For the upper bound, as previously,

$$P(|C(1)|\ge k_n) \overset{(3.8)}{\le} P^{BP}_{n,p}(T\ge k_n) \overset{(2.47)}{\le} P^{BP}_{\lambda}(T\ge k_n) + \frac{k_n\lambda^2}{n} =$$

$$= P^{BP}_{\lambda}(T=\infty) + P^{BP}_{\lambda}(k_n\le T<\infty) + O(\tfrac{k_n}{n})$$

$$\overset{(3.29)\,(2.26)}{\le} \zeta_\lambda + c_\lambda e^{-I_\lambda k_n} + O(k_n/n) \le \zeta_\lambda + O(n^{-1}) + O(\tfrac{k_n}{n}).$$

where we used $I_\lambda = \sup_t (t - \log E[e^{t\,\mathrm{Pois}(\lambda)}])$, recall (3.12), and $k_n \ge a\log n$, $a > \frac{1}{I}$.

For the lower bound, setting $\lambda_n = \frac{n-k_n}{n}\cdot\lambda = \lambda(1-\frac{k_n}{n})$

$$P(|C(1)|\ge k_n) \ge P^{BP}_{n-k_n,p}(T\ge k_n) \ge P^{BP}_{\lambda_n}(T\ge k_n) + O(\tfrac{k_n}{n}).$$

By (2.26) again

$$P^{BP}_{\lambda_n}(T\ge k_n) = \zeta_{\lambda_n} + O(e^{-k_n I_{\lambda_n}}) = \zeta_{\lambda_n} + O(\tfrac{1}{n}), \text{ as } a > \tfrac{1}{I_\lambda}$$

Finally, we need to show that $|\zeta_{\lambda_n} - \zeta_\lambda| \le c\frac{k_n}{n}$. This follows from the

(3.34) **Exercise:** Show that $\lambda \mapsto \zeta_\lambda$ is differentiable on $\lambda\in(1,\infty)$ with $\frac{d}{d\lambda}\zeta_\lambda\big|_{\lambda=1} = 2$. (Hint: recall (2.4), (2.10))

$\square$

Step 2: We now show that the number of vertices in "large" clusters concentrates. We will need another estimate on the variance of $Z_{\ge k}$, cf. (3.24)

(3.35) **Lemma:** For every $n\ge 1$ and $k\in[n]$

$$\mathrm{Var}^{ER}_{n,\frac{\lambda}{n}}(Z_k) \le (\lambda k+1)\,n\cdot E_{n,\frac{\lambda}{n}}\left[|C(1)|\,\mathbb{1}_{|C(1)|<k}\right]$$

!

(3.36) **Remark:** In the case when one expects that $C(1)\sim n$ with positive probability this is much better than (3.24), which would give $\mathrm{Var}(Z_n)\sim n^2$, which is essentially trivial.

Proof of (3.35): We set $Y_k = n - Z_k = \#\{x\in[n] : |C(x)|<k\}$

As $\mathrm{Var}(Y_k) = \mathrm{Var}(Z_k)$, it suffices to show (3.35) for Y_k.

Similarly as in (3.24),

(3.37) $\mathrm{Var}(Y_k) = \sum_{x,y\in[n]} \left(P(|\mathcal{C}(x)|<k, |\mathcal{C}(y)|<k) - P(|\mathcal{C}(x)|<k)^2 \right)$

We split the sum according to $x \leftrightarrow y$ again.

$\sum_{x,y} P(|\mathcal{C}(x)|<k, |\mathcal{C}(y)|<k, x\leftrightarrow y) =$ (if $x\leftrightarrow y$, then $\mathcal{C}(x)=\mathcal{C}(y)$)

(3.38) $= \sum_{x,y} P(|\mathcal{C}(x)|<k, x\leftrightarrow y)$

$= n\, E\left[1_{|\mathcal{C}(1)|<k} \sum_y y\leftrightarrow 1 \right] = n\cdot E\left[|\mathcal{C}(1)|\, 1_{|\mathcal{C}(1)|<k} \right].$

On the other hand, for $l<k$,

(3.39) $P(|\mathcal{C}(x)|=l, |\mathcal{C}(y)|<k, x\not\leftrightarrow y)$

$= P(|\mathcal{C}(x)|=l) \cdot \underbrace{P(x\not\leftrightarrow y \mid |\mathcal{C}(x)|=l)}_{\leq 1} \cdot P(|\mathcal{C}(y)|<k \mid |\mathcal{C}(x)|=l, x\not\leftrightarrow y)$

On $\{|\mathcal{C}(x)|=l, x\not\leftrightarrow y\}$, law of $|\mathcal{C}(y)|$ coincides with the law of $|\mathcal{C}(1)|$ in $ER(n-l,p)$. Viewing $ER(n-l,p)$ as a subset of $ER(n,p)$

$P_{n-l,p}(|\mathcal{C}(1)|<k) - P_{n,p}(|\mathcal{C}(1)|<k) = P_{n,p}(|\tilde{\mathcal{C}}(1)|<k, |\mathcal{C}(1)|\geq k)$,

where $\tilde{\mathcal{C}}(1) = \{ y\in[n-l], 1 \overset{\text{in } [n-l]}{\longleftrightarrow} y \}$. The last event can occur only if there is at least one edge connecting $\tilde{\mathcal{C}}(1)$ with $[n]\setminus[n-l]$, and this has probability $\leq p\cdot k\cdot l = \frac{\lambda}{n} kl$.

Hence, coming back to (3.39),

$P(|\mathcal{C}(y)|<k \mid |\mathcal{C}(x)|=l, x\not\leftrightarrow y) - P(|\mathcal{C}(y)|<k) \leq \frac{\lambda}{n} kl$

Inserting this into (3.39) and using this in (3.37),

(3.40) $\sum_{x,y} \left\{ P(|\mathcal{C}(x)|<k, |\mathcal{C}(y)|<k, x\not\leftrightarrow y) - P(\mathcal{C}(1)<k)^2 \right\} \leq$

$\leq \sum_{l=1}^{k-1} \sum_{x,y\in[n]} \frac{\lambda k l}{n} P(|\mathcal{C}(1)|=l) = \frac{\lambda k}{n} n^2 E\left[|\mathcal{C}(1)|\, 1_{|\mathcal{C}(1)|<k} \right]$

Combining (3.37), (3.38), (3.40) finishes the proof $\square$.

Step 3: We now show that it is extremely unlikely, that there are "middle ground" clusters of size between k_n and αn.

We start with an auxiliary claim.

(3.41) Lemma: Let S_n be as in (3.4), (3.6) that is $S_0 = 1$, $S_k = S_{k-1} + X_k - 1$, $k \geq 1$, with $X_k \sim Bin(n-(k-1)-S_{k-1}, p)$ (this defines S_k beyond H_0), and set $N_k = n-k-S_k$. Then

$$N_k \sim Bin(n-1, (1-p)^k) \text{ and thus}$$
$$S_k \sim Bin(n-1, 1-(1-p)^k) - (k-1)$$

Proof: By the above recursion,

$$N_k = n-k-S_k = n-k-S_{k-1} - X_k + 1 = N_{k-1} - X_k \approx Bin(N_{k-1}, 1-p)$$

Using induction in k we then obtain $N_k = Bin(n-1, (1-p)^k)$, since $N_0 = n-1$. The second claim then follows since $n-1-N_k \sim Bin(n-1, 1-(1-p)^k)$, by usual properties of binomial distribution $\square$.

(3.42) Lemma: Let $k_n = a \log n$ with $a > \frac{1}{I_\lambda}$ and $\alpha < \zeta_\lambda$. Then there is $\delta(a, \alpha, \lambda) > 0$ such that, for n large,

$$P(k_n \leq |\mathcal{C}(1)| \leq \alpha n) \leq e^{-k_n \delta}$$

Proof:
$$P(k_n \leq |\mathcal{C}(1)| \leq \alpha n) = \sum_{l=k_n}^{\alpha n} P(|\mathcal{C}(1)| = l)$$
$$\overset{(3.3)}{\leq} \sum_{l=k_n}^{\alpha n} P(S_l = 0) \overset{(3.41)}{=} \sum_{l=k_n}^{\alpha n} P[Bin(n-1, 1-(1-p)^l) = l-1]$$
$$\leq \sum_l P(Bin(n-1, 1-(1-p)^l) \leq l-1) \qquad (1-p \leq e^{-p})$$
$$\leq \sum_l P(Bin(n, 1-e^{-pl}) \leq l) \qquad (\text{Markov}, s \geq 0)$$
$$\leq \sum_l e^{sl} E[e^{-s\,Bin(n, 1-e^{-pl})}]$$
$$= \sum_l e^{sl} ((1-e^{-pl})e^{-s} + e^{-pl})^n$$
$$= \sum_l e^{sl} (1-(1-e^{-pl})(1-e^{-s}))^n \qquad (1-x) \leq e^{-x}$$

$$\leq \sum_{\ell} \exp\left\{ s\ell - n\left(1-e^{-\frac{\lambda \ell}{n}}\right)(1-e^{-s}) \right\}$$

Optimizing over s, i.e. taking $s = s^* = \log\left(\frac{n}{\ell}\left(1-e^{-\frac{\lambda\ell}{n}}\right)\right)$ and writing $g(s,\lambda) = \frac{1}{s}\left(1-e^{-\lambda s}\right)$ we obtain

(3.43) $$P\left(k_n \leq |\mathcal{C}(1)| \leq \alpha n\right) \leq \sum_{\ell=k_n}^{\alpha n} \exp\left\{ +\ell\left(\log g\left(\tfrac{\ell}{n},\lambda\right) - g\left(\tfrac{\ell}{n},\lambda\right) + 1\right)\right\}$$

Observe now that

$$\lim_{s \downarrow 0} g(s,\lambda) = \lambda > 1$$

(3.44) $$\frac{\partial}{\partial s} g(s,\lambda) = e^{-s\lambda}\,\frac{s\lambda - (e^{s\lambda}-1)}{s^2} < 0$$

$$g(\zeta_\lambda,\lambda) = \frac{1}{\zeta_\lambda}\left(1-e^{-\lambda\zeta_\lambda}\right) \overset{(2.40)}{\underset{(3.29)}{=}} \frac{(1-\eta_\lambda)}{1-\eta_\lambda} = 1.$$

Hence $s^* \geq 0$ only if $\frac{\ell}{n} \leq \zeta_\lambda$ which is the case since $\alpha < \zeta_\lambda$, that is (3.43) indeed holds.

Coming back to (3.43), since $x \mapsto \log x - x + 1$ is strictly decreasing on $x \geq 1$ and $g(\frac{\ell}{n},\lambda) > 1$ for all considered ℓ's,

(3.45) $$P\left(k_n \leq |\mathcal{C}(1)| \leq \alpha n\right) \leq \sum_{\ell \geq k_n} \exp\left\{+\ell\left(\log g(\alpha,\lambda) - g(\alpha,\lambda) + 1\right)\right\}$$
$$\leq e^{-k_n \cdot \sigma(\alpha,\lambda)}$$

where in the last inequality we used that
$\log(g(\zeta_\lambda,\lambda)) - g(\zeta_\lambda,\lambda) + 1 = 0$ and thus
$\log(g(\alpha,\lambda)) - g(\alpha,\lambda) + 1 < 0$ $\square$

<u>Step 4</u> : Conclusion.

Let $k_n = K\log n$ with K large, $\alpha \in (3\zeta_\lambda/4, \zeta_\lambda)$, $\varepsilon \in (0, \frac{\zeta_\lambda}{4})$. Set

(3.46) $$\mathcal{E}_n = \left\{ |Z_{k_n} - n\zeta_\lambda| \leq \varepsilon n \right\} \cap \left\{ \nexists x \in [n] : k_n \leq |\mathcal{C}(x)| \leq \alpha n \right\}.$$

By Lemmas (3.32), (3.35) and Chebyshev inequality

(3.47) $$P\left[|Z_{k_n} - n\zeta_\lambda| \geq \varepsilon n\right] \leq \frac{1}{\varepsilon^2 n^2}\, n\, k_n^2 \xrightarrow[n\to\infty]{} 0$$

and by Lemma (3.42), $P\left[\exists x : k_n \leq |\mathcal{C}(x)| \leq \alpha n\right] \leq n e^{-k_n \sigma} \xrightarrow{n\to\infty} 0$, for K sufficiently large.

Hence $\mathbb{P}^{ER}_{np}[\mathcal{E}_n] \xrightarrow{n\to\infty} 1$.

To conclude, we claim

(3.48) On $\mathcal{E}_n$, $|\mathcal{C}_{max}| = Z_{k_n}$.

Indeed, from the definition of $Z_{k_n} = \#\{x \in [n], |\mathcal{C}(x)| \geq k_n\}$ it follows easily that $|\mathcal{C}_{max}| \leq Z_{k_n}$, since $|\mathcal{C}(x)| = |\mathcal{C}_{max}|$ for all $x \in \mathcal{C}_{max}$. On the other hand, $|\mathcal{C}_{max}| < Z_{k_n}$ implies that there are at least two connected components of size at least k_n. Also, on $\mathcal{E}_n$, there are no connected components of size between k_n and αn. Therefore, there must be two connected components of size at least αn, which implies $Z_{k_n} \geq 2\alpha n \geq \frac{3}{2}\zeta_\lambda n$, which contradicts $|Z_{k_n} - n\zeta_\lambda| \leq \varepsilon n$, since $\varepsilon < \frac{1}{4}$. Hence, we conclude that (3.48) holds $\square$.

IV CRITICAL ER GRAPH

In previous chapter we proved that the Erdős–Rényi graph with $p = \frac{\lambda}{n}$ is "subcritical" for $\lambda < 1$ and "supercritical" for $\lambda > 1$. We now consider the case when p equals or is close to $\frac{1}{n}$. It turns out that in this regime $\mathcal{C}_{max}$ has nontrivial behaviour that we try understand in detail.

We study the size of $\mathcal{C}_{max}$ first.

(4.1) <u>Theorem:</u> Let $p = \frac{\lambda}{n}$ with $\lambda = 1 + \theta n^{-\frac{1}{3}}$, that is

(4.2) $$p = \frac{1}{n} + \frac{\theta}{n^{4/3}}, \qquad \theta \in \mathbb{R}.$$

Then, there exists $b = b(\theta) > 0$ such that for all $A > 1$

(4.3) $$\mathbb{P}^{ER}_{n,p}\left(A^{-1} n^{2/3} \le |\mathcal{C}_{max}| \le A n^{2/3}\right) \ge 1 - \frac{b}{A}.$$

In particular the sequence $\left(\frac{|\mathcal{C}_{max}(n)|}{n^{2/3}}\right)_{n\ge 1}$ is tight.

<u>Proof:</u> The proof follows similar strategy as in Chapter III, so we will skip some parts of the argument and concentrate on key steps that explain the $n^{2/3}$-scaling. We need two lemmas

(4.4) <u>Lemma</u>: (Critical cluster tails, $\lambda = 1 + \frac{\theta}{n^{1/3}}$, $\theta \in \mathbb{R}$)
Let $r > 0$ and $k \le r n^{2/3}$. Then there are $0 < c_1 < c_2 < \infty$ with $c_1 = c_1(r, \theta)$ s.t. for n large

$$\frac{c_1}{\sqrt{k}} \le \mathbb{P}^{ER}_{n,p}\left(|\mathcal{C}(1)| \ge k\right) \le c_2\left((\theta \vee 0) n^{-1/3} + \frac{1}{\sqrt{k}}\right).$$

(4.5) <u>Lemma</u> (Expected critical cluster size, $\lambda = 1 + \frac{\theta}{n^{1/3}}$, $\theta < 0$)
For every $n \ge 1$, $\mathbb{E}^{ER}_{n,p} |\mathcal{C}(1)| \le n^{1/3} / |\theta|$.

(4.6) Remarks:

(a) Lemma (4.4) should be compared with (2.46) where we showed that for Poisson(1)-branching process $P_1^{BP}(T=k) = c k^{-\frac{3}{2}}(1+o(1))$, that is $P^{BP}(T \geq k) = c k^{-1/2}(1+o(1))$.
Of course, one cannot expect this bound to be true for all k, but (4.4) implies that it holds for $k \leq r n^{2/3}$.

(b) (4.5) is "intuitively" consistent with (4.1). If one believes/expects that $E|C(1)|$ is dominated by the contribution of the case when $1 \in C_{max}$, we have

$$(4.7)\quad E\left[|C(1)|\right] \approx E\left[|C(1)| \mathbb{1}_{1\in C_{max}}\right] = E\left[|C_{max}| \mathbb{1}_{v \in C_{max}}\right] = \frac{1}{n} E\left[|C_{max}|^2\right]$$

(Here „$\approx$" is uncontrolled, but can be justified by (4.4)).
When $|C_{max}|$ is typically $\sim n^{2/3}$, then $E(|C_{max}|^2) \sim n^{4/3}$, i.e., by (4.7), $E(|C(1)|) \approx n^{1/3}$ as in (4.5).

Proof of (4.1) using (4.4) & (4.5).

By taking b large we may assume that A is large, since otherwise the RHS of (4.3) is negative. We also w.l.o.g. assume that n is large.

Upper bound: We recall $Z_k = \#\{x \in [n] : |C(x)| \geq k\}$ and thus $\{|C_{max}| \geq k\} = \{Z_k \geq k\}$. (cf. under (3.19)).
Therefore,

$$P(|C(1)| \geq A n^{2/3}) = P(Z_{An^{2/3}} \geq A n^{2/3}) \leq A^{-1} n^{-2/3} E[Z_{An^{2/3}}]$$

$$(4.8)\quad = A^{-1} n^{-2/3} n \cdot P(|C(1)| \geq A n^{2/3}) \overset{(4.4)}{\leq} A^{-1} n^{\frac{1}{3}} \cdot c_2 \left((\theta \vee 0) n^{-\frac{1}{3}} + \frac{1}{\sqrt{A} n^{1/3}}\right)$$

$$= \frac{c_2}{A}\left((\theta \vee 0) + \frac{1}{\sqrt{A}}\right) \leq \frac{1}{A} \cdot (c_2(\theta \vee 0) + 1).$$

<u>Lower bound:</u> Since, $|\mathcal{C}_{max}|$ is "increasing in λ", it is sufficient to establish the LB for $\theta \leq -1$.

Using $\{|\mathcal{C}_{max}| \leq k\} = \{Z_k = 0\}$ and Chebyshev

$$P(|\mathcal{C}_{max}| \leq A^{-1} n^{2/3}) = P(Z_{A^{-1}n^{2/3}} = 0)$$

(4.9) $$\leq \frac{Var(Z_{A^{-1}n^{2/3}})}{E[Z_{A^{-1}n^{2/3}}]^2}$$

By (4.5), as in (4.8),

(4.10) $$E[Z_{A^{-1}n^{2/3}}] = n \cdot P[|\mathcal{C}(1)| \geq A^{-1}n^{2/3}] \geq c_1 \sqrt{A}\, n^{2/3}.$$

By Lemma (3.24), (4.5), $\theta \leq -1$,

(4.11) $$Var(Z_{A^{-1}n^{2/3}}) \leq n \cdot E[|\mathcal{C}(1)|] \overset{(4.5)}{\leq} n^{4/3}.$$

Combining (4.9)–(4.11) yields the lower bound □.

<u>Proof of (4.5).</u> By the comparison with binomial BP, (3.8)

$$E_{n,p}(|\mathcal{C}(1)|) \leq E^{BP}_{n,p}[T] = \sum_{k \geq 0}^{\infty} E^{BP}_{n,p}(Z_k)$$

where Z_k denotes the size of k-th generation, c.f. (2.2

By (2.17), since $\theta < 0$,

$$= \sum_{k \geq 0}^{\infty} (n \cdot p)^k = \frac{1}{1 - np} = \frac{1}{1 - n\left(\frac{1}{n} + \frac{\theta}{n^{4/3}}\right)} = \frac{n^{1/3}}{|\theta|}. \quad \square.$$

<u>Proof of (4.4):</u> Is long and uses the same strategy as before, based on comparison lemmas (3.8), (3.10) and properties of Binomial / Poisson BP. For details see [vdH], pages 154-155. As we will prove better results soon, we omit it here. □