## Limit distribution of cluster sizes

In Theorem 4.1. we proved that $\left(\frac{|C_{max}|}{n^{2/3}}\right)_{n\geq 0}$ is tight in the critical ER graph $\left(p = \frac{1}{n} + \frac{\Theta}{n^{4/3}}\right)$. We now show that $\frac{|C_{max}|}{n^{2/3}}$ converges in distribution. Actually, we will show a stronger result.

(4.12) We write $C_n(1) := |C_{max}| \geq C_n(2) \geq C_n(3) \ldots$

for the ordered component sizes of the graph $ER(n, \frac{1}{n} + \frac{\Theta}{n^{4/3}})$. To describe the limit distribution, let $(W(s))_{s\geq 0}$ be the standard Brownian motion and set

(4.13) $$W^\Theta(s) = W(s) + \Theta s - \frac{1}{2}s^2, \quad s \geq 0,$$

(this is a BM with drift $\Theta - s$), and

(4.14) $$B^\Theta(s) = W^\Theta(s) - \min_{r\leq s} W^\Theta(r), \quad s\geq 0,$$

to be "process $W^\Theta$ reflected at 0". We call an interval $(l,r)$ an excursion of $B^\Theta$ if

(4.15) $$B^\Theta(r) = B^\Theta(l) = 0 \quad \text{and} \quad B^\Theta(s) > 0 \ \forall s \in (l,r).$$

The length of excursion $(l,r)$ is defined as $|r-l|$. Finally, let $\gamma_1 \geq \gamma_2 \geq \ldots$ be the set of the lengths of the excursions of $B^\Theta$, ordered by the length.

(4.16) Theorem: $\left(p = \frac{1}{n} + \frac{\Theta}{n^{4/3}}\right)$. For any $M \geq 0$,

$$n^{-2/3}\left(C_n(1), \ldots, C_n(M)\right) \xrightarrow[n\to\infty]{} (\gamma_1, \ldots, \gamma_M)$$

We start with some preparatory steps which allow also to intuitively understand (4.16). For this recall first the exploration algorithm of $C(1)$ from pages 17-18. In particular, recall that $X_k$ is the number of vertices becoming active in the $k^{th}$ step, and that the algorithm is stopped at the first instant when there are no active vertices.

We now extend this algorithm to discover the whole graph by adding an additional rule.

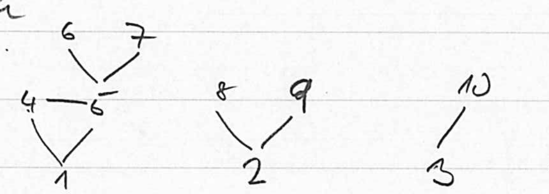
(4.17) When there are no active vertices, we make the neutral vertex with the smallest label active, and continue as previously.

We let $X_k$ keep its meaning and set

(4.18) $$Z_0^n = 0, \quad Z_k^n = Z_{k-1}^n + X_k - 1, \quad k \ge 1$$

(Observe that during the exploration of $\mathcal{C}(1)$, $Z_k^n = S_k - 1$)

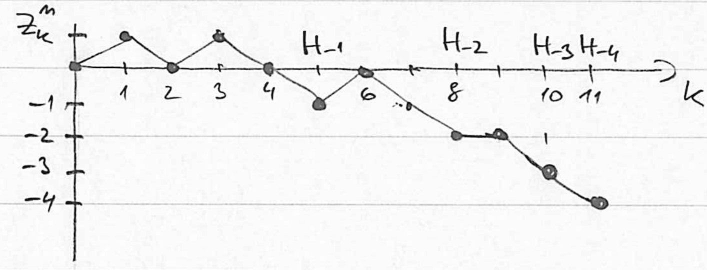
To see how it works consider the following graph




indices of explored vertices.

Here $X_1 \overset{1}{=} 2,\ X_2 \overset{4}{=} 0,\ X_3 \overset{5}{=} 2,\ X_4 \overset{6}{=} 0,\ X_5 \overset{7}{=} 0,\ X_6 \overset{2}{=} 2,$
$X_7 \overset{8}{=} 0,\ X_8 \overset{9}{=} 0,\ X_9 \overset{3}{=} 1,\ X_{10} \overset{10}{=} 0,\ X_{11} \overset{11}{=} 0$

and thus




Observe that $|\mathcal{C}(1)| \overset{(3.3)}{=} H_{-1}$ and that the remaining clusters have sizes $H_{-2}-H_{-1}, H_{-3}-H_{-2}, H_{-4}-H_{-3}$, with

(4.19) $$H_k = \inf\{i \ge 0 : Z_i^n = k\}.$$

We can thus read the cluster sizes from the process $(Z_k^n)_{k \le n}$

(4.20) $(\mathcal{C}_n(1), \mathcal{C}_n(2), \dots) =$ ordered set $(H_{-i} - H_{-i-1} : H_{-i} \le n)$. We set

(4.21) $$\overleftarrow{H}_i = \min\{j : H_{-j} \ge i\},$$

hence at step $i$ we explore $\overleftarrow{H}_i$-th cluster.

Denoting as before by $N_k, S_k$ the numbers of neutral and active vertices at the end of step $k$ we have

(4.22)
$$S_k = Z_k^n + \overleftarrow{H}_k, \quad k \geq 1$$
$$N_k = n - k - S_k = n - k - Z_k^n - \overleftarrow{H}_k .$$

(4.23) <u>Exercise</u>: Provide a more detailed proof of (4.22).

<u>Heuristic considerations:</u> (4.20) expresses the sizes of the clusters with the times $(Z_k)$ needs to reach a new minima. This have a very similar structure as (4.14), (4.15). This suggests that the decisive step on the way to (4.16) should be

(4.24) <u>Theorem</u>: Define

(4.25)
$$\bar{Z}^n(s) = n^{-1/3} Z^n_{\lfloor n^{2/3} s \rfloor} .$$

Then

(4.26)
$$\bar{Z}^n \xrightarrow[n \to \infty]{} W^{\theta},$$

where the convergence is in distribution in distribution on every $(D([0,T], \mathbb{R}), \|\cdot\|_\infty)$, $T \geq 0$.

We first explain why (4.24) should be true. Recall that $X_k \sim \mathrm{Bin}(N_{k-1}, p)$. After $\lfloor n^{2/3} s \rfloor$ steps, by (4.22), $N_k$ is roughly $n - \lfloor n^{2/3} s \rfloor$ (at least if $Z^n_{\lfloor n^{2/3} s \rfloor}$ and $\overleftarrow{H}_{\lfloor n^{2/3} s \rfloor}$ are $\ll n^{2/3}$). Hence,

(4.27)
$$EX_k \sim \left(n - \lfloor n^{2/3} s \rfloor\right) \cdot \left(\frac{1}{n} + \frac{\theta}{n^{4/3}}\right) \cong 1 + \frac{\theta}{n^{1/3}} - \frac{s}{n^{1/3}} + o(n^{-1/3})$$

Recalling (4.18) and (4.26) we see that at time $s \geq 0$, $\bar{Z}_s$ has drift $\theta - s$, c.f. (4.13).

We use martingale techniques to show (4.24). We need

(4.28) <u>Martingale convergence theorem</u>: For $n \geq 1$, let $\{\mathcal{F}_t^n\}_{t\geq 0}$ be a filtration and $(M_n(t))_{t\geq 0}$ a $\{\mathcal{F}^n\}$-local martingale with sample paths in $D = D([0,\infty), \mathbb{R})$, $M_n(0) = 0$. Assume that $(A_n(t))_{t\geq 0}$ is a non-decreasing process with paths in $D$ such that for every $T > 0$

(4.29)
(a) $M_n^2 - A_n$ is $\{\mathcal{F}^n\}$-local martingale.
(b) $\lim_{n\to\infty} E\left[\sup_{t\leq T} |A_n(t) - A_n(t-)|\right] = 0$
(c) $\lim_{n\to\infty} E\left[\sup_{t\leq T} |M_n(t) - M_n(t-)|^2\right] = 0$
(d) $\lim_{n\to\infty} A_n(t) = t$

Then, $M_n \to W$, where $W$ is a standard BM.

<u>Proof</u>: See Ethier-Kurtz: Markov Processes, Thm 7.1.4 (b).

To this end, we write $Z_k$ as
$$Z_k^n = M_k^n + E_k^n,$$
where $M_k$ is a martingale with respect to filtration $\mathcal{G}_k^n = \sigma(Z_i^n, i \leq k)$, and thus
$$E_k^n - E_{k-1}^n = E[Z_k^n - Z_{k-1}^n \mid \mathcal{G}_{k-1}^n] \; ; \; E_0 = 0.$$
and decompose $M_k^n$ as in (4.28)
$$(M_k^n)^2 = N_k^n + A_k^n$$
where $N^n$ is a $\{\mathcal{G}^n\}$-local martingale.

We will show that for $n \to \infty$ and $s_0 \geq 0$ fixed

(4.30) <u>Claim</u>: $n^{-1/3} \sup_{k \leq s_0 n^{2/3}} \left| E_k^n + \frac{1}{n}\frac{k^2}{2} - \frac{1}{n^{1/3}}\theta k \right| \xrightarrow{P} 0$

(4.31) <u>Claim</u>: $n^{-2/3} A^n_{\lfloor s_0 n^{2/3} \rfloor} \xrightarrow{P} s_0$

(4.32) <u>Claim</u>: $\lim_{n\to\infty} n^{-2/3} E\left[\sup_{k\leq s_0 n^{2/3}} |M_k^n - M_{k-1}^n|^2\right] = \lim_{n\to\infty} n^{-2/3} E\left[\sup_{k \leq s_0 n^{2/3}} |A_k^n - A_{k-1}^n|\right] = 0.$

Proof of (4.16) given (4.28)-(4.32).

We define $\bar{A}^n, \bar{E}^n, \bar{M}^n$ analogously to $\bar{Z}^n$, i.e., e.g.,

(4.33) $$\bar{A}^n(s) = n^{-1/3} A^n_{\lfloor n^{2/3} s \rfloor}.$$

It is then easy to see that (4.30)-(4.32) imply

(4.34) $$\sup_{s \le s_0} \left| \bar{E}^n(s) - \left(\theta s - \frac{s^2}{2}\right) \right| \xrightarrow{P} 0$$

and that $\bar{M}^n$ and $\bar{A}^n$ satisfy conditions of theorem (4.28). Hence $\bar{M}^n \longrightarrow W$, where $W$ is a BM and (4.34) yields

$$\bar{Z}^n = \bar{M}^n + \bar{E}^n \xrightarrow[n\to\infty]{d} W + q = W^\theta \text{ with } q(s) = \theta s - \frac{s^2}{2}.$$

This completes the proof □.

It remains to show the claims:

Proof of (4.30):

Since $E^n_k - E^n_{k-1} = E[Z^n_k - Z^n_{k-1} \mid \mathcal{G}^n_{k-1}] =$

(4.35) $$\overset{(4.18)}{=} E[X_k - 1 \mid \mathcal{G}^n_{k-1}] = E[\mathrm{Bin}(N_{k-1}, p)] - 1$$
$$\overset{(4.22)}{=} (n-(k-1) - Z^n_{k-1} - \overleftarrow{H}_{k-1}) \cdot \left(\frac{1}{n} + \frac{\theta}{n^{4/3}}\right) - 1,$$

we have, for $n > \theta^3$

(4.36) $$\left| E^n_k - E^n_{k-1} + \frac{k}{n} - \frac{\theta}{n^{1/3}} + \frac{k\theta}{n^{4/3}} \right| \le \frac{2}{n}\left(\overleftarrow{H}_k + |Z^n_k|\right).$$

Hence, by summing over $k$ yields,

(4.37) $$\left| E^n_k + \frac{k^2}{2n} - \frac{k\theta}{n^{1/3}} + \frac{k^2\theta}{2n^{4/3}} \right| \le \frac{2k}{n} \max_{i \le k}\left(\overleftarrow{H}_i + |Z^n_i|\right) + O\left(\frac{k}{n}\right).$$

Observe now that (4.21) may be written

(4.38) $$\overleftarrow{H}_i = -\min_{k<i} Z_k + 1$$

Hence, (4.37) $\le \frac{4k}{n} \max_{i \le k} |Z^n_i| + O\left(\frac{k}{n}\right)$. Therefore, to show (4.30) it is sufficient to prove that

(4.39) $$n^{-2/3} \sup_{k \le s_0 n^{2/3}} |Z^n_k| \xrightarrow{P} 0$$

To this end we will show a stronger statement:

(4.40) Claim: $n^{-1/3} \sup_{k \le s_0 n^{2/3}} |Z^n_k|$ is stoch. bounded as $n \to \infty$.

With (4.40), (4.30) follows □

Before proving (4.40) we proceed to (4.31), as we will need it there

<u>Proof of (4.31)</u> Since $M^n_\cdot$ is a martingale, we have

$$A^n_k - A^n_{k-1} = E\left[(M^n_k)^2 - (M^n_{k-1})^2 \mid \mathcal{G}^n_{k-1}\right] = E\left[(M^n_k - M^n_{k-1})^2 \mid \mathcal{G}^n_{k-1}\right]$$

$$= E\left[\left(Z^n_k - Z^n_{k-1} - (E^n_k - E^n_{k-1})\right)^2 \mid \mathcal{G}^n_{k-1}\right]$$

(4.41) $$\overset{(4.18),(4.35)}{=} E\left[\left(X_{k} - 1 - E[X_k \mid \mathcal{G}^n_{k-1}]\right)^2 \mid \mathcal{G}^n_{k-1}\right]$$

$$= \mathrm{Var}(X_k \mid \mathcal{G}^n_{k-1}) =$$

$$\overset{(4.22)}{=} (n - (k-1) - Z^n_{k-1} - \overleftarrow{H}_{k-1})\, p\,(1-p).$$

$$\overset{(4.35)}{=} 1 + (1-p)\cdot(E^n_k - E^n_{k-1}).$$

Hence, summing over $k$,

(4.42) $$A^n_k = k + (1-p)E^n_k$$

By (4.30), $n^{-2/3} \sup_{k \le s_0 n^{2/3}} |E^n_k| \longrightarrow 0$, and (4.31) follows $\square$.

<u>Proof of (4.40)</u> We use optional stopping argument.

(4.43) Set $T_n = \min\{k : |Z^n_k| \ge A n^{1/3}\} \wedge s_0 n^{2/3}$.

By optional stopping theorem

$$E\left[(M^n_{T_n})^2\right] = E\left[A^n_{T_n}\right] \overset{(4.41)}{=}$$

(4.44) $$= E\sum_{i=1}^{T_n} \left(n - (i-1) - \underbrace{Z^n_{i-1}}_{|\,| \le A n^{1/3}} - \underbrace{\overleftarrow{H}_{i-1}}_{\ge 0}\right) \underbrace{p}_{\sim \frac{1}{n}} (1-p) \qquad , \quad T_n \le s_0 n^{2/3}$$

$$\le T_n\, s_0\, n^{2/3}$$

$$\le 2 s_0 n^{2/3}, \text{ for } n \text{ large enough.}$$

Further,

$$E\left[|Z^n_{T_n}|\right] \le E|M^n_{T_n}| + E\left[|E^n_{T_n}|\right] \overset{(4.44),\text{ Jensen},(4.41)}{\le}$$

$$\le \sqrt{2 s_0 n^{2/3}} + E\left[|E^n_{T_n}|\right]$$

$$\overset{(4.37)-(4.38)}{\le} \sqrt{2 s_0 n^{2/3}} + \frac{T_n^2}{2n} + \frac{T_n \theta}{n^{1/3}} + \frac{4 T_n}{n} A n^{1/3} \quad \text{using def of } T_n$$

$$\overset{T_n \le s_0 n^{2/3}}{\le} \alpha n^{1/3} + 4 s_0 A,$$

with $\alpha = \alpha(s_0, \theta)$. Hence

$$P\left(\sup_{k \le s_0 n^{2/3}} |Z^n_k| \ge A n^{1/3}\right) = P\left(|Z^n_{T_n}| > A n^{1/3}\right) \le \frac{\alpha}{A} + \frac{4 s_0}{n^{1/3}},$$

proving (4.40). $\square$

(4.45) <u>Exercise</u>: Show (4.32)

We now proceed to the proof of Theorem 4.16.

Proof of (4.16):

We first need to show that excursions of the limit process are matched by the excursion of $Z_n$. To this end we have the following deterministic lemma:

(4.46) Lemma: Let $f: \mathbb{R}_+ \to \mathbb{R}$ be continuous and $\mathcal{E}$ the set of its "excursions above running minimum"

$$\mathcal{E} = \{ e = (l,r) : f(l) = f(r) = \min_{s \le l} f(s),\ f(s) > f(l)\ \forall s \in (l,r) \}$$

Assume that for every $e_1, e_2 \in \mathcal{E}$ with $l_1 < l_2$ we have $f(l_1) > f(l_2)$, and that $\lambda\left(\mathbb{R}_+ \setminus \bigcup_{e \in \mathcal{E}} (l,r)\right) = 0$. Set $\Xi = \{ (l, r-l) : (l,r) \in \mathcal{E} \}$. Let $f_n \to f$ uniformly on compacts and $(t_{n,i}, i \ge 0)$ satisfy

(i) $0 = t_{n,1} < t_{n,2} < \dots$, $\lim_{i \to \infty} t_{n,i} = \infty$

(ii) $f_n(t_{n,i}) = \min_{u \le t_{n,i}} f_n(u)$

(iii) $\max \{ f_n(t_{n,i}) - f_n(t_{n,i+1}) : i \text{ s.t. } t_{n,i} \le s_0 \} \xrightarrow{n \to \infty} 0$ for every $s_0 > 0$.

Set $\Xi^n = \{ (t_{n,i}, t_{n,i+1} - t_{n,i}), i \ge 0 \}$. Then $\Xi^{(n)} \to \Xi$.

(4.47) Remark: The convergence is vague Point process convergence. I.e. for every $f \in C_c(\mathbb{R}_+^2 \to \mathbb{R})$ $\sum_{x \in \Xi^n} f(x) \xrightarrow{n \to \infty} \sum_{x \in \Xi} f(x)$.

Proof: omitted.

(4.48) Lemma: Let $\Xi = \{ (l, r-l) : (l,r) \text{ is an excursion of } B^\theta \}$ and $\Xi_n = \{ ( n^{-2/3} H^n_{-i+1},\ \underbrace{n^{-2/3} (H^n_{-i} - H^n_{-i+1})}_{\text{size of } i\text{-th explored cluster}} ) : i \ge 1 \}$

Then $\Xi_n \xrightarrow{d} \Xi$.

Proof: (sketch) We use the following claim

(4.49) Claim: Let $X_n, X$ be r.v. taking values in some "nice" space $\mathcal{X}$ such that $X_n \xrightarrow{d} X$. Let $h: \mathcal{X} \to \mathcal{Y}$ and $D_h$ the set of discontinuity points of $h$: If $P(X \in D_h) = 0$, then $h(X_n) \xrightarrow{d} h(X)$.
(see Billingsley: Convergence of probability measures, Thm 5.1).

With this claim, (4.48) is a consequence of (4.46) and (4.14), (4.15), if one shows that $W^\theta$ satisfies the assumptions of (4.46) a.s. This is left for exercise. □

We now need to deal with ordering of components sizes, i.e. of second components of process $\Xi_n$. We should check that large excursions typically occur at the "beginning", i.e. no large excursions are overlooked because we have processes on the infinite half-line. We set.

$$T(y) = \min\{s : W^\theta(s) = -y\}$$
$$T_n(y) = \min\{i : Z_i^n = -\lfloor y n^{1/3}\rfloor\}$$

By (4.24), $T_n(y) \xrightarrow{d} T(y)$. Moreover, by the construction, at time $T_n(y)$ all vertices with labels $\le y n^{1/3}$ are explored. So Lemma (4.48) implies that a "limited version" of (4.16) is true, where one only considers components containing at least one of vertices $\{1,2,\dots,\lfloor y n^{1/3}\rfloor\}$. We need this lemma whose proof we omit.

(4.50) Lemma: Let $p(n,y,\delta) = P\left[\begin{array}{l}\text{there is component } C : |C| \ge \delta n^{2/3}, \\ |C| \cap \{1,\dots,\lfloor y n^{1/3}\rfloor\} = \phi\end{array}\right]$

Then $\lim_{y\to\infty} \limsup_{n\to\infty} p(n,y,\delta) = 0 \qquad \forall \delta > 0.$

□