Weak subcritical percolation on finite mean-field graphs.

Olivier Hénard *

September 17, 2021

Abstract

We consider the largest connected components in the percolation of a (large) finite vertex-transitive graph. A geometrical condition reminiscent of the one in Nachmias [Nac09] is formulated. Under this condition, the many-to-two formula allows one to compute the size of the largest components in the weak subcritical regime up to fluctuations. The result applies to Hamming graphs up to dimension 3 and expander.

1 Introduction

Let G_n be a vertex-transitive graph on n vertices, and call $G_n(p)$ the random subgraph of G_n obtained by deleting every edge of G_n independently and with the same probability $1-p$ for $p \in (0,1)$. The number of vertices in the connected components of $G_n(p)$ is one of the simplest functionals of the graph G_n. Following a long tradition, we study the number of vertices, or sizes, of the largest of these connected components. In the setting of an arbitrary vertex-transitive graph, the problem has been started in the series of papers [BCvdH$^+$05a, BCvdH$^+$05b, BCvdH$^+$06] and continued in [Nac09, vdHN15, HN20]. Specific graphs G_n had been considered much before: the case of $K_n(p)$ with K_n the complete graph on n vertices is the classical Erdös-Rényi [ER60] random graph model that has been examined in detail by Erdös and Rényi in the '60s. A convenient parametrisation for the probability p of retaining an edge is in this case

$$p = \frac{\lambda}{n} \quad \text{for } \lambda > 0 \text{ a constant independent of } n \tag{1}$$

For large values of n, the size of the largest component of $K_n(p)$ expects a double jump when λ is increased: it is $\Theta_{\mathbb{P}}(\log(n))$ [1] when $\lambda < 1, \Theta_{\mathbb{P}}\left(n^{2/3}\right)$ when $\lambda = 1$, and $\Theta_{\mathbb{P}}(n)$ when $\lambda > 1$. As for the size of the second largest component of $K_n(p)$: it is $\Theta_{\mathbb{P}}(\log(n))$ when

*LMO, Université Paris-Saclay, 91405 Orsay Cedex, France. Partial support from grant ANR-14-CE25-0014 (ANR GRAAL) and EPSRC grant EP / J004022/2

[1] see Janson [Jan11] for a definition of the probabilistic symbols $\Theta_{\mathbb{P}}$ and $o_{\mathbb{P}}(n)$ used in the Introduction only.

$\lambda < 1, \Theta_{\mathbb{P}}\left(n^{2/3}\right)$ when $\lambda = 1$, and $\Theta_{\mathbb{P}}(\log(n))$ again when $\lambda > 1$. The three regimes are respectively called the subcritical, the critical and the supercritical regime. A distinctive feature of the critical regime is the non-concentration of the size of the largest component, that weakly converges as $n \to \infty$ towards a non-degenerate random variable. Aldous [Ald97] gives a construction of that random variable from the sample path of a Brownian motion with a quadratic drift.

Always in the $K_n(p)$ case, Bollobás [Bol84] and Łuczak [Łuc90] discovered new regimes of interest around the critical value $\lambda = 1$: these regimes are parametrized by sequences $\lambda = \lambda(n)$ with limit 1. Many qualitative features of the critical case $\lambda = 1$ are retained when $\lambda = \lambda(n)$ approaches 1 fast enough: $|\lambda - 1| = O\left(n^{-1/3}\right)$, and this regime is called the critical regime. An essential feature of that regime is the non-concentration of the sizes $\Theta_{\mathbb{P}}\left(n^{2/3}\right)$ of the largest components. The two remaining regimes are parametrised by positive sequences $\varepsilon = \varepsilon(n)$ satisfying [2]

$$\lambda = 1 \pm \varepsilon \quad \text{with} \quad \varepsilon \to 0 \quad \text{and} \quad \varepsilon^3 n \to +\infty.$$

Choosing the *minus* sign defines the weak subcritical regime in which the largest component has size $(2 + o_{\mathbb{P}}(1))\, \varepsilon^{-2} \log(\varepsilon^3 n)$, a quantity that interpolates between $\log(n)$ and $n^{2/3}$ for ε in the range described above. Choosing the *plus* sign defines the weak supercritical regime, in which the size of the largest component is $(2 + o_{\mathbb{P}}(1))\, \varepsilon n$, a quantity that interpolates between $n^{2/3}$ and n. In the weak supercritical regime, the second largest component is negligible with respect to the largest component, therefore called the giant component.

In this paper, we consider a sequence $(G_n)_{n\in\mathbb{N}}$ of vertex-transitive graphs on n vertices in place of the complete graph K_n. By vertex-transitivity, the degree ℓ of a vertex in G_n is the same for all the vertices, and we shall (mainly) consider sequences $\ell = \ell(n)$ diverging with n. We choose the percolation probability to be

$$p = \frac{\lambda}{\ell - 1} \tag{2}$$

and, again, we consider $\lambda = \lambda(n)$. The choice (2) is the natural generalisation of (1): the expected total number of edges in $G_n(p)$ is for instance in both cases equal to $\lambda n/2$. In the same setting, Nachmias [Nac09] investigates the critical and in the weak supercritical regime. Under condition (63) mentioned at the end of this article, he proves that the largest components in the critical regime have $\Theta_{\mathbb{P}}\left(n^{2/3}\right)$ many vertices; in the weak supercritical regime, under a slightly stronger condition, he gives a lower bound on the size of the largest component, $\delta\varepsilon n/\log(\varepsilon^3 n)$ for some fixed constant $\delta > 0$ independent of n; this latter result has been since superseded by the more general work [vdHN15], that catches the correct asymptotic value $2\varepsilon n\,(1 + o_{\mathbb{P}}(1))$. If the choice (2) is natural, it is not always adequate: for many interesting graphs, at the percolation probability $1/(\ell - 1)$, $|\mathcal{C}_1| = o_{\mathbb{P}}\left(n^{2/3}\right)$ so this value does not lie in the so-called critical window. A line of research investigates the analogy between the percolation of those graphs with the percolation of $K_n(p)$ by first defining an appropriate notion of critical probability, see the series [BCvdH$^+$05a, BCvdH$^+$05b, BCvdH$^+$06] and the recent works [vdHN15, HN20] for significant successes.

[2] ε shall always denote a positive quantity in this paper

2 Statement of the results and discussion.

Let G_n be a sequence of vertex transitive graphs on n vertices. The degree of a vertex is $\ell = \ell(n) \geqslant 3$. We will mainly work under the assumption

$$\ell(n) \to +\infty \tag{3}$$

although the case where $\ell(n)$ has a finite limit is also discussed at the end of the article. The edges of G_n are retained independently with probability p, and we denote by $G_n(p)$ the resulting random subgraph of G_n. We let

$$p_\pm = \frac{1 \pm \varepsilon}{\ell - 1}, \text{ for } \quad \varepsilon = \varepsilon(n) \to 0, \quad \text{ and } \varepsilon^3 n \to \infty \tag{4}$$

a positive sequence, and we call weak sub-critical the regime in which $p = p_-$ and weak super-critical regime the regime in which $p = p_+$

A function that is useful in describing the sizes of the largest components of $G_n(p)$ is

$$\delta_\pm(\varepsilon, \ell - 1) = -\log(1 \pm \varepsilon) - (\ell - 2)\log\left(1 - \frac{\pm\varepsilon}{\ell - 2}\right) \tag{5}$$

It is connected to the tail of the size of a $\mathrm{Bin}\,(\ell - 1, p_\pm)$-Galton-Watson tree (GW-tree hereafter), see (29) and (30). Taking the $\ell \to \infty$ limit in (5) we find that $\lim_{\ell\to\infty} \delta_\pm(\varepsilon, \ell - 1) = -\log(1 \pm \varepsilon) \pm \varepsilon$, a function that is ubiquitous in the study of the Erdös-Rényi random graph. It is useful to keep in mind the equivalent

$$\delta_\pm(\varepsilon, \ell - 1) \sim \frac{\varepsilon^2}{2} \text{ when } \varepsilon \to 0 \text{ and } \ell \to \infty \tag{6}$$

The probability that the non-backtracking random walk on G_n (a random walk on the vertices of G_n not allowed to traverse the same edge on two consecutive steps) started at a vertex $v \in G_n$ returns at v after $k \geqslant 2$ steps is denoted by $P^k(v, v)$. By vertex-transitivity, this quantity does not depend on the vertex v, and we write P^k for $P^k(v, v)$. We also introduce the two parameters

$$t_\pm = \delta_\pm^{-1} s \text{ and } s = \log\left(\varepsilon^3 n\right)$$

Notice that, when n and ℓ diverge, $t_\pm \sim 2\varepsilon^{-2}\log(\varepsilon^3 n)$. Our key (asymptotic) condition on the (sequence of) graphs G_n then writes:

$$\exists \text{ a constant } c < 1/2 \text{ such that: } \ (t_\pm)^{1/2} \sum_{k\geqslant 3} k \mathrm{e}^{-ck^2/t_\pm} P^k = o\left(\frac{1}{s}\right) \text{ as } n \to \infty \tag{7}$$

Condition (7) contains two conditions, the one with t_+ and the one with t_-, we distinguish them by writing $(7)_+$ and $(7)_-$ when needed. Condition (7) is a condition on the geometry of the graphs G_n, where by "geometry" we simply mean the collection of numbers $\left(P^k, k \in \mathbb{N}\right)$. We stress on the implicit dependence of the parameters $t_\pm, P^k$ and s on ε and/or n. The

quantity $k\mathrm{e}^{-ck^2/t}$ is an upper bound on the expected number of vertices at distance k from the root in a (critical) GW-tree with finite variance and t vertices, and the bound is uniform in $k \in \{1, \ldots, t\}$. See Addario-Berry et al [ABDJ13] for related estimates for GW-trees whose offspring distribution is critical and has a finite variance. The quantity $t^{1/2}$ that multiplies the sum in (7) is the expected distance of a random vertex to the root in such a GW-tree. We will clarify in Section why $o(1/s)$ in the RHS of (7) is indeed the precision needed to state the following Theorem :

Theorem 2.1. *Assume ε satisfies (4), ℓ satisfies (3) and let $p_- = (1-\varepsilon)/(\ell-1)$. Assume also that the non-backtracking random walk on G_n satisfies condition (7)$_-$. Let $(\mathcal{P}_j, j \in \mathbb{N})$ be the points, arranged in non-increasing order, of a Poisson point measure with intensity*

$$(4\sqrt{\pi})^{-1}\mathrm{e}^{-x}dx \tag{8}$$

on the real line $\mathbb{R}$. The sizes $(|\mathcal{C}_j|, j \in \mathbb{N})$ of the largest components of $G_n(p_-)$ arranged in non-increasing order converge in distribution for the product topology, in the sense that for each fixed $k \in \mathbb{N}$, as $n \to \infty$, we have:

$$\left(\delta_-(\varepsilon, \ell-1)\,|\mathcal{C}_j| - \left(\log\left(\varepsilon^3 n\right) - \frac{5}{2}\log\log\left(\varepsilon^3 n\right)\right), 1 \leqslant j \leqslant k\right) \Longrightarrow (\mathcal{P}_j, 1 \leqslant j \leqslant k)\cdot \tag{9}$$

Remark 2.2 (On the right-hand side in (9)). The computation:

$$\mathbb{P}(\mathcal{P}_1 < y) = \mathrm{e}^{-\int_{[y,\infty)}(4\sqrt{\pi})^{-1}\mathrm{e}^{-x}dx} = \mathrm{e}^{-(4\sqrt{\pi})^{-1}\mathrm{e}^{-y}}, \quad y \in \mathbb{R}$$

ensures the rightmost point of the Poisson measure in (8) exists and is Gumbel distributed. The convergence in distribution of a vector entails the convergence in distribution of its first coordinate. Also, the Gumbel distribution has no atom. Therefore (9) implies, for $y \in \mathbb{R}$:

$$\mathbb{P}\left(\delta_-(\varepsilon, \ell-1)\,|\mathcal{C}_1| - \left(\log\left(\varepsilon^3 n\right) - \frac{5}{2}\log\log\left(\varepsilon^3 n\right)\right) > y\right) \to 1 - \mathrm{e}^{-(4\sqrt{\pi})^{-1}\mathrm{e}^{-y}}$$

as $n \to \infty$. In the same way, it is possible to write down explicitly the limiting distribution of the size of the k-th largest component for $k \in \mathbb{N}$ fixed. Another byproduct is that the (rescaled) spacing between the two largest random variables weakly converges towards an exponential random variable with parameter 1 : $\mathbb{P}(\delta_-(\varepsilon, \ell-1)(|\mathcal{C}_1| - |\mathcal{C}_2|) > x) \to \mathrm{e}^{-x}$ as $n \to \infty$.

Remark 2.3. (On the left-hand side in (9)) From (6) and (9), we deduce the first order asymptotics: $|\mathcal{C}_j| = (1 + o_{\mathbb{P}}(1))\,\delta^{-1}(\varepsilon, \ell-1)\log(\varepsilon^3 n) = (1 + o_{\mathbb{P}}(1))\,2\varepsilon^{-2}\log(\varepsilon^3 n)$. Beware one cannot in general replace $\delta_-(\varepsilon, \ell-1)$ by $\varepsilon^2/2$ in (9). Replacing $\delta_- = \delta_-(\varepsilon, \ell-1)$ by another quantity $\tilde{\delta}_-$ on the left-hand side of (9) is possible as long as $\left(\delta_- - \tilde{\delta}_-\right)/\tilde{\delta}_- = o(1/s)$ as $n \to \infty$. From the expansion:

$$\delta_- = \sum_{k \geqslant 2} \frac{\varepsilon^k}{k}\left(1 + \frac{(-1)^k}{(\ell-2)^{k-1}}\right) \tag{10}$$

we see that $\delta_- = (1 + O(\varepsilon + 1/\ell))\varepsilon^2/2$ as $\ell \to \infty$ and $\varepsilon \to 0$. The choice $\tilde{\delta}_- = \varepsilon^2/2$ is thus possible under the conditions ($\varepsilon \log(\varepsilon^3 n) = o(1)$ and $\log(\varepsilon^3 n) = o(\ell)$).

Remark 2.4. (Weakly dependent random variables) For $v \in G_n$, call $\mathcal{C}(v)$ the component that contains the vertex v in the random subgraph $G_n(p)$. The n coordinates of the vector $(|\mathcal{C}(v)|, v \in G_n)$ are identically distributed, with the tail of their common distribution given by (31), but not quite independent. Were these random variables also independent, then the statement of Theorem 2.1 would hold with only $\varepsilon^3 n$ replaced by εn on the LHS of (9). The slight difference reflects the (weak) dependence between the random variables $(|\mathcal{C}(v)|, v \in G_n)$ that is caused by repetitions : if $|\mathcal{C}(v)| = t$ for some vertex v in G_n, then $|\mathcal{C}(u)| = t$ for at least $t-1$ other vertices u, the ones that belong to that component.

Remark 2.5. (Further illustration of the weak dependence) The difference is also visible when computing the expected number of components with size larger than t. It is in general bounded by (and, for the aforementioned choice of t, in fact equivalent to) $(n/t)\mathbb{P}(|\mathcal{C}(v)| > t)$ and *not* to $n\mathbb{P}(|\mathcal{C}(v)| > t) = \Theta(1)$ as it would be in the independent case: the division by t precisely accounts for the aforementioned repetitions. The proof will show it is possible to deduce the correct order of magnitude for t from this first moment argument : setting $t = \delta^{-1}s$ in the equation $(n/t)\mathbb{P}(|\mathcal{C}(v)| > t) = 1$, the LHS is by (31) equivalent to $(\varepsilon^3 n) \cdot s^{-5/2} \cdot \mathrm{e}^{-s}$ up to multiplicative constant, and this is 1 when $s = \log(\varepsilon^3 n) - 5/2 \log\log(\varepsilon^3 n) + O(1)$. Notice the $\log(\varepsilon^3 n)$ term comes from the exponential decay, and the $\log\log(\varepsilon^3 n)$ from the polynomial correction.

With the help of the estimates on the kernel of the non-backtracking random walk computed in [Nac09], one can check condition $(7)_-$ holds for some new classes of graphs. A sequence of connected graphs G_n is called an expander family if the largest eigenvalue in absolute value of the transition matrix of the simple random walk on G_n, distinct from ± 1, is strictly smaller than 1, uniformly in n. The girth of a graph is the length of a shortest cycle in the graph. The d-dimensional Hamming graph is the cartesian products of d complete graphs, it has vertex set $V = \{1, \dots, n\}^d$ and two vertices are linked by an edge iff the associated d-tuples differ at a single coordinate. Also $\lfloor x \rfloor \in \mathbb{Z}$ denotes the integer part of $x \in \mathbb{R}$.

Proposition 2.6. *Assume the sequence ε satisfies (4). Condition (7) holds for the following graphs G_n :*

- *the transitive expander graphs with girth $g = g(n)$ and vertex degree $\ell = \ell(n)$ that satisfy*
$$\left(\frac{1}{\ell(n)-1}\right)^{\lfloor g(n)/2 \rfloor} n^{1/3} \log^2(n) = O(1) \tag{11}$$
- *the Hamming graph in dimension $d = 3$.*

Remark 2.7 (Hamming graphs). The Hamming graphs are expanders that in dimension 1 and 2 satisfy condition (11) in the first statement. The Hamming graph in dimension 1 is the complete graph, hence the result on the Erdös-Rényi random graph by Łuczak [Łuc90] is recovered, whereas the result in dimension 2 and 3 is new. See Section 2.1 below for recent results on these graphs that are valid in *any* dimension.

Theorem 2.1 (together with the verification of condition (7)) are the highlights of this paper. Slight improvements are conceivable: for instance the assumption of transitivity of the graph could be slightly relaxed to include regular graphs. Simlarly, in our theorem, P^k could possibly be replaced by the (slightly smaller) probability that the random walk draws a self-avoiding loop [3] of length k : in view of the form of condition (7), one expects little benefit on the applicability of Theorem 2.1 however. Our results follows the line of inquiry set up by Nachmias [Nac09]. Because the precise study of the weak subcritical regime is not relevant for studying the width of the critical window, this work paid little attention to this regime, see the bottom of p.1173 in [Nac09]. That left aside a key feature of the weak subcritical regime, namely that it allows for a sharp estimate of the size of the largest components under general assumptions. That omission was subsequently repaired by the important work [HN20] that tackles the more general setting where the critical probability is *implicitly* defined only. If the largest components in the weak subcritical regime are notably smaller, hence "easier" to deal with, than their counterparts in the critical and supercritical regimes, finding a sharp estimate with the right multiplicative constant may represent a technical challenge, see [HN20] again. Our setting, where the critical probability is explicit, is easier, and allows one to obtain the optimal precision (fluctuations) by carefully controlling the difference between a (conditioned) GW-tree and the component containing a given vertex in $G_n(p)$. This crucially requires to work with GW-trees conditioned by the size and not by the height like in [Nac09, HN20]: the former conditioning better approximates the geometry of the largest components in $G_n(p)$ in our regime. A key technical tool to succeed is the "many-to-two" formula, that we rederive from scratch, see formula (19). If the use of a many-to-k formula seems new in the context of random graphs, the tool, that goes back to back to Ikeda et al [INW69] in the case $k = 2$, has already proved very useful in the study of branching Brownian motion. A generic version, the many-to-k (or many-to-few) formula is discussed in Harris and Roberts [HR15]; when $k = 1$, it reduces to the standard many-to-one formula, see [ABDJ13] or chapter 12 of the book [LP17].

2.1 Previous work

Let us try to summarize the state of affairs concerning the study of the size of the largest components in the weak subcritical regime:

- The complete graph model $K_n(p)$: The fluctuations of the random variable $|\mathcal{C}_j|$, $j \in \mathbb{N}$, have been identified by Łuczak [Łuc90] in a sharpening of a result by Bollobás [Bol84]. We warn the reader of a small typo in the statement of the result in [Łuc90]: $\varepsilon^2/2$ should be replaced by δ for the result to hold along the critical window, see Remark 1.3 or [BR09] p.50.

- The configuration model: The fluctuations of $|\mathcal{C}_1|$ are (among other) given in Riordan [Rio12], they involve the same Gumbel distribution as in our result.

[3] a path $v_0, v_1, \ldots, v_k = v_0$ where $0 \leq i < j < k \Rightarrow v_i \neq v_j$

- The Achlioptas bounded-size rules: Riordan and Warnke [RW17] offer a complete study of the transition phase in the context of iteratively constructed random graphs, in which edges come by pair (say) and some amount of choice in the edge to be added is allowed. Again, the authors are able to compute the fluctuations of $|\mathcal{C}_1|$, see their Theorem 2.7.

- The $G_n(p)$ model: Deterministic graphs G_n distinct from the complete graph require very different methods; upper and lower bounds on $|\mathcal{C}_j|, j \in \mathbb{N}$, have been known for some time now under a finite size version of the triangle condition known as the "strong" triangle condition, see Theorem 1.2 in Borgs, Chayes, van der Hofstad and Spencer [BCvdH$^+$05a] but these bounds were separated by a multiplicative $\log(\varepsilon^3 n)$ factor. Only very recently has this gap been reduced to a multiplicative constant factor, see [HN20] : $|\mathcal{C}_1| = \Theta_{\mathbb{P}}(\varepsilon^{-2}\log(\varepsilon^3 n))$. A key difficulty in these works is that the critical probability is only defined implicitly; we should also mention in this direction some recent progress concerning the asymptotic expansion of the critical probability in the special case of the Hamming graph in (fixed) dimension $d \geqslant 1$, see [FVDHDHH20].

An important convention concerning the $\pm$ index: when the index is omitted like in δ, t, one should understand that the statement holds for both values $\delta_\pm, t_\pm$. An identity with such a quantity therefore contains two identities: the one with the $\bullet_+$ index and the one with the $\bullet_-$ index.

2.2 Ideas of proof

The key parameters involved in the study are, in term of $\delta = \delta(\varepsilon, \ell - 1), \varepsilon$ and n :

$$t = \delta^{-1}s, s = \log\left(\varepsilon^3 n\right) - \frac{5}{2}\log\log\left(\varepsilon^3 n\right) + u \tag{12}$$

We do not show the dependence of these parameters on ε and n; u will be either a fixed constant, or a very slow function of n. We denote by $\mathrm{t} : \mathbb{R} \to \mathbb{R}$ the function defined by $t = \mathrm{t}(u)$ and by t^{-1} its inverse function. Write $\mathcal{C}_j$ for the j-th largest component of $G_n(p)$ as measured by its total number of vertices $|\mathcal{C}_j|$ (ties are broken in an arbitrary way). Key to the proof of Theorem 2.1 is the understanding of the following random measure:

$$N(dx) = \sum_{j \geqslant 0} \delta_{\mathrm{t}^{-1}(|\mathcal{C}_j|)}(dx) \tag{13}$$

and of its convergence in distribution in particular. Define $J = (u_1, u_2)$ with $-\infty < u_1 \leqslant u_2 < +\infty$, and let $\mathrm{t}(J) = (t_1, t_2)$. We first want to check the convergence of the random variable $N_-(J)$ (the minus index refers to the subcritical regime) as n tends to ∞. Let $\mathcal{C}(v)$ denote the component of $G_n(p)$ that contains the vertex v. Using the estimate $t_1 \sim t_2$ as $n \to \infty$, we find that the first moment of $N(J)$ satisfies:

$$\mathbb{E}(N(J)) = \mathbb{E}\left(\sum_{v \in G_n} \frac{\mathbf{1}_{\{|\mathcal{C}(v)| \in \mathrm{t}(J)\}}}{|\mathcal{C}(v)|}\right) \sim n\frac{\mathbb{P}\left(|\mathcal{C}(v)| \in (t_1, t_2)\right)}{t_1} \text{ as } n \to \infty \tag{14}$$

Another easy fact is that the tail $\mathbb{P}(|\mathcal{C}(v)| > t)$ is bounded from above by the quantity $\mathbb{P}(|T^m| > t)$, where T^{m} is a "modified" GW-tree with Bin $\left(\ell - \mathbf{1}_{\{v=\rho\}}, p\right)$ offspring distribution. The complementary lower bound is the difficult step. Different ideas can be developed to estimate it, let us review three different possibilities.

First, writing $T^{\mathrm{m}} = T^{\mathrm{m}}_{\ell}$ to show the dependence of T^{m} on ℓ, a step-by-step exploration of $\mathcal{C}(v)$ reveals that:

$$\mathbb{P}\left(\left|T^{\mathrm{m}}_{\ell-t}\right| \geqslant t\right) \leqslant \mathbb{P}(|\mathcal{C}(v)| \geqslant t) \leqslant \mathbb{P}\left(|T^{\mathrm{m}}_{\ell}| \geqslant t\right)$$

In case t is small with respect to ℓ, like in the complete graph K_n, $(\ell = n-1)$, the two bounds are equivalent sequences as $n \to \infty$. The computation of the successive moments of $|\mathcal{C}(v)|$ does not raise additional difficult, and one finds sharp asymptotics on $|\mathcal{C}_1|$. The same strategy works for "mean-field" model like the configuration model, see [Rio12] around formula (7.5). There are other classes of graphs for which a uniform control of the number of already explored vertices is possible, for instance the Hamming graph in dimension 2 [vdHL10]. Second, Nachmias [Nac09] introduced the idea of pruning off the upper bound tree T^{m} from the so-called path-impure vertices to find a sub-tree of T^{m} that is stochastically smaller than $|\mathcal{C}(v)|$. The path-impure vertices are the vertices that are present in the GW-tree but have no counterpart in the exploration of $\mathcal{C}(v)$. The set of path-impure vertices is denoted by $I^1\left(T^{\mathrm{m}}\right)$. A simple bound is, for $1 \leqslant t \leqslant t'$:

$$\mathbb{P}\left(|T^{\mathrm{m}}| \geqslant t', \left|I^1\left(T^{\mathrm{m}}\right)\right| \leqslant t' - t\right) \leqslant \mathbb{P}(|\mathcal{C}(v)| \geqslant t) \leqslant \mathbb{P}\left(|T^{\mathrm{m}}| \geqslant t\right) \tag{15}$$

Based on this bound, plus the second moment method, Nachmias derives in his Lemma 14 a lower bound on the probability that $|\mathcal{C}(v)|$ exceeds ε^{-2}, that is not sharp in the weak subcritical regime of interest to us (essentially because the method of proof relies on conditioning a GW-tree by the height). The third method is the one developed in this paper, it is based on conditioning the GW-trees by the size. Consider $J' = (u, \infty)$. From (14), the estimate (27) on the total progeny of a GW-tree gives that $\mathbb{E}\left(N_-\left(J'\right)\right) \leqslant \mathrm{e}^{-u}/(4\sqrt{\pi})$. One asks under what condition $(1 - o(1))\mathrm{e}^{-u}/(4\sqrt{\pi})$ is a lower bound. Examining (27) again, we see that if t' close to $t = \delta^{-1}s$ in the sense that $t' - t = o\left(\delta^{-1}\right)$ then the ratio $\mathbb{P}\left(T^{\mathrm{m}} \geqslant t'\right)/\mathbb{P}\left(T^{\mathrm{m}} \geqslant t\right)$ has limit 1 as n diverges. But the difference between $\mathcal{C}(v)$ and T^{m} is controlled by the set $I^1\left(T^{\mathrm{m}}\right)$ of path-impure vertices, so, by the first moment method, we only need to show $\mathbb{E}\left(\left|I^1\left(T^{\mathrm{m}}\right)\right| \,||T^{\mathrm{m}}| = t\right) = o\left(\delta^{-1}\right)$. The latter is proved in Proposition 3.8 under condition (7), and the proof of that Proposition is in turn based on a many-to-two formula, Lemma 3.1 with $k = 2$, that is indeed our key tool. Under (7) the lower bound matches the upper bound, and:

$$\mathbb{E}\left(N_-(J)\right) = (1 + o(1))\frac{1}{4\sqrt{\pi}}\left(\mathrm{e}^{-u_1} - \mathrm{e}^{-u_2}\right). \tag{16}$$

The rest of the proof is routine: the RHS is also the expectation of a Poisson random variable with parameter the integral over $J = (u_1, u_2)$ of the intensity measure (8). To claim the convergence in distribution through the method of moments, see *e.g.* Section 6.1 of [JLR11] it remains to check the convergence of the higher factorial moments $\mathbb{E}\left(\prod_{0 \leqslant i < k}\left(N_-(J) - i\right)\right)$, which require a last technical point: (16) has to be proved with $G \backslash G^0$ the graph induced by

G on $V(G)\backslash V\left(G^0\right)$, for $G^0 \subseteq G$ a subgraph with size $O(t)$. The next and last section contains the (self-contained) proof; it starts in Section 3.1 with the many-to- k formula. The required estimates on the size of the GW-trees of interest are stated in Section 3.2. In Section 3.3 GW-trees are randomly embedded in the graph G_n and the number of path-impure vertices in a GW-tree with a given size is estimated. The short subsection 3.4 is a remark on how to translate our estimates in term of modified GW-trees. We start to work with the $G_n(p)$ model itself only at Section 3.5 where we compute the moment of the number of components in a given interval, and also give the few additional ingredients needed to prove the main theorem. The extension to bounded degree graphs is discussed in Section 3.6. In Section 3.7 we obtain, as a by-product of our analysis in the weak subcritical regime, a lower bound on the expected number of components in certain intervals in the weak *supercritical* regime: yet we fail to get the lower bound on the size of the second largest component that is suggested by this estimate. In the last Section 3.8 we prove the Proposition 2.6 on condition (7).

3 Proof

The set of integers is denoted by $\mathbb{Z}$, the subset of non-negative integers $\{0,1,2,\ldots\}$ by $\mathbb{Z}^+$, the subset of positive integers $\{1,2,\ldots\}$ by $\mathbb{N}$ and the set of real numbers by $\mathbb{R}$. Unless explicitly advertised, all the limits and asymptotic are as n the size of the graph G_n goes to ∞. Also, we use the Landau notation o and O (but no more probabilistic counterpart from now on). Sums over an empty set are 0, and products over an empty set are 1. If $G=(V,E)$ is a graph, $|G|$ will denote its number of vertices, or size, of G. For $V^0 \subseteq V$, the graph induced by G on V^0 is the graph with vertex set V^0 and edge set the restriction of the original edge set E to $V^0 \times V^0$. Assume another graph $G'=(V',E')$ is given. A map $f: V \to V'$ is called a graph homomorphism when adjacent vertices in G are mapped to adjacent vertices in $G', (u,v) \in E \Rightarrow (f(u),f(v)) \in E'$ for each $u,v \in V$ It is called a graph isomorphism when f is a bijection from V to V' and two vertices are adjacent in G iff their images are adjacent in G', that is $(u,v) \in E$ iff $(f(u),f(v)) \in E'$.

We may need to attach several distinguished vertices to a graph, that we shall call the pointed vertices: for $k \in \mathbb{N}$, and $u_1,u_2,\ldots,u_k \in V, G=(V,E,u_1,u_2,\ldots,u_k)$ is a pointed graph, and if $G'=(V',E',u_1',u_2',\ldots,u_k')$ for $u_1',u_2',\ldots,u_k' \in V'$ is another pointed graph, we say a graph-homomorphism f from (V,E) to (V',E') preserves the pointed vertices when $f\left(u_i\right)=u_i'$ for $1 \leqslant i \leqslant k$. The trees we will encounter will be planar and rooted. Such trees are embedded in the so-called Ulam tree: this is the graph with vertex set the finite sequences of integers

$$\mathcal{U}=\bigcup_{n \geqslant 0} \mathbb{N}^n.$$

The root of the Ulam tree is the vertex $\mathbb{N}^0$, that we shall denote by ρ. A vertex distinct from the root, $u=(u(1),u(2),\ldots,u(k)) \in \mathcal{U}, k \geqslant 1$, has a unique father $a(u):=(u(1),u(2),\ldots,u(k-1))$, and there is one edge between every vertex distinct from the root and its father. Notice that the Ulam tree is a locally-infinite. The integer k is the generation of u, denoted by $|u|$. By convention, $|\rho|=0$. For $i \leqslant k$, define $u^{(i)}=(u(1),u(2),\ldots,u(i))$ the

ancestor of u at generation i. When u is an ancestor of v we write $u \preceq v$, and $u \prec v$ when also $u \neq v$; in the latter case, u is called a strict ancestor of v. $\preceq$ defines a partial order on $\mathcal{U}$ called the ancestral order. A subset $\mathbf{t}$ of $\mathcal{U}$ is a planar rooted tree when: (i) t contains ρ. (ii) $v = (v(1), \ldots, v(k-1), v(k)) \in \mathbf{t}$ implies $(v(1), \ldots, v(k-1), i) \in \mathbf{t}$ for any $i \in \{1, \ldots, v(k)\}$ (iii) $v = (v(1), \ldots, v(k)) \in \mathbf{t}$ implies $v^{(j)} = (v(1), \ldots, v(j)) \in \mathbf{t}$, for any $j \in \{1, \ldots, k\}$ A vertex $v \in \mathbf{t}$ is a leaf when its set of children $c_{\mathbf{t}}(v) := \{w \in \mathbf{t}, a(w) = v\}$ is empty. We write $c(v)$ for $c_{\mathbf{t}}(v)$ when $\mathbf{t}$ is clear from the context. The number $|c(v)|$ of children of $v \in \mathbf{t}$ is called the outdegree of v in t. We call $\mathcal{T}$ the set of planar rooted trees.

There is a second natural order defined on $\mathcal{U}$: the breadth-first order. We write $u \prec_{\text{bfs}} v$ when $|u| < |v|$, or $|u| = |v|$ and there exists $i < |u|$ such that $(j \leqslant i \Rightarrow u(j) = v(j))$ and $u(i+1) < v(i+1)$). We write $u \preceq_{\text{bfs}} v$ when $u \prec_{\text{bfs}} v$ or $u = v$. Unlike the ancestral order, the breadth-first order is a total order. Again, we shall not distinguish between the order $\preceq_{\text{bfs}}$ and its restriction to $\mathbf{t} \in \mathcal{T}$.

If k is an integer and $u_1, \ldots, u_k$ are k distinct vertices of $\mathbf{t} \in \mathcal{T}$ distinct from the root ρ, we call $\mathbf{t}_{\text{k}} = (\mathbf{t}, u_1, \ldots, u_k)$ a pointed planar rooted tree, and $\mathcal{T}_{\text{k}}$ the set of pointed planar rooted trees. Let us stress that the pointed vertices come in a specific order: $(\mathbf{t}, u_1, u_2)$ and $(\mathbf{t}, u_2, u_1)$ are for instance two distinct elements of $\mathcal{T}_2$. In case every vertex in $\mathbf{t}_{\text{k}} \in \mathcal{T}_{\text{k}}$ is an ancestor of a pointed vertex (possibly itself), we say that $\mathbf{t}_{\text{k}}$ is spanned by its pointed vertices, or simply that $\mathbf{t}_{\text{k}}$ is spanned. Notice that the set of pointed vertices of a spanned tree $\mathbf{t}_{\text{k}}$ contains its set of leaves. Two pointed planar rooted trees $\mathbf{t}_{\text{k}}, \mathbf{t}'_{\text{k}} \in \mathcal{T}_{\text{k}}$ are equivalent when there exists a graph-isomomorphism between $\mathbf{t}_{\text{k}}$ and $\mathbf{t}'_{\text{k}}$ that preserves the root and the pointed vertices. We call $\overline{\mathcal{T}^{\text{s}}_{\text{k}}}$ the set of equivalence classes of spanned pointed rooted trees (we emphasise that those trees are no more ordered.)

For $\mathbf{t}_{\text{k}} \in \mathcal{T}_{\text{k}}$ and $0 \leqslant i \leqslant k$, let $V_i \subseteq V$ be the set of vertices with precisely i children that are ancestors of pointed vertices: $V_i = \{v \in \mathbf{t}, |\{w \in c(v), \exists j, w \preceq u_j\}| = i\}$. In the special case $\mathbf{t}_{\text{k}} \in \overline{\mathcal{T}^{\text{s}}_{\text{k}}}$, every vertex is the ancestor of a pointed vertex, and V_i is the subset of the vertices with i children. The following is a partition of the set of vertices, $V = \bigcup_{0 \leq i \leqslant k} V_i$, and we set

$$\ell_i = |V_i| \,. \tag{17}$$

The definitions of ℓ_i and V_i extend to a tree $\mathbf{t}_{\text{k}} \in \overline{\mathcal{T}^{\text{s}}_{\text{k}}}$.

Let $\text{p} = (\text{p}_k, k \geqslant 0)$ be a distribution on the non negative-integers, and $(X_u, u \in \mathcal{U})$ be a collection of i.i.d. random variables indexed by $\mathcal{U}$ with distribution p. Call T the random tree in which the number of children of a vertex u is given by X_u in T, provided $X_v \neq 0$ for every strict ancestor v of u. The tree T is distributed as the GW-tree with offspring distribution p. Formally, $\rho \in T$ and for every $u \in \mathcal{U} \backslash \{\rho\}, u \in T$ iff for every $1 \leqslant i \leqslant |u|$

$$u(i) \leqslant X_{a(u^{(i)})}.$$

3.1 Many-to-k formula

Certain functions on $\mathcal{T}$ can be decomposed as a sum over the different k-tuples of the vertices of t. The expected value of such functions evaluated on GW-trees is then computed using a many-to- k formula (19).

Consider $\mathrm{p} = (\mathrm{p}_k, k \geqslant 0)$ a probability distribution on the set $\mathbb{Z}^+$. Let $k_{\max}(\mathrm{p}) = \max\left\{\ell \geqslant 0, \Sigma_k k^\ell \mathrm{p}_k < \infty\right\}$, and $k_0(\mathrm{p}) = \max\{k \geqslant 0, \mathrm{p}_k \neq 0\} \in \mathbb{Z}^+ \cup \{\infty\}$. If $i \in \mathbb{Z}^+$ satisfies $i \leqslant \min\{k_{\max}, k_0\}$, the i-th factorial moment and the i-th size-biased probability distribution $\mathrm{p}^{(i)} = \left(\mathrm{p}_k^{(i)}, k \geqslant 0\right)$ are defined by:

$$m_i = \sum_{k \geqslant i} \left[\prod_{0 \leqslant j < i} (k-j)\right] \mathrm{p}_k, \text{ and } \mathrm{p}_k^{(i)} = \frac{\prod_{0 \leqslant j < i}(k-j)}{m_i} \mathrm{p}_k \tag{18}$$

According to our convention that products over an empty set are $1, \mathrm{p}^{(0)} = \mathrm{p}$. Also, for $k < i, \mathrm{p}_k^{(i)} = 0$ since the product in the RHS of (19) contains a null factor. If $i > k_0(\mathrm{p})$ $m_i = 0$, and $p^{(i)}$ is not defined. Fix $\mathbf{t}_{\mathrm{k}} = (\mathbf{t}, v_1, \ldots, v_k) \in \overline{\mathcal{T}}_{\mathrm{k}}$ with $k \leqslant k_{\max}(\mathrm{p}), k_0(\mathrm{p})$. From p and t_{k}, we build $T_{\mathrm{k}} = T_{\mathrm{k}}(\mathrm{t}_{\mathrm{k}}) = (T, u_1, \ldots, u_k) \in \mathcal{T}_{\mathrm{k}}$ a random tree that contains t_{k} as a subtree. Formally, there is a copy [4] $T' \subseteq T$ of t embedded in T; for a vertex in T' the offspring distribution is $\mathrm{p}^{(i)}$ the i-th size-biased distribution, where i is the number of children of the corresponding vertex in t. For a vertex in $T \backslash T'$, the offspring distribution is $\mathrm{p}^{(0)} = \mathrm{p}$. A precise definition uses an induction: The trees T' and $T, T' \subseteq T$, and the graph isomorphism $\omega : \mathbf{t} \to T'$ are defined by the following steps:

- Initialization: The root ρ belongs to T and T', it is the image of ρ_{t} the root in t : $\rho = \omega(\rho_{\mathbf{t}})$.
- Induction 1 : If $u \in T'$, its number of children in T is distributed as $\mathrm{p}^{(i)}$ where $i = |c(v)|$ is the number of children of $v = \omega^{-1}(u)$ in $\mathbf{t}$. Furthermore, if $(vj, 1 \leqslant j \leqslant i)$ are the ordered children of v in $\mathbf{t}$, then $(\omega(vj), 1 \leqslant j \leqslant i)$ is a random sequence of i distinct children [5] of u in T, with uniform distribution. These i children of u belong to T', and the remaining children of u belong to $T \backslash T'$.
- Induction 2 : If $u \in T \backslash T'$, its number of children is distributed as $\mathrm{p} = \mathrm{p}^{(0)}$, all of them belong to $T \backslash T'$.
- Induction 3 : distinct vertices in T have an independent number of children.
- The pointed vertices in T_{k} are $(u_1, \ldots, u_k) = (\omega(v_1), \ldots, \omega(v_j))$

In the next formula, $m_i = m_i(\mathrm{p})$ is the i-th factorial moment of the distribution p, see (18), and $\ell_i = \ell_i(\mathbf{t}_{\mathrm{k}})$ simply counts the number of vertices with precisely i children since $\mathbf{t}_{\mathrm{k}} \in \mathcal{T}_{\mathrm{k}}^{\mathrm{s}}$ or $\mathbf{t}_{\mathrm{k}} \in \overline{\mathcal{T}_{\mathrm{k}}^{\mathrm{s}}}$, see (17).

Lemma 3.1 (Many-to-k formula)**.** *Let p be a distribution on $\mathbb{Z}^+$, and let $k \in \mathbb{N}$ satisfy $k \leqslant k_{\max}(p), k_0(p)$. For F a non-negative measurable function on the set of pointed trees $\mathcal{T}_k$,*

[4] T' does not belong to $\mathcal{T}$ since item (ii) in the definition of $\mathcal{T}$ is not satisfied, so the word "tree" that we use here is an abuse

[5] this is a.s. possible since the number of children of v in T, with distribution $\mathrm{p}^{(i)}$, is a.s. $\geqslant i$

and $T \in \mathcal{T}$ a GW-tree with offspring distribution p*, it holds:*

$$\mathbb{E}\left(\sum_{u_1,\dots,u_k} F\left(T, u_1, \dots, u_k\right)\right) = \sum_{\mathbf{t}_k} \left(\prod_{i \geqslant 1} m_i^{\ell_i}\right) \mathbb{E}\left(F\left(T_k\left(\mathbf{t}_k\right)\right)\right) \tag{19}$$

where:

- *the first sum is over the vertices $u_1, \dots, u_k$ of T such that no one is an ancestor of the other, and the second sum is over the trees $\mathbf{t}_k$ in $\overline{\mathcal{T}_k^s}$ that are spanned by k pointed leaves,*
- *or the first sum is over the pairwise distinct vertices $u_1, \dots, u_k$ of T, and the second sum is over the trees $\mathbf{t}_k$ in $\overline{\mathcal{T}_k}^s$*

The so-called many-to-one formula is the $k = 1$ case. The generation of the leaf uniquely specifies a tree in $\overline{\mathcal{T}_1^{\mathrm{s}}}$, which identifies the latter set with the set of non-negative integers $\mathbb{Z}^+$, while the product on the RHS reduces to $m_1^{\ell_1} = m_1^{h_1}$.

The restriction to $k \leqslant k_0(\mathrm{p})$ is for the sake of simplicity: in case $k > k_0(\mathrm{p})$, the correct formula is obtained by discarding those trees $\mathbf{t}_{\mathrm{k}}$ that have vertices with outdegree larger than k_0 in the sums (i) and (ii).

The sequence $(\ell_i, i \geq 1)$ has finitely many non-null terms, hence the product $\left(\prod_{i \geqslant 1} m_i^{\ell_i}\right)$ is well defined.

Proof. Let $\mathbf{t}_{\mathrm{k}}^0 = (\mathbf{t}^0, v_1, \dots, v_k) \in \mathcal{T}_{\mathbf{k}}$ be a pointed planar rooted tree. We first check (19) for $F = 1_{\mathbf{t}_{\mathrm{k}}^\circ}$. The subtree of $\mathbf{t}_{\mathrm{k}}^0$ spanned by $v_1, \dots, v_k$ is denoted $\mathbf{t}_k \in \overline{\mathcal{T}_k^{\mathrm{s}}}$. For $i \geqslant 0$, set $V_i^0 = V_i\left(\mathbf{t}_{\mathrm{k}}^0\right), \ell_i^0 = \ell_i\left(\mathbf{t}_{\mathrm{k}}^0\right)$ and $\ell_i = \ell_i\left(\mathbf{t}_{\mathrm{k}}\right)$. For $i \geqslant 1$, notice that $\ell_i^0 = \ell_i$. Using the definition (18) of $\mathrm{p}^{(i)}$, we find:

$$\begin{aligned}
\mathbb{P}\left(T = \mathbf{t}^0\right) &= \prod_{u \in V(\mathbf{t}^0)} \mathrm{p}_{|c(u)|} \\
&= \left(\prod_{1 \leqslant i \leqslant k_0} m_i^{\ell_i}\right) \prod_{0 \leqslant i \leqslant k_0} \prod_{u \in V_i^0} \frac{1}{\prod_{0 \leqslant j < i}(|c(u)| - j)} \mathrm{p}_{|c(u)|}^{(i)} \qquad (20) \\
&= \left(\prod_{1 \leqslant i \leqslant k_0} m_i^{\ell_i}\right) \mathbb{P}\left(T_{\mathrm{k}}\left(\mathbf{t}_{\mathrm{k}}\right) = \mathbf{t}_{\mathrm{k}}^0\right)
\end{aligned}$$

Formula (20) is formula (19) with the choice $F = 1_{\mathrm{t}_{\mathrm{k}}^\circ}$, since one single tree, the tree t_{k} defined above, contributes to the sum on the RHS of (19). A function F on $\mathcal{T}_{\mathrm{k}}$ may be decomposed as a sum of indicator functions, and relation (19) is linear in F : Formula (20) therefore implies (19), with the sum as in (ii). Adding the restriction that, in the collection $v_1, \dots, v_k$, no one is an ancestor of another is equivalent to summing over the trees t_{k} spanned by pointed leaves, giving (i). □

3.2 Asymptotics for the size of GW-trees

We restrict in this section our views to a particular class of GW-trees. These are, for $\varepsilon > 0$ and $\ell \geqslant 2$ an integer

$$T = T_{\pm} \in \mathcal{T} \text{ the GW-tree with } \mathrm{Bin}\,(\ell - 1, p_{\pm}) \text{ offspring distribution.} \tag{21}$$

Although the notation does not show it, T depends on ε and ℓ. The next Lemma bounds the probability that a forest of GW-trees distributed as (21) has a given size in term of the probability that a single GW-tree has that same size. We stress the statement is valid for ε fixed (i.e., condition (4) is not assumed).

Lemma 3.2. *Let $(T_i, i \geqslant 1)$ be independent GW-trees distributed as $T = T_{\pm} \in \mathcal{T}$ the GW-tree in (21). Let $c < 1/2$. There exists $\ell_0 = \ell_0(c)$ such that, for $\ell \geqslant \ell_0$, and for $1 \leqslant h \leqslant j$:*

$$\mathbb{P}\left(\sum_{1 \leqslant i \leqslant h} |T_i| = j\right) \leqslant h \left(\frac{1-p}{1 \pm \varepsilon}\right)^{h-1} \mathrm{e}^{-c\frac{h(h-1)}{j}} \mathbb{P}(|T| = j) \tag{22}$$

Remark 3.3. For *critical* GW-trees with a finite variance, similar bounds may be deduced from the estimates in Addario-Berry, Devroye and Janson [ABDJ13], see formulae (17) and (19) in this article.

Proof. Let $(Z_t, t \geqslant 0)$ be a random walk started at $Z_0 = 0$ with independent increments distributed as $X - 1, X$ a $\mathrm{Bin}(\ell - 1, p)$ random variable. For $h \in \mathbb{N}$, we let $H_{-h}(Z) = \inf\{j \geqslant 1, Z_j = -h\}$ be the hitting time of $-h$. Also, let $T \in \mathcal{T}$ be a GW-tree with offspring distribution the distribution of X. If $\rho = v_1, \ldots, v_{|T|}$ is the sequence of the vertices of T arranged in breadth-first order, then there is the identity in distribution

$$(Z_j, 1 \leqslant j \leqslant H_{-1}(Z)) = \left(\Big|\bigcup_{i \leqslant j} \{c\,(v_i)\} \setminus \bigcup_{i \leqslant j} \{v_i\}\Big|, 1 \leqslant j \leqslant |T\,|\right) \tag{23}$$

where the RHS counts the number of vertices among the children of the vertices $v_1 \ldots v_j$ that do not belong to $v_1 \ldots v_j$. The identity (23) implies in particular:

$$|T| = H_{-1}(Z) \tag{24}$$

which extends in a straightforward way to a collection of $h \geqslant 1$ independent GW-trees $(T_i, 1 \leqslant i \leqslant h)$ distributed as T :

$$\sum_{1 \leqslant i \leqslant h} |T_i| = H_{-h}(Z)$$

We also use the following combinatorial identity, known as Spitzer lemma [Spi56], that connects the distribution of the hitting time of a random walk with the marginal distribution of that random walk [6]

$$j\mathbb{P}\,(|H_{-h}(Z)| = j) = h\mathbb{P}\,(Z_j = -h)$$

[6] the proof only requires invariance by cyclic shift of the distribution of the increments of the random walk, see *e.g.* Pitman [Pit02]

We deduce from this and the convolution property of the Binomial distribution that:

$$\begin{aligned}\mathbb{P}\left(\sum_{1\leqslant i\leqslant h}|T_i|=j\right) &= \frac{h}{j}\mathbb{P}\left(Z_j=-h\right)\\ &= \frac{h}{j}\mathbb{P}(\mathrm{Bin}((\ell-1)j,p)=j-h)\\ &= \frac{h}{j}\binom{(\ell-1)j}{j-h}p^{j-h}(1-p)^{(\ell-1)j-(j-h)} \qquad (25)\end{aligned}$$

Expanding the binomial coefficient in the last expression allows to relate the above probability to the same probability when h is set to $1, \mathbb{P}\left(|T_1|=j\right)$:

$$\begin{aligned}h\left(\frac{1-p}{p}\right)^{h-1}\frac{\prod_{1\leqslant i<h}(j-i)}{\prod_{j-h\leqslant i<j-1}(\ell-1)j-i}\cdot\left(\frac{1}{j}\frac{\prod_{0\leqslant i<j-1}(\ell-1)j-i}{(j-1)!}p^{j-1}(1-p)^{(\ell-1)j-(j-1)}\right)\\ = h\left(\frac{1-p}{p}\right)^{h-1}\frac{\prod_{1\leqslant i<h}(j-i)}{\prod_{j-h\leqslant i<j-1}(\ell-1)j-i}\mathbb{P}\left(|T_1|=j\right)\end{aligned}$$

Using the definition of $p=p_\pm$ in (4), one can estimate the prefactor:

$$\begin{aligned}&h\left(\frac{1-p}{p}\right)^{h-1}\frac{\prod_{1\leqslant i<h}(j-i)}{\prod_{j-h\leqslant i<j-1}(\ell-1)j-i}\\ &= h\left(\frac{1-p}{1\pm\varepsilon}\right)^{h-1}\frac{\prod_{1\leqslant i<h}\left(1-\frac{i}{j}\right)}{\prod_{j-h\leqslant i<t-1}\left(1-\frac{i}{j(\ell-1)}\right)} \qquad (26)\\ &= h\left(\frac{1-p}{1\pm\varepsilon}\right)^{h-1}\mathrm{e}^{-\frac{h(h-1)}{2j}\left(1-O\left(\frac{1}{\ell}\right)\right)}\end{aligned}$$

For ℓ large enough, $O\left(\frac{1}{\ell}\right)<1-2c$, and the inequality (22) is proved.

□

Another key ingredient is a precise estimate on the tail of the size of a single tree T. Such estimates generally follow from two ingredients: a local central limit theorem on the random walk, plus a centering (or tilting) operation, see Riordan [Rio12] for an implementation of that combination. Our example allows for direct computations. Unlike the previous Lemma, we now assume that ε and ℓ depend on n in a way specified by conditions (3) and (4)

Proposition 3.4. *Let $T=T_\pm\in\mathcal{T}$ be the GW-tree given by (21), with sequences ε and ℓ satisfying (4) and (3) respectively. Let v be a sequence that satisfies $v_n=o\left(\log\log\left(\varepsilon^3 n\right)\right)$. Then, for $t=t(u)$ given by (12), it holds that:*

$$\frac{n}{t}\mathbb{P}(t\leqslant|T|<\infty)=(1+o(1))\frac{1}{4\sqrt{\pi}}\mathrm{e}^{-u}\ \text{as}\ n\to\infty \qquad (27)$$

with $o(1)$ uniform over the sequences $u=(u_n)$ such that $|u_n|\leqslant v_n, n\in\mathbb{N}$.

Under (4), the sequence $(\varepsilon^3 n)_{n\geqslant 1}$ diverges and the condition $|u_n| \leqslant v_n$ allows for constant sequences (u_n). Also, the restriction to finite trees on the LHS of (27) is necessary in the case of supercritical GW-trees only, since subcritical GW-trees are a.s. finite. Last, (27) entails $|T_+|$ and $|T_-|$ have equivalent tail (not equal, though).

Proof. We first consider a given GW-tree, associated with a fixed n and ℓ. The symbol $o_j(1)$ stand for a sequence with null limit as $j \to \infty$. Formula (25) with $h = 1$ reads:

$$\mathbb{P}(|T| = j) = \frac{1}{j}\binom{(\ell-1)j}{j-1} p^{j-1}(1-p)^{(\ell-1)j-(j-1)} \tag{28}$$

Stirling formula under the form

$$j! = (1 + o_j(1))\,\sqrt{2\pi} j^{j+1/2} \mathrm{e}^{-j}$$

allows to approximate the Binomial coefficient and to estimate:

$$\begin{aligned}
\mathbb{P}(|T| = j) &= \frac{1}{j}\frac{((\ell-1)j)!}{(j-1)!((\ell-2)j+1)!} p^{j-1}(1-p)^{(\ell-1)j-(j-1)} \\
&= (1+o_j(1))\,\frac{1}{j}\frac{1}{\sqrt{2\pi}}\frac{((\ell-1)j)^{(\ell-1)j+1/2}}{(j-1)^{j-1/2}((\ell-2)j+1)^{(\ell-2)j+3/2}} p^{j-1}(1-p)^{(\ell-1)j-(j-1)} \\
&= (1+o_j(1))\,\frac{1}{j}\frac{1}{\sqrt{2\pi}}\left[\frac{j^{j-1}}{(j-1)^{j-1/2}}\left(\frac{(\ell-1)j}{(\ell-2)j+1}\right)^{(\ell-2)j+3/2}\right](1\pm\varepsilon)^{j-1}(1-p)^{(\ell-2)j+1}
\end{aligned}$$

The term under bracket requires some care, and may be evaluated using:

$$\begin{aligned}
&\tfrac{j^{j-1}}{(j-1)^{j-1/2}}\left(\tfrac{(\ell-1)j}{(\ell-2)j+1}\right)^{(\ell-2)j+3/2} \\
&= j^{-1/2}\left(\tfrac{j}{j-1}\right)^{j-1/2}\left(\tfrac{(\ell-2)j}{(\ell-2)j+1}\right)^{(\ell-2)j+3/2}\left(\tfrac{(\ell-1)j}{(\ell-2)j}\right)^{(\ell-2)j+3/2} \\
&= j^{-1/2}\left(1+\tfrac{1}{j-1}\right)^{j-1/2}\left(1-\tfrac{1}{(\ell-2)j+1}\right)^{(\ell-2)j+3/2}\left(1+\tfrac{1}{\ell-2}\right)^{(\ell-2)j+3/2} \\
&= j^{-1/2}\left(\mathrm{e}^1+o_j(1)\right)\left(\mathrm{e}^{-1}+o_j(1)\right)\left(1+\tfrac{1}{\ell-2}\right)^{(\ell-2)j+3/2}
\end{aligned}$$

This entails

$$\begin{aligned}
\mathbb{P}(|T| = j) &= (1+o_j(1))\,\frac{1}{\sqrt{2\pi}}\frac{1}{j^{3/2}}\left(1+\frac{1}{\ell-2}\right)^{(\ell-2)j+3/2}(1\pm\varepsilon)^{j-1}(1-p)^{(\ell-2)j+1} \\
&= (1+o_j(1))\,q_1\frac{1}{\sqrt{2\pi}}\frac{1}{j^{3/2}}q_2^j
\end{aligned} \tag{29}$$

with the notation

$$q_1 = \left(1+\frac{1}{\ell-2}\right)^{3/2}(1\pm\varepsilon)^{-1}(1-p) \text{ and } q_2 = \left(1+\frac{1}{\ell-2}\right)^{\ell-2}(1\pm\varepsilon)(1-p)^{\ell-2}$$

We now recognise that

$$q_2 = \mathrm{e}^{-\delta}, \text{ with } \delta = \delta(\ell-1,\varepsilon) \text{ defined in (5)} \tag{30}$$

by the following computation:

$$\begin{aligned}
\left(1+\frac{1}{\ell-2}\right)(1-p) &= \left(1+\frac{1}{\ell-2}\right)\left(1-\frac{1\pm\varepsilon}{\ell-1}\right) \\
&= 1+\frac{1}{\ell-2}-\frac{1\pm\varepsilon}{\ell-1}-\frac{1\pm\varepsilon}{(\ell-1)(\ell-2)} \\
&= 1+\frac{(\ell-1)-(1\pm\varepsilon)(\ell-2)}{(\ell-1)(\ell-2)}-\frac{1\pm\varepsilon}{(\ell-1)(\ell-2)} \\
&= 1+\frac{(\ell-1)(1-(1\pm\varepsilon))+(1\pm\varepsilon)}{(\ell-1)(\ell-2)}-\frac{1\pm\varepsilon}{(\ell-1)(\ell-2)} \\
&= 1-\frac{\pm\varepsilon}{\ell-2}
\end{aligned}$$

Summing the equivalents in (29) we find that:

$$\mathbb{P}(|T|\geqslant j) = (1+o(1))\frac{q_1}{\sqrt{2\pi}}\sum_{i\geqslant j} i^{-3/2}\mathrm{e}^{-\delta i}$$

We now set $j=t$ and assume the parameters ε and ℓ depend on n, with $\varepsilon\to 0$ and $\ell\to\infty$. The next $o(1)$ are $o_n(1)$ as usual. Plainly, $q_1 = 1+o(1)$. The map $x\mapsto x^{-3/2}\mathrm{e}^{-\delta x}$ is non increasing on $[t,\infty)$, this implies:

$$\frac{1}{\sqrt{2\pi}}\sum_{i\geqslant t} i^{-3/2}\mathrm{e}^{-\delta i} = (1+o(1))\frac{1}{\sqrt{2\pi}}\int_{x\geqslant t} x^{-3/2}\mathrm{e}^{-\delta x}dx$$

Integrating the latter expression we find that

$$\mathbb{P}(t\leqslant |T|<\infty) = (1+o(1))\frac{1}{\sqrt{2\pi}}\delta^{-1}t^{-3/2}\mathrm{e}^{-\delta t} \tag{31}$$

The upper bound follows from the bound $x^{-3/2}\leqslant t^{-3/2}$ in the integrand, and a lower bound is given by the same integral over $[t,t']$ instead of $[t,\infty)$. Then we choose t' in such a way that $\delta\,(t'-t)\to\infty$ with $t' = (1+o(1))t$. To complete the proof, it remains to expand $t=\delta^{-1}s$:

$$\frac{n}{t}\delta^{-1}t^{-3/2}\mathrm{e}^{-\delta t} = \frac{n}{\delta^{-1}s}\delta^{-1}\left(\delta^{-1}s\right)^{-3/2}\mathrm{e}^{-\delta\left(\delta^{-1}s\right)} = \frac{n\delta^{3/2}}{s^{5/2}}\mathrm{e}^{-s}$$

and then $s=s(u)$, using also $\delta = (1+o(1))\varepsilon^2/2$ from (5) (recall $\varepsilon\to 0$ and $\ell\to\infty$) as well as $\varepsilon^3 n\to\infty$

$$\begin{aligned}
&2^{-3/2}\frac{\varepsilon^3 n}{\left(\log(\varepsilon^3 n)-\frac{5}{2}\log(\log(\varepsilon^3 n))+u\right)^{5/2}}\mathrm{e}^{-\left(\log\left(\varepsilon^3 n\right)-\frac{5}{2}\log\left(\log\left(\varepsilon^3 n\right)\right)+u\right)} \\
&= \left(2^{-3/2}+o(1)\right)\mathrm{e}^{-u}
\end{aligned}$$

with $o(1)$ in the last expression uniform in $|u|\leqslant v$ with $v = o\left(\log\log\left(\varepsilon^3 n\right)\right)$. Bearing in mind the $1/\sqrt{2\pi}$ factor in (31), the Proposition is proved. □

Remark 3.5 (Duality of GW-trees.). Let $p_+ = (1+\varepsilon)/(\ell-1)$ and $\bar{p}_- = (1-\bar{\varepsilon})/(\ell-1)$. We have, from (5), that $p_+ (1-p_+)^{\ell-2} = \bar{p}_- (1-\bar{p}_-)^{\ell-2}$ if and only if $\delta_+(\varepsilon, \ell-1) = \delta_-(\bar{\varepsilon}, \ell-1)$ For a fixed $\varepsilon > 0$, the latter equation has a unique solution $\bar{\varepsilon} > 0$, from the definition of the function $\delta_\pm$. (28) then implies that, for the GW-trees T_+ and $\bar{T}_-$ respectively associated with p_+ and $\bar{p}_-$ as above:

$$(T_+ \mid |T_+| < \infty) \text{ is distributed as } \bar{T}_-$$

It is a general fact that supercritical GW-trees conditioned on being finite have the same distribution as certain subcritical GW-trees. The specificity of the Binomial GW-trees is that the corresponding subcritical GW-tree again are Binomial GW-trees.

We notice the following stability property of the $\mathrm{Bin}(\ell-1, p)$ distribution under size-biaising.

Lemma 3.6. *Let $\ell \in \mathbb{N}, i \in \{0, 1, \ldots, \ell-1\}$, and $p \in (0,1]$. If* p *is the* $\mathrm{Bin}(\ell, p)$ *distribution, then* $\left(\mathrm{p}^{(i)}_{i+k}, k \in \mathbb{Z}^+\right)$ *is the* $\mathrm{Bin}(\ell-i, p)$ *distribution.*

Proof. The generating function of p the $\mathrm{Bin}(\ell, p)$ distribution is given by:

$$\sum_{k \geqslant 0} s^k \mathrm{p}_k = (1 - p(1-s))^\ell$$

The i-th factorial moment of p is:

$$m_i = p^i \prod_{0 \leqslant j < i} (\ell - j)$$

and the size-biased distribution $\mathrm{p}^{(i)}$ satisfies:

$$\begin{aligned} \mathrm{p}^{(i)}_k &= \frac{\prod_{0 \leqslant j < i}(k-j)}{p^i \prod_{0 \leqslant j < i}(\ell - j)} \left(\begin{array}{c} \ell \\ k \end{array} \right) p^k (1-p)^{\ell-k} \\ &= \left(\begin{array}{c} \ell - i \\ k - i \end{array} \right) p^{k-i} (1-p)^{(\ell-i)-(k-i)} \end{aligned}$$

Therefore $\left(\mathrm{p}^{(i)}_{i+k}, k \in \mathbb{Z}^+\right)$ is the $\mathrm{Bin}(\ell-i, p)$ distribution. Consider a tree $\mathbf{t}_\mathrm{k} \in \overline{\mathcal{T}_\mathrm{k}}^\mathrm{s}$, and the corresponding random tree $T_\mathrm{k}(\mathbf{t}_\mathrm{k})$. The next Lemma estimates the number of vertices in the latter tree, that is the random variable $|T_\mathrm{k}(\mathbf{t}_\mathrm{k})|$. We set $h+1 := |\mathbf{t}_\mathrm{k}|$. Recall the definition of ℓ_i in (17). We have

$$h = \sum_{i \geqslant 1} i \ell_i \tag{32}$$

since both sides count the number of non-root vertices in $\mathbf{t}_\mathrm{k}$. □

Lemma 3.7. *Let $T = T_\pm \in \mathcal{T}$ be the GW-tree in (21). Let $j \geqslant 3$, and $c < 1/2$. There exists $\ell_0 = \ell_0(c)$ such that, for $\ell \geqslant \ell_0$*

$$\mathbb{P}\left(|T_\mathrm{k}(\mathbf{t}_\mathrm{k})| = j\right) \leqslant (h+1)(1 \pm \varepsilon)^{-h} \mathrm{e}^{-c\frac{h(h+1)}{j}} \mathbb{P}(|T| = j) \tag{33}$$

Proof. We first compare $\mathrm{p}^{(i)}$ and p, for p the $\mathrm{Bin}(\ell-1,p)$ distribution. Using Lemma 3.6, and expanding the binomial coefficient, we find, for $1 \leqslant k \leqslant \ell - i$, that:

$$\begin{aligned}
\mathrm{p}^{(i)}_{k+i} &= \binom{\ell-1-i}{k} p^k (1-p)^{\ell-1-i-k} \\
&= \frac{\prod_{0\leqslant j<i}(\ell-1-j-k)}{\prod_{0\leqslant j<i}(\ell-1-j)} \cdot \frac{1}{(1-p)^i} \cdot \binom{\ell-1}{k} p^k (1-p)^{\ell-1-k} \\
&\leqslant \frac{1}{(1-p)^i} \cdot \mathrm{p}_k \qquad (34)
\end{aligned}$$

where the inequality follows by bounding the first factor of the product by 1.

Let $\mathbf{t}_k = (\mathbf{t}, u_1, \ldots, u_k) \in \mathcal{T}_{\mathrm{k}}$. Denote by $(v_j)_{1\leqslant j\leqslant h+1}$ the ancestors (large or strict) of the k pointed vertices (including the pointed vertices themselves) in $\mathbf{t}_k$ ranked by breadth-first order. (This definition of h is consistent with the one in (32)). Consider, for $i \in \{1,\ldots,h+1\}, U_i$ the set of vertices of $\mathbf{t}_k$ whose most recent common ancestor in $\{(v_j)_{1\leqslant j\leqslant h+1}\}$ is v_i, and define $\mathbf{t}^{[i]} \in \mathcal{T}$ the tree induced by $\mathbf{t}$ on U_i and rooted at v_i. We apply this construction to $T_{\mathrm{k}}(\mathrm{t}_{\mathrm{k}}) \in \mathcal{T}_{\mathrm{k}}$, shortened in T_{k} in the following lines. Recall T' is the random tree embedded in T_{k}. Conditionally on T', the random tree $T^{[i]}_{\mathrm{k}}$ is a GW-tree with offspring distribution p, except for the root that has offspring distribution $\mathrm{p}^{(j)}_{j+\cdot}$, for $j = |c_{T'}(v_i)|$ the number of children of v_i in T'. With the help of (34), we see that:

$$\mathbb{P}\left(T^{[i]}_{\mathrm{k}} = \mathbf{t}\right) \leqslant (1-p)^{-|c_{T'}(v_i)|}\mathbb{P}(T=\mathbf{t})$$

Now, the trees $T^{[i]}_{\mathrm{k}}$ are independent. Let $(T_i)_{1\leqslant i\leqslant h+1}$ be a collection of independent trees distributed as T. The latter identity and $\sum_{1\leqslant i\leqslant h+1}|c_{T'}(v_i)| = h$ together imply that

$$\mathbb{P}(|T_{\mathrm{k}}| = j) = \mathbb{P}\left(\sum_{1\leqslant i\leqslant h+1}|T^{[i]}_{\mathrm{k}}| = j\right) \leqslant (1-p)^{-h}\mathbb{P}\left(\sum_{1\leqslant i\leqslant h+1}|T_i| = j\right)$$

Combined with (22) and the inequality $1-p \leqslant 1$, the latter gives (33) □

3.3 The number of path-impure vertices in one large GW-tree.

Let G_n be a vertex transitive graph on n vertices with a pointed vertex called the root, and let $\mathbf{t} \in \mathcal{T}$. Call ℓ the common degree of the vertices in G_n, and assume that the number of children of every vertex in $\mathbf{t}$ is $\leqslant \ell - 1$, except the root of $\mathbf{t}$ that may have up to ℓ children. Conditionally on $\mathbf{t}$, we first define $\iota : \mathbf{t} \to G_n$ a random graph homomorphism by induction:

- $\iota(\rho)$ is a random vertex x in G_n
- If $v \in \mathrm{t}$ has i children, denoted by $v1, \ldots, vi$, then $(\iota(vj), 1 \leqslant j \leqslant i)$ is a random, uniformly distributed, sequence of i distinct elements of the set of neighbours of $\iota(v)$ in G_n, also distinct from $\iota(a(v))$ in case $v \neq \rho$.

- The vertices of $\mathbf{t}$ are seen in the breadth-first order.

Conditionally on $\mathbf{t}$ and ι, a vertex $w \in \mathbf{t}$ is called *impure* if there exists a vertex v smaller than w in the *breadth-first* order, $v \prec_{\text{bfs}} w$, such that $\iota(v) = \iota(w)$. In that case, say that v makes w impure; also, a vertex $w \in \mathbf{t}$ is called *path-impure* if it has an ancestor in $\mathbf{t}$ that is impure. This means that there exists $v \preceq w$, and a vertex $u \prec_{\text{bfs}} v$ such that $\iota(u) = \iota(v)$; say that u makes w path-impure in this case. We denote by $I^1(\mathbf{t})$ the subset of path-impure vertices of $\mathbf{t}$.

Fix a graph $G^0 \subseteq G = G_n$. Define $G \backslash G^0$ the graph induced by G on the vertex set $V(G) \backslash V(G^0)$. Conditionally on $\mathbf{t}$ and ι, a vertex of $\mathbf{t}$ is called G^0-impure if it is mapped by ι to a vertex in G^0, and is called G^0-path-impure if it has an ancestor (strict or large) in t that is G^0-impure. We denote by $I^0(\mathbf{t})$ the subset of G^0-path-impure vertices of $\mathbf{t}$.

For v a vertex of G, we let $\mathcal{C}^0(v)$ be the component that contains v in the percolation of $G \backslash G^0$. We set, for $\mathbf{t} \in \mathcal{T}$,

$$I(\mathbf{t}) = I^0(\mathbf{t}) \cup I^1(\mathbf{t}) \tag{35}$$

We shall consider successively the expected number of path-impure vertices and of G^0- path-impure vertices in a large GW-tree. To bound the expected number of path-impure vertices, we need the special case $k = 2$ in formula (19) (many-to-two formula).

Proposition 3.8. *Let $T \in \mathcal{T}$ be the GW-tree given by (21). Assume ε satisfies (4) and ℓ satisfies (3). Let also $I^1 = I^1(T)$ be the subset of the path-impure vertices of T, and let P^k be the kernel of the non-backtracking random walk on G_n. Let $c < 1/2$. There exists $\ell_0 = \ell_0(c)$ such that, for $\ell \geqslant \ell_0$ and for any $j \geqslant 1$*

$$\mathbb{E}\left(\left|I^1(T)\right| \,\middle|\, |T| = j\right) \leqslant \frac{\pi^{1/2}}{2^{5/2} c^{3/2}} \left(\sum_{k \geqslant 3} k e^{-c\frac{k^2}{j}} P^k \right) j^{3/2} \tag{36}$$

Notice that, for the RHS of (36) to be $o(j)$, we need the term in parenthesis to be $o\left(j^{-1/2}\right)$ We first fix some notation. Let $\mathbf{t}_2 \in \overline{\mathcal{T}_2}^{\text{s}}$ be a tree spanned by 2 pointed vertices. We denote by h_0 the generation of the most common ancestor of the two pointed vertices, and by $h_0 + h_1$ and $h_0 + h_2$ the generations of the two pointed vertices, with $h_0 + h_1 \leqslant h_0 + h_2$ In this way, the triplet $\mathrm{h} = (h_0, h_1, h_2)$ uniquely defines a tree $\mathbf{t}_2 \in \overline{\mathcal{T}_2}^{\text{s}}$, and we abuse notation by writing $T_2(\mathrm{h})$ for $T_2(\mathrm{t}_2)$ the associated GW-tree in this case. Let us point that the number h of non-root vertices in $\mathbf{t}_2$ then satisfies:

$$h = h_0 + h_1 + h_2 \tag{37}$$

Proof. Let $j \in \mathbb{N}$. We use the many-to-two formula in Lemma 3.1 with the index of summation (ii). For v, w distinct vertices of T, the choice

$$F(T, v, w) = \mathbf{1}\{w \text{ makes } v \text{ path-impure in } T, |T| = j\}$$

allows to estimate the size of $I^1(T)$ the subset of the path-impure vertices in the GW-tree T. (The function F is, through ι, a random function, but the many-to-two formula still holds

for such a function.) Recall the notation $T_2(\mathrm{h}) = (T(\mathrm{h}), u_1, u_2)$ and the notation u_0 for the most recent common ancestor of u_1 and u_2. Recall the definition of h in (37) and the equality (32) on $\ell_1 + 2\ell_2$. We point out that $m_1 = p(\ell-1) = 1 \pm \varepsilon$ and $m_2 = p^2(\ell-1)(\ell-2) \leqslant (p(\ell-1))^2 \leqslant (1\pm\varepsilon)^2$, so the product that appears in (19) satisfies

$$\prod_{1\leqslant i\leqslant k_0} m_i^{\ell_i} = m_1^{\ell_1} m_2^{\ell_2} \leqslant (1\pm\varepsilon)^{\ell_1+2\ell_2} = (1\pm\varepsilon)^h. \tag{38}$$

Recall $u_1^{(i)}$ denotes the ancestor of u_1 at generation i. Fix j and apply (19) to find:

$$\begin{aligned}
\mathbb{E}\left(\left|I^1\right|, |T| = j\right) &\leqslant \mathbb{E}\left(\sum_{v\neq w\in T\setminus\{\rho\}} F(T,v,w)\right) \qquad (39)\\
&\leqslant \sum_{h_0,h_1,h_2} (1\pm\varepsilon)^h \mathbb{P}\left(u_2 \text{ makes } u_1 \text{ path-impure in } T_2(h), |T_2(\mathrm{h})| = j\right)\\
&\leqslant \sum_{h_0,h_1,h_2,i} (1\pm\varepsilon)^h \mathbb{P}\left(u_2 \prec_{\mathrm{bfs}} u_1^{(i)}, \iota(u_2) = \iota(u_1^{(i)}), |T_2(\mathrm{h})| = j\right)\\
&\leqslant \sum_{h_0,h_1,h_2,i} (1\pm\varepsilon)^h \mathbf{1}_{\{h_0+h_2\leqslant i\leqslant h_0+h_1\}} \mathbb{P}\left(\iota(u_2) = \iota(u_1^{(i)}), |T_2(\mathrm{h})| = j\right)\\
&= \sum_{h_0,h_1,h_2,k} (1\pm\varepsilon)^h \mathbf{1}_{\{h_2+h_2\leqslant k\leqslant h_2+h_1\}} P^k \mathbb{P}\left(|T_2(\mathrm{h})| = j\right). \qquad (40)
\end{aligned}$$

At the third line, we use the definition of path-impurity of u_1 in term of its ancestors. For any two vertices u and v, $u \prec_{\mathrm{bfs}} v$ implies $|u| \leqslant |v|$, whence the inequality at the fourth line. We set $k = h_2 + i - h_0$ at the fifth line, k is the graph distance between the vertices u_2 and $u_1^{(i)}$. Also we use that, for two vertices of $T_2(h)$ that are mapped to the same vertex of G_n by ι, the image in G_n of the unique path in $T_2(h)$ between these vertices is distributed as a loop of the non-backtracking random walk: this follows by construction of ι. We now fix k, and, motivated by (33), compute a sum over h_0, h_1, h_2 in (40): Fix $c > 0$ an arbitrary positive number, and set

$$A = \sum_{h_0,h_1,h_2} \mathbf{1}_{\{h_2+h_2\leqslant k\leqslant h_2+h_1\}} (h_0+h_1+h_2)\, \mathrm{e}^{-c\frac{(h_0+h_1+h_2)^2}{j}}.$$

First we can sum over $h_0 \geqslant 1$, using a simple comparison with an integral. This leaves

$$A \leqslant \frac{j}{2c} \sum_{h_1,h_2} \mathbf{1}_{\{h_2+h_2\leqslant k\leqslant h_2+h_1\}} \mathrm{e}^{-c\frac{(h_1+h_2)^2}{j}}.$$

There remains two sums to perform. Set $i = h_1 + h_2$; due to the restriction, $h_2 + h_2 \leqslant k$ a given value of i appears at most $k/2$ times:

$$A \leqslant \frac{j}{2c} \sum_{h_1,h_2} \mathbf{1}_{\{h_2+h_2\leqslant k\leqslant h_2+h_1\}} \mathrm{e}^{-c\frac{(h_1+h_2)^2}{j}} \leqslant \frac{j}{2c} \sum_{i\geqslant k} \frac{k}{2} \mathrm{e}^{-c\frac{i^2}{j}}$$

The last sum is estimated writing $i = k + i'$

$$\sum_{j' \geqslant 0} \mathrm{e}^{-c\frac{(k+i')^2}{j}} \leqslant \mathrm{e}^{-c\frac{k^2}{j}} \sum_{i' \geqslant 0} \mathrm{e}^{-c\frac{(i')^2}{j}} \leqslant \left(\frac{j\pi}{2c}\right)^{1/2} \mathrm{e}^{-c\frac{k^2}{j}}$$

so that

$$A \leqslant \frac{\pi^{1/2}}{2^{5/2}c^{3/2}} k \mathrm{e}^{-c\frac{k^2}{j}} j^{3/2}.$$

Inserting (33) into (40) and using the bound on A, we finally deduce for any $c < 1/2$:

$$\mathbb{E}\left(\left|I^1\right|, |T| = j\right) \leqslant \frac{\pi^{1/2}}{2^{5/2}c^{3/2}} \left(\sum_{k \geqslant 3} k \mathrm{e}^{-c\frac{k^2}{j}} P^k\right) j^{3/2} \mathbb{P}(|T| = j)$$

and this is estimate (36). In the latter formula, the sum over k starts from $k = 3$ because the non-backtracking walk can not do shorter loops. The Proposition now follows by definition of the conditional expectation.

To bound the expected number of G^0-path-impure vertices, we only need the special case $k = 1$ in formula (19)

Proposition 3.9. *Let $T \in \mathcal{T}$ be the GW-tree given by (21). Let $I^0 = I^0(T)$ be the subset of the G^0-path-impure vertices of T constructed from the random homomorphism ι from T to G_n. Let $c < 1/2$. There exists $\ell_0 = \ell_0(c)$ such that, for $\ell \geqslant \ell_0$, and for any $j \geqslant 1$*

$$\mathbb{E}\left(\left|I^0(T)\right| \,\middle|\, |T| = j\right) \leqslant \frac{\pi^{1/2}}{2c^{3/2}} \frac{j^{3/2}\,|G^0|}{n} \tag{41}$$

Notice that, whenever $\iota(\rho) \in G^0$, we have $I^0(T) = T$. The fact that $\iota(\rho)$ is random is therefore important to avoid starting from G^0 too often.

For a tree $\mathrm{t}_1 \in \overline{\mathcal{T}_1}^{\mathrm{s}}$ spanned by 1 pointed vertex, we denote by h the generation of the pointed vertex, which uniquely defines the tree $\mathbf{t}_1 \in \overline{\mathcal{T}_1^{\mathrm{s}}}$. We abuse notation by writing $T_1(h)$ for $T_1(\mathbf{t}_1)$ in this case.

Proof. We do the choice

$$F(T, v) = \mathbf{1}\left\{v \; G^0\text{-path-impure in } T, |T| = j\right\}$$

and we use the many-to-one formula: this formula involves the tree $T_1(h)$ with one single pointed vertex at generation h. We find, after (38), that the product $\prod_{1 \leqslant i \leqslant k_0} m_i^{\ell_i}$ in (19) simplifies to:

$$m_1^{\ell_1} = (1 \pm \varepsilon)^{h_1}$$

Recall $u_1^{(i)}$ is the ancestor of u_1 at generation i. We apply (19) to the GW-tree T to find:

$$\begin{aligned}
\mathbb{E}\left(\left|I^0(T)\right|, |T| = j\right) &\leqslant \mathbb{E}\left(\sum_{v\in T\setminus\{\rho\}} F(T,v)\right) \\
&= \sum_h (1\pm\varepsilon)^h \mathbb{P}\left(u_1G^0 \text{ -path-impure in } T_1(h), |T_1(h)| = j\right) \\
&\leqslant \sum_{i\leqslant h} (1\pm\varepsilon)^h \mathbb{P}\left(\iota(u_1^{(i)}) \in G^0, |T_1(h)| = j\right) \\
&= \sum_{i\leqslant h} (1\pm\varepsilon)^h \mathbb{P}\left(\iota(u_1^{(i)}) \in G^0\right) \mathbb{P}\left(|T_1(h)| = j\right) \\
&\leqslant \frac{|G^0|}{n} \sum_h (1\pm\varepsilon)^h h \mathbb{P}\left(|T_1(h)| = j\right) \qquad (42)
\end{aligned}$$

For the last estimate, we used that $\mathbb{P}\left(\iota\left(u_1^{(i)}\right) \in G^0\right) = |G^0|/n$ for any $1 \leqslant i \leqslant h$, which holds because $\iota(\rho)$ is a random vertex in G. We now use Lemma 3.7 :

$$\begin{aligned}
\sum_h (1\pm\varepsilon)^h h \mathbb{P}\left(|T_1(h)| = j\right) &\leqslant \sum_h h^2 \mathrm{e}^{-ch^2/j} \mathbb{P}(|T| = j) \\
&\leqslant \frac{j}{2c} \sum_{h\geqslant 1} \mathrm{e}^{-ch^2/j} \mathbb{P}(|T| = j) \\
&\leqslant \frac{\pi^{1/2}}{2c^{3/2}} j^{3/2} \mathbb{P}(|T| = j) \qquad (43)
\end{aligned}$$

The estimate (41) follows from (42) and (43). □

We now bound $I^1(T)$ and $I^0(T)$ using the basic

Lemma 3.10. *Let $X_n \geqslant 0$ be a sequence of non-negative random variables with a finite first moment, and d_n be a sequence such that $\mathbb{E}(X_n) = o(d_n)$. There exists a sequence b_n satisfying*

$$b_n = o(d_n) \text{ and } \mathbb{P}(X_n \geqslant b_n) = o(1)$$

Proof. Set $a_n = \mathbb{E}(X_n)$ and choose $b_n = \sqrt{a_n d_n} = o(d_n)$. By Markov inequality, $\mathbb{P}(X_n \geqslant b_n) \leqslant \mathbb{E}(X_n)/b_n = (a_n/d_n)^{1/2} = o(1)$. □

Lemma 3.10 entails the following Corollary to Proposition 3.8.

Corollary 3.11. *Let t be given by (12) and β' be a non-negative sequence such that $t+\beta' \sim t$. In the setting of Proposition 3.8, and under condition (7), there exists a sequence β_1 such that*

$$\beta_1 = o\left(\delta^{-1}\right) \text{ and } \sup_{t\leqslant t'\leqslant t+\beta'} \mathbb{P}\left(\left|I^1\right| \geqslant \beta_1 \,\middle|\, |T| = t'\right) = o(1). \qquad (44)$$

Proof. Using condition (7) and the estimate (36) we find the bound:

$$\sup_{t\leqslant t'\leqslant t+\beta'} \mathbb{E}\left(\left|I^1\right| \big| |T| = t'\right) \leqslant \frac{\pi^{1/2}}{2^{5/2}c^{3/2}} \left(\sum_{k\geqslant 3} k\mathrm{e}^{-c\frac{k^2}{t'}} P^k\right) (t+\beta')^{3/2}$$

The assumption $t+\beta' \sim t$ and the condition (7) ensure the RHS is $o\left(\delta^{-1}\right)$. The existence of a sequence β_1 satisfying (44) now follows from Lemma 3.10. □

Corollary 3.12. *Let t be given by (12), and β' be a non-negative sequence such that $t+\beta' \sim t$. In the setting of Proposition 3.9 if ε satisfies (4), there exists a sequence β_0 such that*

$$\beta_0 = o\left(\delta^{-1}\right) \text{ and } \sup_{t',G^0} \mathbb{P}\left(\left|I^0\right| \geqslant \beta_0 \big| |T| = t'\right) = o(1) \tag{45}$$

with the sup over (t', G^0) such that $t \leqslant t' \leqslant t+\beta'$ and $G^0 \leqslant ct$ for c a finite constant.

Proof. We set $j = t$ in (41) and observe, using (12) that

$$\frac{t^{5/2}}{n} = \frac{\delta^{-5/2}s^{5/2}}{n} = o\left(\delta^{-1}\right).$$

Together with (41), we obtain: $\sup_{t',G^0} \mathbb{E}\left(|I^0(T)| \,||T| = t'\right) = o\left(\delta^{-1}\right)$ with the sup as indicated above. Lemma 3.10 now applies to give (45). □

The following Proposition is key to the proof of Theorem 2.1. Notice Proposition 3.4 makes a similar statement without taking into account the path-impure vertices.

Proposition 3.13. *Let $T = T_\pm \in \mathcal{T}$ be the GW-tree in (22). Assume ε satisfies (4), ℓ satisfies (3), and $t = t(u)$ is given by (12). There exists a sequence β such that $\beta = o\left(\delta^{-1}\right)$ and*

$$\frac{n}{t}\mathbb{P}(t+\beta \leqslant |T| < \infty, |I(T)| \leqslant \beta) = (1+o(1))\frac{1}{4\sqrt{\pi}}\mathrm{e}^{-u} \tag{46}$$

with $o(1)$ uniform over the sequences $|u| \leqslant v$ such that $v(n) = o\left(\log\log\left(\varepsilon^3 n\right)\right)$.

Crucial in this estimate is the choice of β : it should be larger than the typical values of $I(T)$, but small enough so the replacement of t by $t+\beta$ on the LHS in (27) and (46) does not change the limit in the RHS.

Proof. We set $\beta = \beta_0 + \beta_1$ given by the two Corollaries, and we notice that $t+\beta = \delta^{-1}(s+o(1))$. Also, we set $u' = \left(\log\log\left(\varepsilon^3 n\right)\right)^{1/2}$ and[7] $\beta' = \delta^{-1}u'$. With this definition, we have that $t+\beta' \sim t+\beta \sim t$. Proposition 3.4 applies and we find that the three quantities:

$$\mathbb{P}\left(t+\beta \leqslant |T| \leqslant t+\beta'\right), \mathbb{P}(t+\beta \leqslant |T| < \infty), \text{ and } \mathbb{P}(t \leqslant |T| < \infty) \tag{47}$$

are equivalent as $n \to \infty$. We have the lower and upper bounds:

$$\begin{array}{ll} \left(1 - \sup_{t+\beta\leqslant t'\leqslant t+\beta'} \mathbb{P}\left(|I| > \beta \big| |T| = t'\right)\right) \mathbb{P}\left(t+\beta \leqslant |T| \leqslant t+\beta'\right) & \leqslant \mathbb{P}(t+\beta \leqslant |T| < \infty, |I| \leqslant \beta) \\ & \leqslant \mathbb{P}(t+\beta \leqslant |T| < \infty) \end{array}$$

By (44), (45), the definition of β and (47), the two bounds are equivalent as $n \to \infty$, moreover $(n/t)\mathbb{P}(t+\beta \leqslant |T| < \infty, |I| \leqslant \beta) = (1+o(1))\mathrm{e}^{-u}/(4\sqrt{\pi})$. □

[7] any sequence u' satisfying $1 \ll u' \ll \log\log\left(\varepsilon^3 n\right)$ works as well

3.4 Modified GW-trees.

In the modified GW-tree $T_{\pm}^{\mathrm{m}} \in \mathcal{T}$, every vertex v has a random independent number of offspring distributed as $\mathrm{Bin}\left(\ell - \mathbf{1}_{v\neq\rho}, p\right)$. It is called "modified" because the distinct offspring distribution is different at the root. In the case (3) that ℓ diverges, this change affects the asymptotic of the tree only slightly, and the following dominations hold between the tails of $|T^m|$ and $|T|$.

Lemma 3.14. *Let $j \in \mathbb{N}$. The tail of the random variables $|T|$ and $|T^m|$ satisfy:*

$$\mathbb{P}(|T| \geqslant j) \leqslant \mathbb{P}\left(|T^m| \geqslant j\right) \leqslant (1 + O(1/\ell))\mathbb{P}(|T| \geqslant j) \tag{48}$$

Proof. We call natural coupling of T and T^{m} the coupling that uses the same Bernoulli random variables to define the Binomial number of offspring at each vertex. In this coupling, only the number of children of the root may differ, by at most 1. We have $T \subseteq T^{\mathrm{m}}$, whence the lower bound in (50). If we consider T_1 and T_2 two independent copies of T and B an independent random variable such that $\mathbb{P}(B = 1) = 1 - \mathbb{P}(B = 0) = 1/(\ell - 1)$ then we have

$$|T^{\mathrm{m}}| = |T_1| + B\,|T_2|\,. \tag{49}$$

in this coupling. We also point out that $\mathbb{P}\left(|T_1| + |T_2| \geqslant j\right) \leqslant 2(1 + O(\varepsilon))\mathbb{P}(|T| \geqslant j)$ follows from (23) with $h = 2$. These two equations entail the upper bound in (48).

$$\begin{aligned}\mathbb{P}\left(|T^{\mathrm{m}}| \geqslant j\right) &\leqslant \mathbb{P}\left(|T_1| \geqslant j\right) + \mathbb{P}(B = 1)\mathbb{P}\left(|T_1| + |T_2| \geqslant j\right)\\ &\leqslant (1 + O(1/\ell))\mathbb{P}(|T| \geqslant j)\end{aligned}$$

□

Recall the definition of $\mathcal{C}^0(v)$ a few lines before (35).

Lemma 3.15. *There exists a coupling in which, if v is a uniformly chosen random vertex independent of the percolation of G :*

$$|T^m \backslash I\left(T^m\right)| \leqslant \left|\mathcal{C}^0(v)\right| \leqslant |T^m| \tag{50}$$

This is essentially the statement of Proposition 11 in [Nac09], with the only difference that we take v a uniformly chosen random vertex in G (consider the case of a fixed, deterministic $v \in G^0$ to see why this is needed). We do not repeat the proof. The lower bound in (50) may be strict: this is because a vertex $w \in T$ can be path-impure because of a vertex $v \in T$ that is itself path-impure.

Pruning off the path-impure vertices does not necessarily preserve the inclusion of trees: $T \subseteq T^{\mathrm{m}}$ does not imply in general $T\backslash I(T) \subseteq T^{\mathrm{m}}\backslash I\left(T^{\mathrm{m}}\right)$. However, in the coupling (49), T and T^{m} agree in distribution on the event $B = 0$, and we always have the lower bound:

$$\begin{aligned}\mathbb{P}\left(|T^{\mathrm{m}}\backslash I\left(T^{\mathrm{m}}\right)| \geqslant j\right) &\geqslant \mathbb{P}(|T\backslash I(T)| \geqslant j, B = 0)\\ &\geqslant (1 - O(1/\ell))\mathbb{P}(|T| \geqslant j + \beta, |I(T)| \leqslant \beta)\end{aligned}$$

3.5 The number of components of $G_n(p)$ with size in a given interval.

The results collected so far are applied in this section to the random graph of interest. For $G^0 \subseteq G = G_n$ recall $G\backslash G^0$ is the graph induced by G on the vertex set $V(G)\backslash V(G^0)$. Call $\left(\mathcal{C}_j^0, j \geqslant 1\right)$ the largest components of $(G\backslash G^0)(p)$ arranged in non-increasing order of size. Recall the definition of the map $\mathbb{R} \to \mathbb{R}, u \mapsto \mathrm{t}(u)$ in (12), and set t^{-1} for the inverse function of t. A point measure recording the sizes of the components in $(G\backslash G^0)(p)$ is defined by:

$$N^0 = \sum_{j \geqslant 1} \delta_{\mathrm{t}^{-1}(|\mathcal{C}_j^0|)}$$

and we set N for N^0 when G^0 is the empty graph. The goal of this section is to show the convergence, in a sense to be precised, of the random measure N towards a Poisson point measure whose intensity measure μ is given by (8). A first step is to compute the first moment of the random measure N^0 evaluated on intervals. Call an interval $J = (u_1, u_2)$ bounded when $-\infty < u_1 \leqslant u_2 < \infty$, and bounded from the left when $-\infty < u_1 \leqslant u_2 \leqslant \infty$. In the following, we set $t_1 = \mathrm{t}(u_1)$ and $t_2 = \mathrm{t}(u_2)$. We stress that the next two propositions are concerned with the weak subcritical regime.

Proposition 3.16. *Let J be a bounded interval. Assume ε satisfies (4), ℓ satisfies (3), t is given by (13) and P^k satisfies condition (7). The first moment of the random variable $N_-^0(J)$ satisfies:*

$$\mathbb{E}\left(N_-^0(J)\right) \to \int_J \mu(dx) \text{ as } n \to \infty$$

and the convergence is uniform over the graphs $G^0 \subseteq G$ such that $|G^0| \leqslant Ct$ for any constant C independent of n.

Proof. Let $j, j' \in \mathbb{N}$ with $j \leqslant j'$. The inequality in Lemma 3.15 :

$$\mathbb{P}\left(|T^{\mathrm{m}}\backslash I(T^{\mathrm{m}})| \geqslant j\right) \leqslant \mathbb{P}\left(\left|\mathcal{C}^0(v)\right| \geqslant j\right) \leqslant \mathbb{P}\left(|T^{\mathrm{m}}| \geqslant j\right)$$

can also be written[8] in term of T:

$$(1 - O(1/\ell))\mathbb{P}(|T\backslash I(T)| \geqslant j) \leqslant \mathbb{P}\left(\left|\mathcal{C}^0(v)\right| \geqslant j\right) \leqslant (1 + O(1/\ell))\mathbb{P}(|T| \geqslant j)$$

using the natural coupling of T and T^{m} for the lower bound, see the proof of Lemma 3.14 , and again Lemma 3.14 for the upper bound. We can now put the pieces together. We multiply the last inequality by n/j and then observe that $\mathbb{P}(|T\backslash I(T)| \geqslant j) \geqslant \mathbb{P}(j' \leqslant |T| < \infty, |I(T)| \leqslant j' - j)$. Also we choose $j = t$ and $j' = t + \beta$. The estimates in (27),(46) entail that

$$\lim_{n\to\infty} \frac{n}{t}\mathbb{P}\left(\left|\mathcal{C}_-^0(v)\right| \geqslant t\right) = \frac{1}{4\sqrt{\pi}}\mathrm{e}^{-u} \tag{51}$$

The first moment of $N_-^0(J)$ then satisfies:

[8] this step could have been avoided by stating a many-to-two formula for modified GW-trees

$$\mathbb{E}\left(N_-^0(J)\right) = \mathbb{E}\left(\sum_v \frac{1_{\{|\mathcal{C}^0(v)|\epsilon(t_1,t_2)\}}}{|\mathcal{C}^0(v)|}\right) = (1+o(1))\frac{n}{t_1}\mathbb{P}\left(\left|\mathcal{C}^0(v)\right| \in (t_1,t_2)\right)$$
$$= (1+o(1))\frac{1}{4\sqrt{\pi}}\left(\mathrm{e}^{-u_1} - \mathrm{e}^{-u_2}\right) \tag{52}$$

when $n \to \infty$. Notice we used $t_1 \sim t_2$ for the second equality. The RHS of (52) is the integral of μ given by (8) on J, which concludes the proof. □

Recall that N stands for N^0 when G^0 is the empty graph.

Proposition 3.17. *Let J be a bounded interval. Assume ε satisfies (4), ℓ satisfies (3), t is given by (12) and P^k satisfies condition (7). The second factorial moment of $N_-(J)$ satisfies:*

$$\mathbb{E}\left(N_-(J)\left(N_-(J)-1\right)\right) \to \left(\int_J \mu(dx)\right)^2 \tag{53}$$

Proof. We start by writing:

$$\mathbb{E}(N(J)(N(J)-1)) = \mathbb{E}\left(\sum_w \frac{\mathbf{1}_{\{|\mathcal{C}(w)|\in \mathrm{t}(J)\}}}{|\mathcal{C}(w)|}(N(J)-1)\right)$$
$$= n\sum_{j\in \mathrm{t}(J)} \frac{\mathbb{E}(N(J)-1, |\mathcal{C}(w)|=j)}{j}$$

where the sum at the first equality is over the vertices w of G_n, and w may be any vertex (by vertex transitivity) at the second equality. The latter expression may be written

$$(1+o(1))\frac{n}{t_1}\sum_{G^0\subseteq G_n, |G^0|\in(t_1,t_2)} \mathbb{E}\left(N(J)-1 \mid \mathcal{C}(w)=G^0\right)\mathbb{P}\left(\mathcal{C}(w)=G^0\right) \tag{54}$$

In term of N_-^0, this also writes:

$$\mathbb{E}\left(N_-(J)-1 \mid \mathcal{C}(w)=G^0\right) = \mathbb{E}\left(N_-^0(J)\right) = (1+o(1))\int_J \mu(dx) \tag{55}$$

using Proposition 3.16 for the latter identity, where $o(1)$ is uniform over the graphs $G^0 \subseteq G_n$ such that $|G^0| \in (t_1, t_2)$. Putting this into (54) and using (55) with G^0 reduced to the empty graph, we conclude that (53) holds. □

Let $\mathcal{J}$ denote the set of finite unions of intervals bounded from the left.

Lemma 3.18. *Propositions 3.16 and 3.17 are valid for $J \in \mathcal{J}$.*

We stress that Propositions 3.16 and 3.17 are stated in the weak subcritical regime. Proposition 3.16 does not hold for every $J \in \mathcal{J}$ in the weak supercritical regime.

Proof. We start with Proposition 3.16. First consider the case when J is an interval, $J = (u_1, u_2)$ with $u_2 = \infty$. The equivalent $t_1 \sim t_2$ does no more hold in the equality (52). An inequality replaces that equality:

$$\mathbb{E}\left(N_-^0(J)\right) \leqslant \frac{n}{t_1}\mathbb{P}\left(\left|\mathcal{C}^0(v)\right| \in J\right) = (1+o(1))\frac{1}{4\sqrt{\pi}}\mathrm{e}^{-u_1}. \tag{56}$$

The converse inequality is proved by approximating J by an increasing sequence of finite intervals, and we obtain $\liminf_{n\to\infty}\mathbb{E}\left(N_-^0(J)\right) \geqslant \mathrm{e}^{-u_1}/4\sqrt{\pi}$. Linearity of the expectation entails that (52) extends to an arbitrary $J \in \mathcal{J}$. We turn to Proposition 3.17. First consider $J = (u_1, u_2)$ with $u_2 = \infty$. As before, a lower bound is achieved by approximation through an increasing sequence of finite intervals: $\liminf_{n\to\infty}\mathbb{E}\left(N_-(J)\left(N_-(J)-1\right)\right) \geqslant \left(\mathrm{e}^{-u_1}/4\sqrt{\pi}\right)^2$. For the upper bound, we notice that $|\mathcal{C}^0(v)| \leqslant |\mathcal{C}(v)|$ and this entails, since $u_2 = \infty$, that:

$$\mathbb{E}\left(N^0(J)\right) \leqslant \frac{n}{t_1}\mathbb{P}\left(\left|\mathcal{C}^0(v)\right| \in J\right) \leqslant \frac{n}{t_1}\mathbb{P}(|\mathcal{C}(v)| \in J).$$

We compute as in (54)

$$\begin{aligned}
\mathbb{E}(N(J)(N(J)-1)) &\leqslant \frac{n}{t_1}\sum_{|G^0|\in\mathbf{t}(J)}\mathbb{E}\left(N^0(J)\right)\mathbb{P}\left(\mathcal{C}(w)=G^0\right)\\
&\leqslant \left(\frac{n}{t_1}\right)^2\mathbb{P}(|\mathcal{C}(v)|\in \mathrm{t}(J))\sum_{|G^0|\in\mathrm{t}(J)}\mathbb{P}\left(\mathcal{C}(w)=G^0\right)\\
&\leqslant \left(\frac{n}{t_1}\mathbb{P}(|\mathcal{C}(v)|\in \mathrm{t}(J))\right)^2
\end{aligned}$$

But we know from (56) with G^0 the empty graph that $(n/t_1)\,\mathbb{P}(|\mathcal{C}(v)| \in \mathrm{t}(J)) = (1+o(1))\mathrm{e}^{-u_1}/4\sqrt{\pi}$ in the weak subcritical regime, and the upper bound $\limsup_{n\to\infty}\mathbb{E}\left(N_-(J)\left(N_-(J)-1\right)\right) \leqslant \left(\mathrm{e}^{-u_1}/4\sqrt{\pi}\right)^2$ follows. The case of an arbitrary $J \in \mathcal{J}$ is similar. □

Let $\mathcal{M}(\mathbb{R})$ be the set of locally finite measures on the Borel sigma-algebra of $\mathbb{R}$. A measure $M \in \mathcal{M}(\mathbb{R})$ is called a point measure when $M(J)$ takes values in $\mathbb{N}$ for any J bounded Borel set; a point measure is further called simple when $M(\{x\}) \in \{0,1\}$ for any $x \in \mathbb{R}$. A random element of $\mathcal{M}(\mathbb{R})$ is called a random measure. A sequence $(M_n)_{n\in\mathbb{N}}$ of random measures weakly converges (resp. vaguely converges) towards a random measure M when the sequence of random variables $\int M_n(dx)f(x)$ weakly converge to $\int M(dx)f(x)$ for each f continuous and bounded (resp. continuous bounded and compactly supported). Proposition 16.17 in [Kal02], reproduced below, gives a criterion for the vague convergence of probability measures in term of the void probabilities.

Proposition 3.19. *Let $(M_n)_{n\geqslant 1}$ be a sequence of random point measures on $\mathbb{R}$, and let M be a random simple point measure. Then $(M_n)_{n\geqslant 1}$ vaguely converges to M if the following two conditions hold:*

- $\lim_{n\to\infty}\mathbb{P}\left(M_n(J)=0\right) = \mathbb{P}(M(J)=0)$, *for any J finite union of bounded intervals of $\mathbb{R}$.*

- $\limsup_{n\to\infty} \mathbb{E}\left(M_n(K)\right) \leqslant \mathbb{E}(M(K))$, *for any compact set* K.

Theorem 2.1 requires the weak convergence of the measures restricted to intervals bounded from the left. The latter may be turned into vague convergence by compactification of the space.

Proof of Theorem 2.1. To stress on the dependence in n, we write N_n for N in this proof. Let $k \geqslant 1$, and $J \in \mathcal{J}$. By an induction argument, we arrive at the following generalisation of Proposition 3.17 : the k-th factorial moment satisfies

$$\mathbb{E}\left(\prod_{0\leqslant i\leqslant k-1} (N_n(J) - i)\right) = (1 + o(1)) \left(\int_J \mu(dx)\right)^k \tag{57}$$

For $k \in \mathbb{N}$, the k-th moment of a random variable is a linear combination of the i-th factorial moments for $1 \leqslant i \leqslant k$, hence the convergence (57) of the factorial moments entails that of the usual moments. Moreover, the Poisson distribution is uniquely determined by its moments. By the method of moments, see *e.g.* Section 6.1 of [JLR11] and the Theorem 6.1 in particular, $N_n(J)$ weakly converges towards the Poisson distribution with parameter $\int_J \mu(dx)$. This implies in particular the convergence of the void probabilities:

$$\lim_{n\to\infty} \mathbb{P}\left(N_n(J) = 0\right) = \mathrm{e}^{-\int_J \mu(dx)} \tag{58}$$

The proof is concluded applying Proposition 3.19 to N_n^ϕ the push-forward of N_n by an increasing diffeomorphism $\phi : \mathbb{R} \to (-\infty, 0)$, e.g. $x \mapsto -\mathrm{e}^{-x}$. Call μ^ϕ the push-forward measure μ by the map ϕ. Let J be a finite union of bounded intervals. Equation (58) is equivalent to:

$$\lim_{n\to\infty} \mathbb{P}\left(N_n^\phi(J) = 0\right) = \mathrm{e}^{-\int_J \mu^\phi(dx)}$$

Let $\varepsilon > 0$, and K be a compact set of the real line. For $x \in \mathbb{R}$, set $d(x, K) = \inf\{d(x, y), y \in K\}$ for the distance of x to K. For $\eta > 0$, the set $O_\eta = \{x; d(x, K) < \eta\}$ is an open set of the real line, hence it can be written as a union of open intervals, that is furthermore finite. By monotone convergence there exists $\eta > 0$ so that $\mu^\phi\left(O_\eta \backslash K\right) < \varepsilon$. Proposition 3.16 applies:

$$\mathbb{E}\left(N_n^\phi(K)\right) \leqslant \mathbb{E}\left(N_n^\phi\left(O_\eta\right)\right) \to \int_{O_\eta} \mu^\phi(dx) \leqslant \int_K \mu^\phi(dx) + \varepsilon$$

and this proves, since ε is arbitrarily small, that

$$\limsup_{n\to\infty} \mathbb{E}\left(N_n^\phi(K)\right) \leqslant \int_K \mu^\phi(dx), \tag{59}$$

for K a compact set. (58) and (59) are the assumptions to apply Proposition 3.19, which entails the vague convergence of N_n^ϕ towards the Poisson measure with intensity μ^ϕ. The latter in turn is equivalent to the weak convergence of N_n to the Poisson measure with intensity μ, when both measures are restricted to intervals bounded from the left. This is the statement of Theorem 1.1. □

3.6 Adaptation to the case of graphs with bounded degrees

Interesting examples of expander graphs with bounded degrees are known (the Ramanajuan graphs constructed in [LPS88] for instance), and we would like them to be included in our analysis. This requires modifications in the both the statement and the proof of Theorem 2.1. Precisely, assume

$$\ell(n) \geqslant 3, \quad \lim_{n\to\infty} \ell(n) = L < \infty$$

which means that $\ell(n)$ is constant equal to L for n large enough. Theorem (2.1) then holds with the intensity of the Poisson point measure in that Theorem replaced by its multiple

$$\frac{1}{4\sqrt{\pi}} \frac{L}{L-1} \mathrm{e}^{-x} dx$$

Observe that, when $L = \infty$, the convention $L/(L-1) = 1$ allows to recover the original intensity measure. If $\ell = \ell(n)$ and $\varepsilon = \varepsilon(n)$ satisfy (4) and n goes to infinity, one has

$$\delta_{\pm}(\varepsilon, \ell - 1) \sim \frac{(L-1)(L-3)}{(L-2)^2} \frac{\varepsilon^2}{2} \text{ as } n \to \infty :$$

and the first order of the size of the largest components is now $2\varepsilon^{-2} \log(\varepsilon^3 n) (L-2)^2/((L-1)(L-3)$. To prove Theorem 2.1 in this new setting, we first observe using the decomposition (51) and the exact computation (31) for the size of the union of $h = 2\mathrm{GW}$ trees that the estimate (50) on the tail of the modified GW tree has to be replaced by

$$\frac{n}{t}\mathbb{P}\left(\left|T_-^{\mathrm{m}}\right| \geqslant t\right) = (1 + o(1)) \frac{1}{4\sqrt{\pi}} \frac{L}{L-1} \mathrm{e}^{-u} \text{ as } n \to \infty \tag{60}$$

in other words, T_- and T_-^{m} have no more equivalent tails (in the scale $t = \mathrm{t}(u)$). We then need an estimate similar to (60) with $T_-^{\mathrm{m}} \backslash I\left(T_-^{\mathrm{m}}\right)$ in place of T_-^{m} in order to prove the analogue of (51). There is no way round but to find a many-to-k formula in the context of modified GW-trees. This modification is achieved as follows: in Lemma 3.1, if $\tilde{m}_i$ denotes the i-th factorial moment of the offspring distribution at the root, the RHS in (22) is multiplied by $\tilde{m}_{c_{\mathrm{t_k}}(\rho)}/m_{c_{\mathrm{t_k}}(\rho)}$ (remember that the root only has a distinct offspring distribution in modified GW-trees). In our case, the ratio $\tilde{m}_i/m_i$ is $\ell/(\ell - i)$, and the upper bounds in the estimates (36) and (41) are multiplied by a positive constant independent of n. Since condition (7) is not sensitive to constants, we conclude that, under (7)

$$\frac{n}{t}\mathbb{P}\left(\left|T_-^{\mathrm{m}} \backslash I\left(T_-^{\mathrm{m}}\right)\right| \geqslant t\right) = (1 + o(1)) \frac{1}{4\sqrt{\pi}} \frac{L}{L-1} \mathrm{e}^{-u} \text{ as } n \to \infty \tag{61}$$

and the rest of the proof follows unchanged from this point on.

3.7 A remark on the second largest component in the weak supercritical regime.

Recall the definition of the quantity $\bar{\varepsilon}$ in Remark 3.5. There is a conjectured parallel, known as the discrete duality principle, between the largest components in $G_n(\bar{p}_-), \bar{p}_- = (1-\bar{\varepsilon})/(\ell-1)$

and the largest components from the second one in $G_n(p_+)$ $p_+ = (1+\varepsilon)/(\ell-1)$. This principle has been proved for a few graphs, among which the complete graph [NP07] and the configuration model [Rio12]. The proof usually relies on the possibility to characterise the (random) graph induced by $G_n(p_+)$ on the complement of $\mathcal{C}_1$ in a simple way. We believe that an analogous result should hold in our case, yet the assumption we make or the methods we use in this paper only give a lower bound for the expected number of components in certain intervals, as shown below. The random measure of interest is again

$$N = \sum_{j \geqslant 1} \delta_{\mathrm{t}^{-1}(|\mathcal{C}_j|)}$$

and the lower bound is on the expected number of components with size in the interval $\mathrm{t}(u_1, u_2) = (t_1, t_2)$, as follows.

Proposition 3.20. *Assume ε satisfies (4), ℓ satisfies (3), and let $p_+ = (1+\varepsilon)/(\ell-1)$. Assume the non-backtracking random walk on G_n satisfies (7). It holds, for $J = (u_1, u_2)$, $-\infty < u_1 \leqslant u_2 < \infty$, that:*

$$\mathbb{E}(N(J)) \geqslant (1+o(1)) \frac{1}{4\sqrt{\pi}} \left(\mathrm{e}^{-u_1} - \mathrm{e}^{-u_2} \right)$$

as $n \to \infty$.

Proof. The set $I^1(T)$ is the subset of path-impure vertices in T defined from the random homomorphism ι from T to G_n. For $\beta = o(\delta^{-1})$ as in Corollary 3.11, we have:

$$\begin{aligned}
\mathbb{E}(N(J)) &= (1+o(1)) \frac{n}{t_1} \mathbb{P}\left(|\mathcal{C}(v)| \in (t_1, t_2) \right) \\
&\geqslant (1+o(1)) \frac{n}{t_1} \mathbb{P}\left(|T| \in (t_1+\beta, t_2), \left| I^1 \right| \leqslant \beta \right) \\
&= (1+o(1)) \frac{1}{4\sqrt{\pi}} \left(\mathrm{e}^{-u_1} - \mathrm{e}^{-u_2} \right)
\end{aligned}$$

using $t_1 \sim t_2$ at the first and second line and Proposition 3.4 (with condition (7)) at the second line. □

With respect to the computation (52), we obtain an inequality in place of an equality at the second line. Let $\eta > 0$ and $j \in \mathbb{N}$ be fixed, independent of n. The lower bound above suggests that, with high probability as $n \to \infty$, the j-th largest component has size at least $\delta_+(\varepsilon, \ell-1)^{-1} \left(\log(\varepsilon^3 n) - (5/2+\eta) \log\log(\varepsilon^3 n) \right)$

$$\mathbb{P}\left(\frac{\delta_+(\varepsilon, \ell-1) |\mathcal{C}_j| - \log(\varepsilon^3 n)}{\log\log(\varepsilon^3 n)} > -(5/2+\eta) \right) = 1 - o(1) \tag{62}$$

To prove this statement, it would be enough to have an upper bound on the second (factorial) moment. For $v, w \in V(G_n)$, we write $v \not\sim w$ if v and w are not connected by a path of open edges. Now, $N(J)(N(J)-1)$ counts the number of ordered pair of distinct components, and

It follows that if V and W stand for two independent uniform vertices in $V(G_n)$ under $\mathbb{P}$,

$$\mathbb{E}(N(J)(N(J)-1)) = (1+o(1))\left(\frac{n}{t_1}\right)^2 \mathbb{P}(|\mathcal{C}(V)| \in \mathrm{t}(J), |\mathcal{C}(W)| \in \mathrm{t}(J), V \not\sim W)$$

We did not find [9] an obvious way to bound $\mathbb{P}(|\mathcal{C}(V)| \in \mathrm{t}(J), |\mathcal{C}(W)| \in \mathrm{t}(J), V \not\sim W)$ by $\mathbb{P}(|\mathcal{C}(V)| \in \mathrm{t}(J))^2$, which is the first step to implement the second moment method and conclude to (67). It may be that further conditions are necessary to prove (62).

3.8 Verification of condition (7)

We rely on [Nac09] to check condition (7). To that aim, it is useful first to relate (7) with the assumption

$$n^{1/3}\sum_{k=1}^{n^{1/3}} kP^k = O(1) \tag{63}$$

made by Nachmias in the study of the critical regime. Assumption (63) alone is not enough to check condition (7). One also needs the following condition [10]: there exists a finite constant c independent of n such that for n large enough,

$$P^k \leqslant \frac{c}{n} \text{ for } k \geqslant n^{1/3}. \tag{64}$$

Lemma 3.21. *Assume the sequence ε satisfies (4). Conditions (63) and (64) imply (7).*

Proof. The first $n^{1/3}$ terms in the sum in (7) are bounded using Nachmias condition (63)

$$t^{1/2}\sum_{k=1}^{n^{1/3}} k\mathrm{e}^{-c\frac{k^2}{t}}P^k \leqslant t^{1/2}n^{-1/3}\left(n^{1/3}\sum_{k=1}^{n^{1/3}} kP^k\right) = O\left(\frac{s^{1/2}}{\left(\delta^{3/2}n\right)^{1/3}}\right)$$

and it is simple to check that $s^{1/2}/\left(\delta^{3/2}n\right)^{1/3} = o(1/s)$. For the subsequent terms in the sum, we have from assumption (64) that:

$$t^{1/2}\sum_{k=n^{1/3}}^{\infty} k\mathrm{e}^{-c\frac{k^2}{t}}P^k = O\left(\frac{t^{3/2}}{n}\right)$$

with room to spare. Then

$$\frac{t^{3/2}}{n} = \frac{\left(\delta^{-1}s\right)^{3/2}}{n} = \frac{s^{3/2}}{\delta^{3/2}n} = o\left(\frac{1}{s}\right) \tag{65}$$

[9] The possibility of closed edges with endvertices in both $|\mathcal{C}(V)|$ and $|\mathcal{C}(W)|$ prevents us from using the van den Berg-Kesten-Reimer (BKR) inequality

[10] In practice, checking (64) usually does not raise additional difficulties with respect to (63). In [Nac09] for instance, (63) is checked by proving in the first place that the bound on P^k in (64) holds for $k \geqslant \log(n)$ which entails (64).

using for the last estimate that $\delta^{3/2}n = \left(2^{-3/2} + o(1)\right)\varepsilon^3 n$, as follows from (10), and then $\log^{5/2}(\varepsilon^3 n) = o(\varepsilon^3 n)$, as follows from (4). Both terms in the sum in the LHS of (7) are $o(1/s)$, hence the condition is satisfied. □

Estimates on the kernel of the non-backtracking walk computed in [Nac09] then yield Proposition 2.6.

Proof of Proposition 2.6. For the two graphs, condition (63) is checked in Theorem 2 and, under condition 11, in Theorem 6 of [Nac09]. Also, condition (64) is checked along the proofs of these two theorems: see p.1177 for the expander graphs and p.1178 for the Hamming graph in that same reference. Lemma 3.21 concludes the proof. □

In the case of the Hamming graph in dimension $1, 2$ and 3, one may check (7) by hand, without the intermediate step of checking condition (63). Linking conditions (7) and (63) allows us not to display tedious but straightforward computations.

Acknowledgments. We are grateful to M.J. Luczak for asking us a question about the percolation of Hamming graphs that started this work and for sharing her expertise in the field.

References

[ABDJ13] Louigi Addario-Berry, Luc Devroye, and Svante Janson. Sub-Gaussian tail bounds for the width and height of conditioned Galton–Watson trees. *The Annals of Probability*, 41(2):1072–1087, 2013.

[Ald97] David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25(2):812–854, 1997.

[BCvdH+05a] Christian Borgs, Jennifer T. Chayes, Remco van der Hofstad, Gordon Slade, and Joel Spencer. Random subgraphs of finite graphs. I. The scaling window under the triangle condition. *Random Structures Algorithms*, 27(2):137–184, 2005.

[BCvdH+05b] Christian Borgs, Jennifer T. Chayes, Remco van der Hofstad, Gordon Slade, and Joel Spencer. Random subgraphs of finite graphs. II. The lace expansion and the triangle condition. *Ann. Probab.*, 33(5):1886–1944, 2005.

[BCvdH+06] Christian Borgs, Jennifer T. Chayes, Remco van der Hofstad, Gordon Slade, and Joel Spencer. Random subgraphs of finite graphs. III. The phase transition for the n-cube. *Combinatorica*, 26(4):395–410, 2006.

[Bol84] Béla Bollobás. The evolution of random graphs. *Trans. Amer. Math. Soc.*, 286(1):257–274, 1984.

[BR09] Béla Bollobás and Oliver Riordan. Random graphs and branching processes. In *Handbook of large-scale random networks*, volume 18 of *Bolyai Soc. Math. Stud.*, pages 15–115. Springer, Berlin, 2009.

[ER60] Paul Erdös and Alfréd Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[FVDHDHH20] Lorenzo Federico, Remco Van Der Hofstad, Frank Den Hollander, and Tim Hulshof. Expansion of percolation critical points for hamming graphs. *Combinatorics, Probability and Computing*, 29(1):68–100, 2020.

[HN20] Tim Hulshof and Asaf Nachmias. Slightly subcritical hypercube percolation. *Random Structures & Algorithms*, 56(2):557–593, 2020.

[HR15] Simon Harris and Matthew Roberts. The many-to-few lemma and multiple spines. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 2015.

[INW69] Nobuyuki Ikeda, Masao Nagasawa, and Shinzo Watanabe. Branching Markov processes iii. *Journal of Mathematics of Kyoto University*, 9(1):95–160, 1969.

[Jan11] Svante Janson. Poset limits and exchangeable random posets. *Combinatorica*, 31(5):529–563, 2011.

[JLR11] Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *Random graphs*, volume 45. John Wiley & Sons, 2011.

[Kal02] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[LP17] Russell Lyons and Yuval Peres. *Probability on trees and networks*, volume 42. Cambridge University Press, 2017.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[Łuc90] Tomasz Łuczak. Component behavior near the critical point of the random graph process. *Random Structures & Algorithms*, 1(3):287–310, 1990.

[Nac09] Asaf Nachmias. Mean-field conditions for percolation on finite graphs. *Geom. Funct. Anal.*, 19(4):1171–1194, 2009.

[NP07] Asaf Nachmias and Yuval Peres. Component sizes of the random graph outside the scaling window. *ALEA Lat. Am. J. Probab. Math. Stat.*, 3:133–142, 2007.

[Pit02] J. Pitman. Combinatorial stochastic processes. Lectures from the 32nd Summer School on Probability Theory held in saint-flour, july 7–24, 2002. *Lecture Notes in Mathematics*, 2002.

[Rio12] Oliver Riordan. The phase transition in the configuration model. *Combinatorics, Probability and Computing*, 21(1-2):265–299, 2012.

[RW17] O. Riordan and L. Warnke. The phase transition in bounded-size Achlioptas processes. *ArXiv e-prints*, April 2017.

[Spi56] Frank Spitzer. A combinatorial lemma and its application to probability theory. *Transactions of the American Mathematical Society*, 82(2):323–339, 1956.

[vdHL10] Remco van der Hofstad and Malwina J. Luczak. Random subgraphs of the 2D Hamming graph: the supercritical phase. *Probab. Theory Related Fields*, 147(1-2):1–41, 2010.

[vdHN15] Remco van der Hofstad and Asaf Nachmias. Hypercube percolation. *Journal of the European Mathematical Society*, 2015.