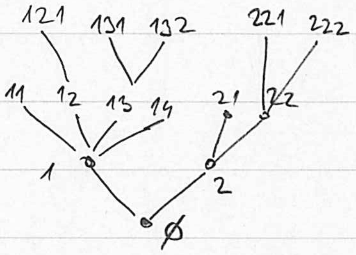
Shape of large critical GW-trees.

We now understand sizes of clusters in critical ER graph. To understand what is the typical shape of them we need to come back to GW trees.

We start by introducing another ways how to encode finite tree by a stochastic process. We work with rooted, ordered trees and label them according to their genealogy as on the figure

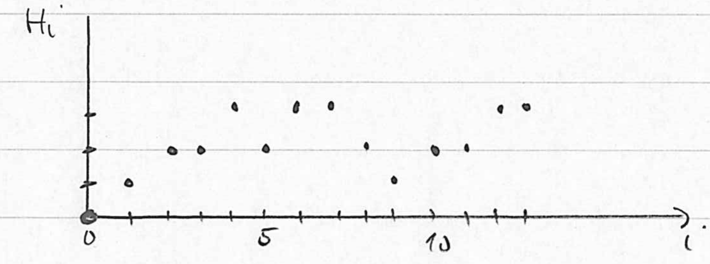


(4.51) Figure: tree with 14 vertices labeled according genealogy.

Height function: Assume the tree has n vertices. List them as $v_0, \dots, v_{n-1}$ in lexicographical order (w.r.t. their labels)

(4.52) Set $H(i) = \mathrm{dist}(v_i, \emptyset)$, $i \in \{0, \dots, n-1\}$.

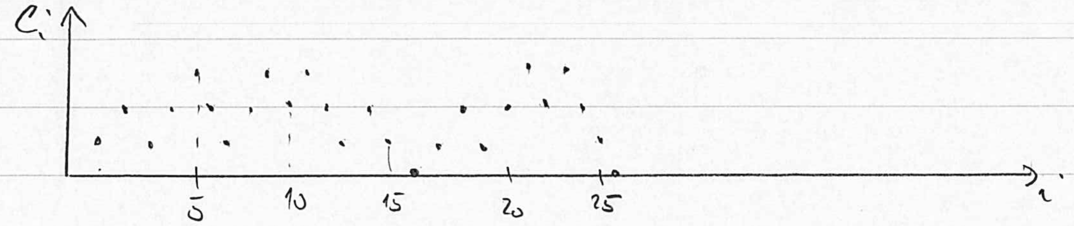
(4.53) Exercise: Show that the tree can be recovered from its height function



(4.54) Figure: The hight function of the tree from (4.51)

Contour function: Trace the "contour" of the tree from left to right, so that we pass every edge twice and record the distance from the root. We get

$$C(i) : i \in \{0, \dots, 2(n-1)\}.$$



(4.55) Figure: The contour function of (4.51)

(4.56) Exercise (easy): Show that the contour function determines the tree.

Depth-first walk (Łukasiewicz path)

Let $v_0, \dots, v_{n-1}$ be as above, and set $X_i = \#$children of v_i. Set

$$S(0) = 0, \quad S(i) = \sum_{j=0}^{i-1} (X_j - 1) \qquad 1 \le i \le n.$$

This is the exploration path considered previously.

It is less clear that S encodes the shape of the tree. We show

(4.57) Lemma: For every finite tree

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

Proof: By the same arguments as on page 10, for any subtree of the original tree, the value of X once we have finished exploring it is one less than its value when we visited the root of this subtree, whereas within the subtree S has at least value of its root. Now $\mathrm{dist}(v_i, \emptyset)$ is equal to the number of subtrees we begun but not completed exploring before reaching v_i. The roots of these subtrees are times $j < i$ with $S(j) = \min\{S(k) : j \le k \le i\}$ □

Recall that if T is a GW tree, then S is a random walk stopped at hitting (-1). Sometimes it is technically easier to deal with a sequence of iid GW trees and concatenate their height processes, resp depth-first walks. As on p. 34, the first tree ends when S hits (-1), the second when it hits (-2), etc.

(4.58) Exercise: Check that (4.57) holds also in this case.

We now explore critical GW trees. We assume that

(4.59) $$EX_i = 1 \quad \text{and} \quad \operatorname{Var} X_i = \sigma^2 < \infty$$

It is then obvious that

(4.60) Proposition (convergence of the depth first walk). Under (4.59)

$$\left(\frac{1}{\sqrt{n}} S(\lfloor nt \rfloor) : t \geq 0\right) \xrightarrow[n\to\infty]{d} (\sigma W(t) : t \geq 0)$$

where W is a standard Brownian motion.

(4.61) It is much more delicate to show that

(4.62) Theorem: (convergence of the height function) Under (4.59)

$$\left(\frac{1}{\sqrt{n}} H(\lfloor nt \rfloor), t \geq 0\right) \xrightarrow[n\to\infty]{d} \left(\frac{2}{\sigma} B(t), t \geq 0\right)$$

where $B(t) = W(t) - \inf_{s \leq t} W(s)$ is a reflected BM.

Each excursion of H corresponds to one tree in "GW forest"; the length of this excursion is the total progeny of this tree. If one conditions on the total progeny being n, let $n \to \infty$, we should obtain an excursion of the limit process. This is indeed true but we will not show it.

(4.63) Theorem (Aldous 1991) Under (4.59), the height process of a critical GW tree conditioned to have total progeny n, satisfies

$$\left(\frac{1}{\sqrt{n}} H(\lfloor nt \rfloor), t \in [0,1]\right) \xrightarrow{n\to\infty} \frac{2}{\sigma} \left(e(t) : 0 \leq t \leq 1\right),$$

where e is a standard Brownian excursion.

(4.64) <u>Remark</u>: The law of e can formally be obtained by conditioning the standard BM W on $\{W(1)=1, W(s)>0\ \forall s\in(0,1)\}$. As this event has probability 0, this is slightly involving, see e.g. Revuz-Yor, Chapter $\overline{XII}$. Otherwise we can obtain e by conditioning a SRW S on $\{S_i>0\ i\le n, S_n=0\}$ letting $n\to\infty$ and rescaling as in (4.60).

Actually more is true

(4.65) <u>Theorem</u>. (Marckert - Mokkadem (2003)). Conditioned on the total progeny n, for critical GW tree s.t. (4.59)

$$\left(n^{-1/2} S(\lfloor n\cdot\rfloor), n^{-1/2} H(\lfloor n\cdot\rfloor), n^{-1/2} C(\lfloor 2n\cdot\rfloor)\right) \xrightarrow[d]{n\to\infty} \left(\sigma e, \frac{2}{\sigma} e, \frac{2}{\sigma} e\right).$$

This suggests the existence of a 'limit tree' coded by Brownian excursion. We now construct it.

(4.66) <u>Def</u>: A compact metric space $(\mathcal{T}, d)$ is a real tree if for all $x,y\in\mathcal{T}$

(a) exists a unique shortest path γ_{xy} from x to y (and $\text{length}(\gamma_{xy}) = d(x,y)$)

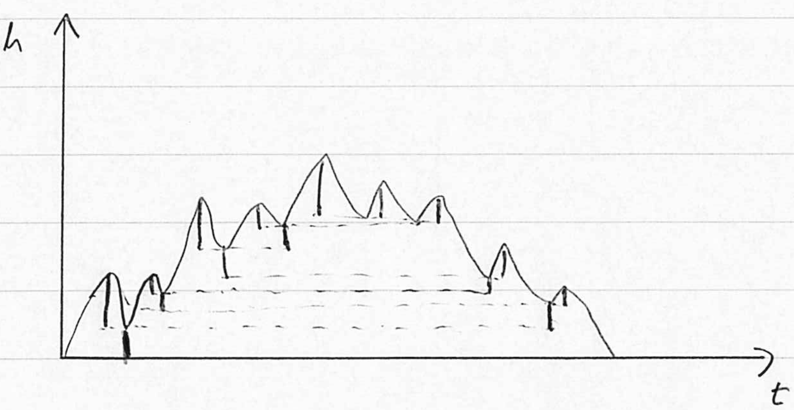
(b) γ_{xy} is the only non selfintersecting paths from x to y.

- Rooted real tree is a real tree with a distinguished vertex ρ.

Elements of $\mathcal{T}$ are called vertices of $\mathcal{T}$. A leaf is a vertex v such that $v\notin\gamma_{\rho w}$ for any $w\neq v$.

Coding real trees:

Assume that $h: [0,\infty) \to [0,\infty)$ is continuous with compact support. h will play the rôle of contour of real tree.



(4.67) Figure: h and corresponding real tree

Use h to define a distance d_h on supp h.

(4.68) $$d_h(x,y) = h(x) + h(y) - 2\inf\{h(z) : z \in [x,y]\}, \quad x \le y$$

Let $\sim$ be an equivalence relation

$$x \sim y \iff d_h(x,y) = 0,$$

and take the quotient $\mathcal{T}_h = \operatorname{supp} h / \sim$

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

(4.69) Definition Let e be the standard Brownian excursion. The random tree $\mathcal{T}_{2e}$ is called Aldous' continuum random tree (CRT).

Measuring distance between metric spaces.

Let (M,δ) be a metric space. The Hausdorff distance between two compacts subsets K, K' is

(4.70) $$d_H(K,K') = \inf\{\varepsilon > 0 : K \subset F_\varepsilon(K'), K' \subset F_\varepsilon(K)\}$$

with $F_\varepsilon(K) = \{y \in M : \delta(y,K) \le \varepsilon\}$.

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

(4.71) Def (Gromov-Hausdorff distance)

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

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

For rooted spaces, we set

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

This seems complicated, but fortunately for real trees $\mathcal{T}_h, \mathcal{T}_{h'}$ with coding functions h, h' we have

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

From (4.72) and (4.65) we can then deduce

(4.73) Theorem (Aldous) Let T_n be the critical GW tree conditioned to have size n, satisfying (4.59), endowed with the usual distance. Then

$$\frac{\sigma}{\sqrt{n}} T_n \xrightarrow[GH]{n\to\infty} \mathcal{T}_{2e}.$$

We will not prove all statements made here but it is useful to try

(4.74) Exercise: Let T_n be GW with offspring distribution $P(X=k) = 2^{-(k+1)}$, conditioned on $|T_n| = n$. Show that C is a SRW conditioned to hit 0 for the first time at $2(n-1)$.

We now come back and show the convergence of the height function of the critical GW forest.

Proof of Theorem (4.62)

We use the so called "ladder time decomposition of RW.

Recall that S is a random walk whose steps $X_i - 1 =: Y_i$ have mean 0 and variance σ, in particular they take values in the set $\{-1, 0, 1, 2, \dots\}$. We define

(4.75) $T = \inf\{k \geq 1 : S(k) \geq 0\}$

to be the first (weak) ascending ladder time.

S is recurrent, so T is a.s. finite

If $Y_0 \geq 0$, then $T = 1$ and $S(T) = Y_0$.

On the other hand, if $Y_0 = -1$, then S might do several excursions going below -1 and eventually return there several times, but finally S leaves -1 (maybe by going downwards) and reaches $\{0, 1, \dots\}$ without hitting -1 again. By strong Markov property, the number of excursions from (-1) to (-1) before hitting $\{0, 1, \dots\}$ is geometrically distributed with parameter $P(X(T) > 0)$

(4.76) Exercise: (a) Use above considerations to show that for $k \geq 0$

$$P(S(T) = k) = P(Y_0 = k) + P(Y_0 = -1) \cdot P(S(T) = k+1 \mid S(T) > 0)$$

(b) Use (a) to show that for $k \geq 0$

$$P(S(T) = k) = \sum_{j=0}^{\infty} \left(\frac{P(Y_0 = -1)}{P(S(T) > 0)} \right)^j P(Y_0 = k + j)$$

(c) Observe that $\sum_{k=0}^{\infty} \sum_{j=k}^{\infty} P(Y_0 = j) = E[Y_0 + 1] = 1$.

Deduce from this and (b) that

$$P(S(T) > 0) = P(Y_0 = -1)$$

And thus $P(S(T) = k) = \sum_{j \geq k} P(Y_0 = j)$, $k \geq 0$.

Hence also

$$E[S(T)] = \frac{\sigma^2}{2}.$$

We come back to (4.62). We write

(4.77) $$M(n) = \sup_{k\le n} S(k), \quad m(n) = \inf_{k\le n} S(k)$$

and introduce a time reversed walk

(4.78) $$\hat{S}^n(k) = S(n) - S(n-k), \quad k \in \{0, \dots, n\}.$$

It is easy to see that the increments of $\hat{S}^n$ have the same distribution as the increments of S, ie.

(4.79) $$(\hat{S}^n_k, k \le n) \stackrel{d}{=} (S_k, k \le n)$$

By (4.57)

(4.80) $$\begin{aligned} H(n) &= \#\{0 \le k < n : S(k) = \min_{k \le j \le n} S(j)\} \\ &= \#\{1 \le i \le n : \hat{S}^n(i) = \sup_{0 \le \ell \le i} \hat{S}^n(\ell)\} \end{aligned}$$

By analogy define

(4.81) $$J(n) = \#\{1 \le i \le n : S(i) = M(i)\}$$

and observe that

(4.82) $$\sup_{0 \le \ell \le n} \hat{S}^n(\ell) = S(n) - \inf_{0 \le \ell \le n} S(\ell) = S(n) - m(n).$$

From (4.79)-(4.82) we get that for every n fixed

(4.83) $$(M(n), J(n)) \stackrel{d}{=} (S(n) - m(n), H(n)).$$

We now claim

(4.84) <u>Claim</u>: $$\frac{H_n}{S(n) - m(n)} \xrightarrow{P} \frac{2}{\sigma^2}.$$

By (4.60), for $t_1 \le \dots \le t_m$

(4.85) $$\frac{1}{\sqrt{n}} \left(S(\lfloor nt_1 \rfloor) - m(\lfloor nt_1 \rfloor), \dots, S(\lfloor nt_m \rfloor) - m(\lfloor nt_m \rfloor) \right) \xrightarrow{d} \sigma \left(W(t_1) - \inf_{s \le t_1} W(s), \dots, W(t_m) - \inf_{s \le t_m} W(s) \right)$$

Hence, by (4.84).

$$\frac{1}{\sqrt{n}} \left(H(\lfloor nt_1 \rfloor), \dots, H(\lfloor nt_m \rfloor) \right) \xrightarrow{d} \frac{2}{\sigma} \left(W(t_1) - \inf_{s \le t_1} W(s), \dots, W(t_m) - \inf_{s \le t_m} W(s) \right)$$

proving (4.62) at least at the level of finite dimensional distributions. The proof of the tightness is omitted. $\square$

It remains to show (4.84).

Proof of (4.84): Set $T_0 = 0$,

(4.86) $$T_i = \inf\{k > T_{i-1} : S(k) = M(k)\}, \quad i \geq 1.$$

be the sequence of successive weak ascending ladder times (observe $T_1 = T$). Then,

(4.87) $$M(n) = \sum_{\substack{k\geq 1\\ T_k \leq n}} \big(M(T_k) - M(T_{k-1})\big) \overset{(4.81)}{=} \sum_{k=1}^{J(n)} \big(M(T_k) - M(T_{k-1})\big) = \sum_{k=1}^{J(n)} \big(S(T_k) - S(T_{k-1})\big)$$

Since $J(n) \nearrow \infty$ as $n \nearrow \infty$, by strong LLN and (4.76)

(4.88) $$\frac{M(n)}{J(n)} \xrightarrow{n\to\infty} E[S(T_1)] = \frac{\sigma^2}{2} \quad \text{a.s.}$$

Together with (4.83) it then implies (4.84) $\square$.

Shape of critical ER-graph clusters:

We close this chapter by explaining how the ideas of the previous section can be extended to understand the clusters of the critical ER graph. It is based on paper by Addario-Berry, Broutin and Goldschmidt.

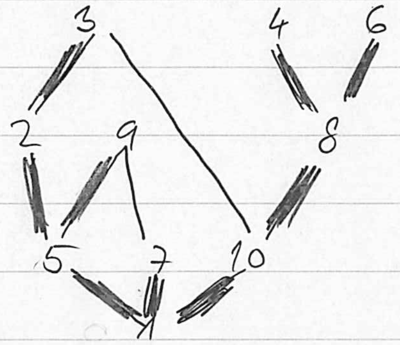
First we should however come back to Aldous (1997) paper. We have seen in (4.16) that rescaled sizes of clusters converge to excursion length of the process B^θ of (4.15). This was proved by replacing cluster by depth-first type walk, i.e. essentially by looking at trees contained in the clusters.

One can ask how much the clusters differ from trees. The first natural question is how many additional edges they contain. We thus define

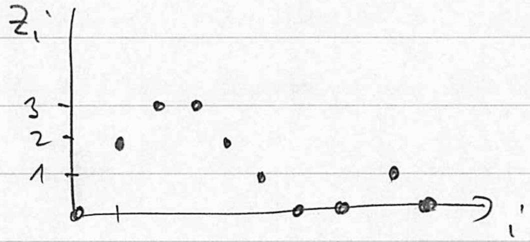
(4.89) $$S_n(i) = \#\text{edges in } i^{th} \text{ largest cluster} - \#\text{vertices} + 1$$

to be the number of "surplus" edges in i^{th} largest cluster. To understand S_n we come back to the exploration, and recall that the exploration checks the state of all possible edges between active and neutral vertices, but never checks edges between active vertices

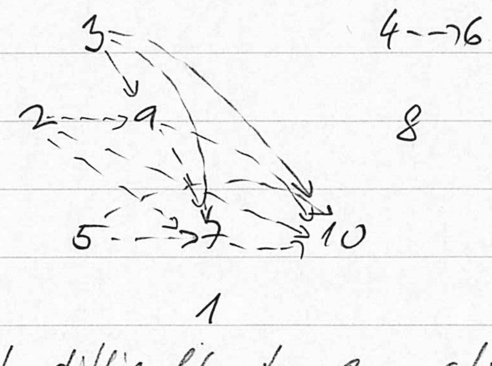
(4.90) Example Consider graph



Its depth-first walk visits vertices 1, 5, 2, 3, 9, 7, 10, 8, 4, 6 in this order. S_i, the number of active vertices at the end of ith step is $S_i = Z_i + 1$ with Z_i as



and the walk only discovers edges that are thick on the figure. But there are many "pending" edges it never really checked. These are all edges between current active and remaining active vertices. In the example these are those



It is not difficult to see after some thinking that number P of pending edges satisfies

(4.91) $$P = \sum_{k=0}^{|G|-1} Z_k .$$

Which is roughly the "area" under the graph of Z.

As consequence, in $ER(n, \frac{1}{n} + \frac{\theta}{n^{4/3}})$, given a component and its depth-first walk, the surplus is a

(4.92) $Bin(P, \frac{1}{n} + \frac{\theta}{n^{4/3}})$ random variable. Taking now the scaling of Z from (4.25) one obtains the following improvement of (4.16)

(4.93) Theorem: (Aldous) Let B^θ be as in (4.15) and consider Poisson point process with rate 1 in $[0,\infty)^2$. Let $\gamma_1 \geq \gamma_2 \geq \ldots$ be the ordered length of excursions of B^θ and let s_i be the # of points of the Poisson process under the excursion corresponding to γ_i. Then

$$(n^{-2/3} C_n(i), S_n(i))_{i\geq 1} \xrightarrow[n\to\infty]{} (\gamma_i, s_i)_{i\geq 1}.$$

To understand the shape one obvious first.

(4.94) Claim: A component of $G(n,p)$ conditioned to have m vertices and s surplus edges is a uniform connected graph with m vertices and $m+s-1$ edges.

For $s=0$ Aldous (1993), Le Gall (2005) proved that

(4.95) Thm: Let T_m be a uniformly chosen tree on m labelled vertices, viewed as metric space. Then

$$\frac{1}{\sqrt{m}} T_m \xrightarrow[GH]{d} \mathcal{T}_{2e} \quad \text{as } m \to \infty.$$

Here $\mathcal{T}_{2e}$ is the CRT.

Surpluses explained on the blackboard...