RANDOM SUBGRAPHS OF THE 2D HAMMING GRAPH: THE SUPERCRITICAL PHASE

REMCO VAN DER HOFSTAD AND MALWINA J. LUCZAK

CDAM Research report LSE-CDAM-2006-15

ABSTRACT. We study random subgraphs of the 2-dimensional Hamming graph $H(2,n)$, which is the Cartesian product of two complete graphs on n vertices. Let p be the edge probability, and write $p = \frac{1+\varepsilon}{2(n-1)}$ for some $\varepsilon \in \mathbb{R}$. In [4, 5], the size of the largest connected component was estimated precisely for a large class of graphs including $H(2,n)$ for $\varepsilon \leq \Lambda V^{-1/3}$, where $\Lambda > 0$ is a constant and $V = n^2$ denotes the number of vertices in $H(2,n)$. Until now, no matching lower bound on the size in the supercritical regime has been obtained.

In this paper we prove that, when $\varepsilon \gg (\log V)^{1/3} V^{-1/3}$, then the largest connected component has size close to $2\varepsilon V$ with high probability. We thus obtain a law of large numbers for the largest connected component size, and show that the corresponding values of p are supercritical. Barring the factor $(\log n)^{1/3}$, this identifies the size of the largest connected component all the way down to the critical p window.

1. INTRODUCTION

We study random subgraphs of the 2-dimensional Hamming graph $H(2,n)$. The d-dimensional Hamming graph is a graph on $V = n^d$ vertices, each corresponding to one of the n^d distinct d-vectors $\mathbf{v} = (v_1, \ldots, v_d) \in \{1, \ldots, n\}^d$. A pair of vertices are connected by an edge if and only if these vertices differ in precisely one co-ordinate. (See for example [9] for more information on the properties of Hamming graphs.) The 1-dimensional Hamming graph $H(1,n)$ is the complete graph; for $d \geq 2$, the graph $H(d,n)$ is the Cartesian product of d complete graphs on n vertices. In particular, it is transitive and the degree of each vertex is $\Omega = d(n-1)$.

We write $\mathbb{P}_p$ for the probability law of the random subgraph of $\mathbb{G}$ resulting when each edge is *occupied* (or *present*) with probability p, and *vacant* (or *absent*) with probability $1-p$, independently of all the other edges. We write $\mathbb{E}_p$ for the expectation with respect to $\mathbb{P}_p$. Also, Var_p will denote the variance under $\mathbb{P}_p$.

Throughout we work with the 2-dimensional Hamming graph $H(2,n)$ unless explicitly stated otherwise, and we shall assume that $p = \frac{1+\varepsilon}{2(n-1)} = \frac{1+\varepsilon}{\Omega}$, where $\varepsilon = \varepsilon(n) \in (0,1)$ tends to 0 in a certain way to be specified below. Our goal is to study properties of random subgraphs of $H(2,n)$ under $\mathbb{P}_p$.

Random subgraphs of finite tori with various edge sets were studied in quite some generality in [4, 5], and we now highlight the key results of these papers. Some of the theorems in [4, 5] apply to a general finite transitive graph, which in what follows will be denoted by $\mathbb{G}$. We also denote the number of vertices in $\mathbb{G}$ by $V = |\mathbb{G}|$ (V for "volume") and the vertex degree by Ω. Given a vertex $\mathbf{v}$ of $\mathbb{G}$, we shall write $C(\mathbf{v})$ for the *connected component* or *cluster* containing $\mathbf{v}$, and

---

$|C(\mathbf{v})|$ for the number of vertices in $C(\mathbf{v})$. Further, we let $\chi(p)$ be the expected size of the cluster containing $\mathbf{v}$, that is

$$\chi(p) = \mathbb{E}_p[|C(\mathbf{v})|]. \tag{1.1}$$

(Note that, by transitivity, this is independent of the choice of $\mathbf{v}$.) Then in [4, 5] the *critical threshold* $p_c = p_c(\mathbb{G}, \lambda)$ of a finite transitive graph $\mathbb{G}$ is defined to be the unique solution to the equation

$$\chi(p_c) = \lambda V^{1/3}, \tag{1.2}$$

where $\lambda > 0$ is a sufficiently small constant. (See [5] for details concerning the precise constraints on the size of λ.)

In [4], cluster sizes were investigated for graphs $\mathbb{G}$ satisfying the so-called *triangle condition*. In [5], the triangle condition was established for certain types of graphs $\mathbb{G}$, including the Hamming graph $H(d, n)$ of a general dimension $d \geq 1$. We shall now describe these results briefly in order to set up our own scene.

Let $\mathcal{C}_{\max}$ denote a cluster of maximum size, where we may pick any such cluster if it is not unique. Then $|\mathcal{C}_{\max}|$ is the maximum cluster size, that is

$$|\mathcal{C}_{\max}| = \max\{|C(\mathbf{v})| : \mathbf{v} \in \mathbb{G}\}. \tag{1.3}$$

The main theorems in [4] concern the scaling of $\chi(p)$ and bounds on $|\mathcal{C}_{\max}|$ in graphs satisfying the triangle condition as $V \to \infty$. Specifically, it is shown in [4, 5] that, if p_c is as in (1.2) and

$$p = p_c + \frac{\varepsilon}{\Omega}, \tag{1.4}$$

then, for all $\omega \geq 1$ and all ε such that $\varepsilon V^{1/3} \to -\infty$, asymptotically the expected cluster size $\chi(p)$ satisfies

$$\chi(p) = \frac{1 + O(\Omega^{-1}) + O(V^{-1/3})}{|\varepsilon|}. \tag{1.5}$$

With regard to the maximum cluster size, as $V \to \infty$,

$$\mathbb{P}_p\Big(\frac{\chi^2(p)}{3600\omega} \leq |\mathcal{C}_{\max}| \leq 2\chi^2(p)\log(V/\chi^3(p))\Big) \geq \Big(1 + \frac{36\chi^3(p)}{\omega V}\Big)^{-1} - \sqrt{e}[2\log(V/\chi^3(p))]^{-3/2}. \tag{1.6}$$

The above describes the behaviour of the mean and maximum cluster sizes for *subcritical* p values, which are p values satisfying $\varepsilon V^{1/3} \to -\infty$; in particular, the bounds apply to $H(2, n)$. For a constant $\Lambda > 0$, the *critical window* is defined as the interval of all $p = p_c + \frac{\varepsilon}{\Omega}$ such that $|\varepsilon| \leq \Lambda V^{-1/3}$. Theorem 1.3 in [4] shows that, for some constant $b = b(\Lambda)$, the maximum cluster size inside the critical window satisfies

$$\mathbb{P}_p\Big(\omega^{-1}V^{2/3} \leq |\mathcal{C}_{\max}| \leq \omega V^{2/3}\Big) \geq 1 - \frac{b}{\omega}. \tag{1.7}$$

The corresponding results in [4, 5] are significantly weaker in the case $p = p_c + \frac{\varepsilon}{\Omega}$ where $\varepsilon^3 V \to \infty$ (that is, when p is *above the critical window* or *supercritical*). In particular, only upper bounds on the maximum cluster size are established therein. More precisely, it is proved in [4] that, for all $\omega \geq 1$,

$$\mathbb{P}_p\Big(|\mathcal{C}_{\max}| \geq \omega(V^{2/3} + \varepsilon V)\Big) \leq \frac{21}{\omega}. \tag{1.8}$$

The problem with this result is that it does not imply that p_c as defined in (1.2) actually *is* the critical value, and thus that $p = p_c + \frac{\varepsilon}{\Omega}$ with $\varepsilon^3 V \to \infty$ really is above the critical window.

Indeed, to prove that this is the case, one additionally needs a *lower bound* on the maximum connected component size. No such results are established in [4, 5], and we expect that the geometry of the graphs under consideration plays a crucial role in lower bounding the largest cluster size.

The aim of this paper is to establish the asymptotics of the maximum supercritical cluster for the 2-dimensional Hamming graph $H(2, n)$. Throughout our proofs we shall use the phrase "with high probability" (abbreviated as "**whp**") to mean "with probability tending to 1 as $V \to \infty$". Also, "with very high probability" (abbreviated as "**wvhp**") will mean "with probability at least $1 - O(V^{-3})$ as $V \to \infty$". All unspecified limits are as $V \to \infty$. Given an event E, $I[E]$ will denote the indicator of E. We write $\mathbb{P}(\cdot)$ for a generic probability measure (for instance, the probability measure corresponding to a sequence of i.i.d. binomial random variables), which may vary from situation to situation. We use the O_p and o_p notations in the standard way (see e.g. Janson, Łuczak and Ruciński [17]). For example, if (X_n) is a sequence of random variables, then $X_n = O_p(1)$ means "X_n is bounded in probability" and $X_n = o_p(1)$ means that X_n converges to zero in probability as $n \to \infty$. We shall also use the asymptotic $o(), O(), \Omega(), \Theta()$ notations (without the subscript "p") in the standard way, and again referring to the regime where $V \to \infty$. We write $f(V) \gg g(V)$ (resp. $f(V) \ll g(V)$) when $g(V) = o(f(V))$ (resp. $f(V) = o(g(V))$) as $V \to \infty$. Throughout, the symbol "$\sim$" refers to estimates of the *leading order* as $V \to \infty$, with unspecified constants and thus uncontrolled error terms. Finally, we denote by C a generic (unspecified) positive constant, which may change from line to line. We shall interchange this use of C with the $O()$ notation.

1.1. **The model.** We consider the Hamming graph $H(d, n)$, and take the edge probability p as in (1.4). Let us note that [4, Theorem 1.5] establishes that, for a graph $\mathbb{G}$ satisfying the triangle condition,

$$1 - \chi(p_c)^{-1} \leq \Omega p_c \leq 1 - \chi(p_c)^{-1} + O(\Omega^{-1}). \tag{1.9}$$

When $\mathbb{G} = H(d, n)$, then $\Omega = d(n - 1)$ and $\chi(p_c) = \lambda V^{1/3} = \lambda n^{d/3}$. Therefore, if $\varepsilon = \Theta(V^{-1/3})$, then

$$p = \frac{1 + \varepsilon}{\Omega} = p_c + \frac{\varepsilon}{\Omega}(1 + O(1)), \tag{1.10}$$

while for p outside the critical window,

$$p = \frac{1 + \varepsilon}{\Omega} = p_c + \frac{\varepsilon}{\Omega}(1 + o(1)). \tag{1.11}$$

Since in the case $d = 2$, $\Omega^{-2} = o(\frac{\varepsilon}{\Omega})$ for $\varepsilon \gg V^{-1/3}$, the critical value defined in [4, 5] agrees asymptotically with the value $\frac{1+\varepsilon}{d(n-1)} = \frac{1+\varepsilon}{\Omega}$. This shows that we are working in the correct range of p values. For $d \geq 3$, (1.10)–(1.11) may not necessarily be valid, and we shall discuss this issue in more detail in Section 1.2.

From now on, we concentrate on the supercritical case, that is $\varepsilon \gg V^{-1/3} = n^{-d/3}$. Our main result is the following:

Theorem 1.1 (The supercritical phase for $H(2, n)$). *Consider the 2-dimensional Hamming graph $H(2, n)$. Let $p = p_c + \frac{\varepsilon}{\Omega}$ and let $n^{-2/3}(\log n)^{1/3} \ll \varepsilon \ll 1$. Then*

$$|\mathcal{C}_{\max}| = 2\varepsilon n^2(1 + o_p(1)). \tag{1.12}$$

Theorem 1.1 shows that, when $n^{-2/3}(\log n)^{1/3} \ll \varepsilon \ll 1$, the largest connected component satisfies a law of large numbers. Barring the factor $(\log n)^{1/3}$ in the lower bound on ε, Theorem 1.1 identifies the asymptotic size of the largest cluster all the way down to the critical threshold. Therefore, our result demonstrates that $p_c = \frac{1}{2(n-1)}$ really *is* the critical value for random subgraphs of the 2-dimensional Hamming graph. We believe that our proof can be adapted to deal with the case where $\varepsilon > 0$ is *fixed*. Here, the corresponding statement would be that $|\mathcal{C}_{\max}| \sim \zeta_{1+\varepsilon} V$ **whp**, where ζ_λ is the survival probability of a Poisson branching process with mean offspring λ. Since the proof of Theorem 1.1 is the most challenging for ε as close as possible to the critical window $V^{-1/3}$, we choose not to consider the case of constant ε in this paper. Prior to giving a proof of Theorem 1.1, we discuss its statement in more detail in Section 1.2 below. Therein we also include some conjectures concerning Hamming graphs of a general dimension d.

1.2. **Discussion and heuristics.** We first sketch an intuitive picture justifying the definition of p_c from [4, 5] given in (1.2). This picture relies on a branching process approximation for $p < p_c$. We expect random clusters in our model to exhibit behaviour similar to that of a subcritical branching process. Therefore, from the theory of branching processes, if $p = p_c + \frac{\varepsilon}{\Omega}$ is just below the critical point (for instance, if $\varepsilon < 0$), then we should have

$$\mathbb{P}_p(|C(\mathbf{v})| \geq k) \sim \frac{C}{\sqrt{k}} e^{-\frac{1}{2}k\varepsilon^2(1+o(1))}, \tag{1.13}$$

which in turn implies that

$$\chi(p) = \mathbb{E}_p[|C(\mathbf{v})|] \sim \int_0^\infty x^{-1/2} e^{-\frac{1}{2}x\varepsilon^2(1+o(1))} dx \sim \int_0^{\varepsilon^{-2}} x^{-1/2} dx \sim \varepsilon^{-1}. \tag{1.14}$$

Thus in fact

$$\mathbb{P}_p(|C(\mathbf{v})| \geq k) \sim \frac{C}{\sqrt{k}} e^{-\frac{k}{\chi^2(p)}\Omega(1)}, \tag{1.15}$$

and hence for subcritical p (possibly up to logarithmic corrections)

$$|\mathcal{C}_{\max}| \sim \chi(p)^2 \quad \textbf{whp}. \tag{1.16}$$

On the other hand, there should be a connected component dominating all the others in size in the case $p > p_c$. One way to express this intuitive statement is to impose that

$$\chi(p) = \mathbb{E}_p[|C(\mathbf{v})|] \sim \mathbb{E}_p[|C(\mathbf{v})|I[\mathbf{v} \in \mathcal{C}_{\max}]] = \frac{1}{V}\mathbb{E}_p[|\mathcal{C}_{\max}|^2]. \tag{1.17}$$

Naturally, the meaning of formula (1.17) is, in essence, that the main contribution to the expected size of a cluster of any particular vertex $\mathbf{v}$ is from those configurations where this vertex lies in the largest component.

Note that (1.17) could be taken as a *defining property* of supercritical behaviour. Then the *critical window* can be defined as the interval of p values where the subcritical and supercritical pictures coincide. In other words, if p lies within the critical window then both (1.16) and (1.17) should be satisfied.

Assume further that sufficient amount of the concentration of measure exhibited by $|\mathcal{C}_{\max}|$ in the subcritical regime (as implied by (1.16)) carries through to the critical window, so that $\mathbb{E}_p[|\mathcal{C}_{\max}|^2] \sim \chi(p)^4$. It then follows that, for p inside the critical window,

$$\chi(p) \sim \frac{1}{V}\mathbb{E}_p[|\mathcal{C}_{\max}|^2] \sim \frac{1}{V}\chi(p)^4; \tag{1.18}$$

and hence, inside the critical window,

$$\chi(p) \sim V^{1/3}. \tag{1.19}$$

This provides a rationale for the definition (1.2) of the critical threshold p_c. In conclusion, the above heuristic demonstrates that branching process approximations in the subcritical case and the domination of the expected cluster size by the maximum cluster size in the supercritical case together imply that (1.2) is the "right" definition for p_c.

At this point, we emphasise that subcritical branching process approximations are only likely to be valid for a random graph sufficiently mean-field in character, in the sense that its geometry is of little significance for the structure of its random subgraphs. This is the case for sufficiently high-dimensional random graphs, but cannot be expected to hold for low-dimensional random graphs, as indicated in [7, 8]. For random subgraphs of the torus with nearest-neighbour bonds in a sufficiently high (but constant) dimension, as well as for the torus with sufficiently spread-out bonds in dimension greater than 6, it is shown in [12] that the largest critical connected component is of order $V^{2/3}$, with logarithmic corrections in the lower bound. Accordingly, assuming universality in high-dimensional finite range percolation, one can expect classical random graph asymptotics at criticality to be valid for random subgraphs of the torus when $d > 6$ for general choices of finite-range edges. On the other hand, the results of [7, 8] suggest that random graph asymptotics at the phase transition threshold are *not* valid for random subgraphs of the d-dimensional torus when $d < 6$.

We close this section with a few conjectures. The present paper investigates the scaling of the largest connected component in supercritical percolation on the Hamming graph $H(2, n)$. Many random graph models are well known to satisfy what is sometimes referred to as a *discrete duality principle* (see for instance [1, Section 10.5]). This is the principle that the size of the second largest supercritical component is asymptotically close in distribution to the size of the largest subcritical component for an appropriate choice of subcritical edge probability. This notion of duality is closely related to the duality exhibited by branching processes [2, 10, 11, 18], or [1, Section 10.4]. We expect the Hamming graph $H(2, n)$ to follow the discrete duality principle. More precisely, letting $|\mathcal{C}_{(2)}|$ to be the size of the second largest component, we conjecture that

$$|\mathcal{C}_{(2)}| = 2\varepsilon^{-2} \log(\varepsilon^3 V)(1 + o_{\mathrm{p}}(1)). \tag{1.20}$$

For the Hamming graph $H(d, n)$ of an arbitrary dimension d, we conjecture that critical p values are of the form

$$p = \sum_{i=1}^{\lceil d/3 \rceil} a_i n^{-i} + \frac{\lambda}{n^{1+d/3}}, \tag{1.21}$$

where λ is an arbitrary constant, and the coefficients $a_i = a_i(d)$ are independent of n. Note that $p_c = \frac{1}{\Omega} = \frac{1}{d(n-1)}$ corresponds to $a_i = a_i(d) = \frac{1}{d}$ for all $i \geq 1$. We believe that, when d is sufficiently large, there exists an i such that $a_i(d) \neq \frac{1}{d}$. In particular, if this is indeed true, then for $\varepsilon = \Theta(V^{-1/3})$ and d sufficiently large, the edge probability $p = p_c + \frac{\varepsilon}{\Omega}$ is *not* the same as $p = \frac{1+\varepsilon}{\Omega}$. For $d = 2$, however, these two choices *do* agree.

For p inside the critical window, $|\mathcal{C}_{\max}|$ should be of the order $V^{2/3} = n^{2d/3}$. Below the critical window, we expect that the average cluster size satisfies $\chi(p) \sim [\Omega(p_c - p)]^{-1}$, while the maximum cluster size satisfies

$$|\mathcal{C}_{\max}| \sim 2\chi(p)^2 \log \left(V/\chi(p)^3 \right) \quad \textbf{whp}. \tag{1.22}$$

Note that [4] establishes in full only the upper bound part of (1.22), the corresponding best lower bound therein being $|\mathcal{C}_{\max}| \geq \chi(p)^2/(3600\omega)$ **whp** for ω large (cf. (1.6)). (It is the upper bound, however, that is relevant for locating the phase transition window.)

We anticipate that above the critical window

$$|\mathcal{C}_{\max}| \sim 2\varepsilon V \quad \textbf{whp}, \tag{1.23}$$

where $\varepsilon = \Omega(p - p_c)$. Possibly, the methodology of [13] can be adapted to prove that (1.21) does indeed hold for p within the critical window, and will further allow one to compute the numerical values of $a_i(d)$. The proof of this conjecture would enable an extension to random subgraphs of $H(d,n)$ of the phase transition description available for the classical binomial random graph $G(n,p)$.

2. OVERVIEW OF THE PROOF OF THEOREM 1.1

This section contains an extensive overview of the proof of our main result, breaking it down into a number of key propositions and lemmas. The first of these estimates the asymptotic probability of a connected cluster being large.

Proposition 2.1 (The cluster tail). *Set $p = p_c + \frac{\varepsilon}{\Omega}$. Let $V^{-1/3} \ll \varepsilon \ll 1$, and let $\eta \ll \varepsilon$ satisfy $\eta V \gg \varepsilon^{-2}$ as $V \to \infty$. Then for every N_V such that $\varepsilon^{-2} \ll N_V \leq \eta V$,*

$$\mathbb{P}_p(|C(\mathbf{v})| \geq N_V) = 2\varepsilon(1 + o(1)). \tag{2.1}$$

Proposition 2.1 consists of two parts, corresponding to the upper and lower bounds. These are re-stated separately in Section 4 as Lemmas 4.2 and 4.3, and proved in Sections 4 and 6 respectively.

Set

$$Z_{\geq k} = \sum_{\mathbf{v} \in H(2,n)} I[|C(\mathbf{v})| \geq k], \tag{2.2}$$

the number of vertices in clusters of size at least k. The following proposition upper bounds the variance of $Z_{\geq N_V}$ for some suitably large N_V ($N_V \gg \varepsilon^{-2}$); the proof can be found in Section 7.

Proposition 2.2 (Concentration of the number of vertices in large components). *Set $p = p_c + \frac{\varepsilon}{\Omega}$ and let $V^{-1/3}(\log V)^{1/3} \ll \varepsilon \ll 1$. Then there exists $N_V \gg \varepsilon^{-2}$ such that, for every $\delta > 0$,*

$$\mathbb{P}_p\Big(|Z_{\geq N_V} - \mathbb{E}_p[Z_{\geq N_V}]| \geq \delta\varepsilon V\Big) = o(1). \tag{2.3}$$

Armed with Propositions 2.1 and 2.2, we may prove the upper bound on $|\mathcal{C}_{\max}|$ in Theorem 1.1.

Proof of the upper bound part of Theorem 1.1. We shall show that **whp**, $|\mathcal{C}_{\max}| \leq 2\varepsilon V(1 + o(1))$. To this end, let us choose η satisfying both $\eta \ll \varepsilon$ and $\eta^3 V \gg 1$, and N_V as in Proposition 2.2. By Proposition 2.1, for any fixed $\delta > 0$,

$$\mathbb{P}_p\Big(|Z_{\geq \eta V} - 2\varepsilon V| \geq \delta\varepsilon V\Big) \leq \mathbb{P}_p\Big(|Z_{\geq N_V} - \mathbb{E}_p[Z_{\geq N_V}]| \geq \delta\varepsilon V/3\Big)$$

$$+ \mathbb{P}_p\Big(|Z_{\geq \eta V} - Z_{\geq N_V}| \geq \delta\varepsilon V/3\Big), \tag{2.4}$$

provided that V is large enough. More precisely, the volume V must be such that

$$\Big| \mathbb{E}_p[Z_{\geq N_V}] - 2\varepsilon V \Big| \leq \delta\varepsilon V/3, \tag{2.5}$$

or, equivalently (using transitivity),

$$\left| \mathbb{P}_p(|C(\mathbf{v})| \geq N_V) - 2\varepsilon \right| \leq \delta\varepsilon/3. \tag{2.6}$$

By Proposition 2.2,

$$\mathbb{P}_p\Big(|Z_{\geq N_V} - \mathbb{E}_p[Z_{\geq N_V}]| \geq \delta\varepsilon V/3\Big) = o(1). \tag{2.7}$$

Further, since $N_V \gg \varepsilon^{-2}$, Proposition 2.1 and Markov's inequality together yield that

$$\mathbb{P}_p\Big(|Z_{\geq \eta V} - Z_{\geq N_V}| \geq \delta\varepsilon V/3\Big) \leq \frac{3\mathbb{E}_p[Z_{\geq N_V} - Z_{\geq \eta V}]}{\delta\varepsilon V} = o(1). \tag{2.8}$$

Thus the random variable $Z_{\geq \eta V}$ is concentrated around $2\varepsilon V$; in other words, the number of vertices in connected components of size at least ηV is close to $2\varepsilon V$ **whp**. But on the event $\{Z_{\geq \eta V} \geq 1\}$, we have that

$$|\mathcal{C}_{\max}| \leq Z_{\geq \eta V}, \tag{2.9}$$

and so it follows that

$$|\mathcal{C}_{\max}| \leq 2\varepsilon V(1 + o_{\mathrm{p}}(1)). \tag{2.10}$$

∎

It remains to establish a matching lower bound for $|\mathcal{C}_{\max}|$, and we shall do this via a "sprinkling" argument. (Sprinkling is sometimes referred to as the "two-round exposure", see [17, Chapter 1].) This part of our proof is based on two results below. Before we state them, we need to introduce some more notation.

For each i, the *i-th horizontal line* of $H(2, n)$ is defined to be the set $\{(i, x) : x = 1, \ldots, n\}$ of vertices with first co-ordinate i; similarly the set $\{(x, i) : x = 1, \ldots, n\}$ of vertices with the second co-ordinate equal to i constitutes the *i-th vertical line*. A vertex belonging to a given line is said to be an *element* of that line.

Proposition 2.3 (Lower bound on the number of line elements in a large cluster)**.** *Set $p = p_c + \frac{\varepsilon}{\Omega}$. There exists a constant $C > 0$ such that the following holds. Choose $V^{-1/3} \ll \varepsilon \ll 1$ and $\eta \ll \varepsilon$ such that $\eta V \gg \varepsilon^{-2}$ and $\eta V/n \geq C \log n$ for n sufficiently large. Then **whp** for every cluster of size at least ηV, there are at least $\frac{5n}{8}$ horizontal lines each with at least $\eta V/(4n)$ elements contained in the cluster.*

The proof of Proposition 2.3 is deferred to Section 6. Assuming it holds, we now prove that the second round exposure will join together any pair of large clusters formed during the first round.

Lemma 2.4 (Sprinkling)**.** *Set $p = p_c + \frac{\varepsilon}{\Omega}$. Choose $V^{-1/3} \ll \varepsilon \ll 1$. There exists $\eta = \eta(V)$ satisfying $\eta \ll \varepsilon$, $\eta^3 V \gg 1$, $\eta V \gg n$, and such that the following holds. Let S_1, S_2 be disjoint sets of vertices both containing at least $\eta V/(4n)$ elements of at least $5n/8$ horizontal lines (possibly different lines for S_1 and S_2). Then*

$$\mathbb{P}_{\eta/\Omega}(S_1 \longleftrightarrow S_2) \geq 1 - o\Big(\Big(\frac{\eta}{\varepsilon}\Big)^2\Big). \tag{2.11}$$

Proof. Choose two disjoint vertex sets S_1 and S_2 each containing at least $\eta V/4$ elements in at least $5n/8$ horizontal lines. Then S_1 and S_2 must have at least $n/4$ such lines in common, that is *both* S_1 and S_2 contain at least $\eta V/(4n) = \eta n/4$ elements of these lines. Along the shared good lines,

there are at least $(\eta n)^2/16$ edges with one endpoint in S_1 and the other in S_2. All of these edges will be occupied independently under $\mathbb{P}_{\eta/\Omega}$, so

$$\mathbb{P}_{\eta/\Omega}(S_1 \longleftrightarrow\!\!\!\!\!/\,\, S_2) \leq \left(1 - \frac{\eta}{2(n-1)}\right)^{n(\eta n)^2/32}. \tag{2.12}$$

Since $1 - x \leq e^{-x}$,

$$\mathbb{P}_{\eta/\Omega}(S_1 \longleftrightarrow\!\!\!\!\!/\,\, S_2) \leq e^{-\eta^3 n^2/64}, \tag{2.13}$$

and so we need η such that

$$e^{-\eta^3 n^2/64} = o\!\left(\frac{\eta^2}{\varepsilon^2}\right). \tag{2.14}$$

We take $\eta = \sqrt{\varepsilon}V^{-1/6}$, so (using $\varepsilon \gg V^{-1/3}$) $\eta \ll \varepsilon$. Further, $\frac{\eta^2}{\varepsilon^2} = (\varepsilon V^{1/3})^{-1}$ and also

$$\eta^3 n^2 = \eta^3 V = \varepsilon^{3/2} V^{1/2} \gg 1. \tag{2.15}$$

We conclude that

$$e^{-\eta^3 n^2/64} = e^{-\varepsilon^{3/2} n/64} = e^{-(\varepsilon^3 V)^{1/2}/64} = o((\varepsilon V^{1/3})^{-1}), \tag{2.16}$$

which completes the proof. $\qquad\square$

We can now do the lower bound part of Theorem 1.1.

Proof of the lower bound part of Theorem 1.1. We choose η as in Proposition 2.3 and Lemma 2.4, and define p_- by the relation

$$p_- + (1 - p_-)\frac{\eta}{\Omega} = p. \tag{2.17}$$

Note that every configuration with edge probability p can be obtained in a unique way as follows. First construct a configuration by throwing in edges independently of one another with probability p_-; subsequently, "sprinkle" extra edges with probability $\frac{\eta}{\Omega}$, independently of one another and of the p_- configuration. In the final configuration, an edge is occupied precisely when it is occupied in either the p_- configuration, or when it is an edge that is added during the sprinkling procedure. Since $\eta \ll \varepsilon$,

$$p_- = p + o\!\left(\frac{\varepsilon}{n}\right). \tag{2.18}$$

Let $Z'_{\geq \eta V}$ denote the number of vertices in connected components of size at least ηV in the p_- configuration. Since δ in Proposition 2.2 is arbitrary, it implies that $Z'_{\geq \eta V} = 2\varepsilon V(1 + o_{\mathrm{p}}(1))$ **whp** after the first round of exposure; and by Proposition 2.3 **whp** every cluster of size at least ηV includes at least $\eta V/(4n)$ elements in at least $\frac{5n}{8}$ lines. Thus, under the measure $\mathbb{P}_{p_-}$, **whp** there are at most $\frac{2\varepsilon}{\eta}$ connected clusters of size at least ηV, and each of these connected components contains at least $\eta V/(4n)$ elements in at least $\frac{5n}{8}$ lines.

It now suffices to prove that **whp** the subsequent sprinkling procedure (second round of exposure) joins together every pair of clusters of size at least ηV. Indeed, if this is the case, then after the sprinkling we end up with a single connected component of size at least

$$Z'_{\geq \eta V} \geq 2\varepsilon V(1 + o_{\mathrm{p}}(1)). \tag{2.19}$$

Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be two vertices such that $C(\mathbf{v}_1) \neq C(\mathbf{v}_2)$, and $|C(\mathbf{v}_1)| \geq \eta V$ and $|C(\mathbf{v}_2)| \geq \eta V$. Let us take $S_1 = C(\mathbf{v}_1)$ and $S_2 = C(\mathbf{v}_2)$. By Proposition 2.3, we may assume that for both S_1

and S_2 one can find at least $\frac{5n}{8}$ lines (not necessarily the same ones for S_1 and S_2) each with at least $\eta V/(4n)$ elements in S_1 and S_2. Then, by Lemma 2.4,

$$\mathbb{P}_{\eta/\Omega}(S_1 \longleftrightarrow\!\!\!\!\!/\ S_2) = o\left(\left(\frac{\eta}{\varepsilon}\right)^2\right). \tag{2.20}$$

But **whp** there are at most $\left(\frac{2\varepsilon}{\eta}\right)^2(1 + o(1))$ choices for $\mathbf{v}_1$ and $\mathbf{v}_2$ with $|C(\mathbf{v}_1)| \geq \eta V$ and $|C(\mathbf{v}_2)| \geq \eta V$, and so a simple union bound implies that after sprinkling **whp** all connected components of size at least ηV are connected. By (2.19), this completes the proof. ∎

The remainder of the paper is organised as follows. In Section 3 we prove some auxiliary results relating to branching processes. In Section 4, we describe a *cluster exploration* procedure, and state key estimates for the tails of the cluster size distribution. In Section 5, we prove the upper bound part of Proposition 2.1, and further establish an upper bound on the number of elements per line in a large cluster. Section 6 contains a proof of the lower bound of Proposition 2.1, as well as a proof of Proposition 2.3. Finally, in Section 7, we give a proof of Proposition 2.2.

3. Total progeny of a Galton-Watson process

This section brings together some useful results from the theory of branching processes, which will play a key role at various stages in our proofs.

We consider a standard Galton-Watson process with offspring distribution of a random variable Z such that $\mathbb{E}[Z^2] < \infty$. We assume that with probability 1 the process begins with one individual. Also, Z will always have a binomial distribution $\mathrm{Bi}(N, p)$, where $N \in \mathbb{N}$ and $p \in [0, 1]$ is the Hamming graph edge probability. We write $\mathbb{P}_{N,p}$ for the probability measure corresponding to this process (implicitly assuming an underlying sample space and σ-field).

Let F be the total progeny or family size. Our aim is to prove the following three results concerning the distribution of F. Proposition 3.1 compares the distribution of F under the measures $\mathbb{P}_{N,p}$, $\mathbb{P}_{\tilde{N},p}$ for different values of N and $\tilde{N}$. Proposition 3.2 estimates the probability that the value of F is between ℓ and 2ℓ, for some (large) integer ℓ. Proposition 3.3 estimates the probability that F takes a value at least ℓ for some large integer ℓ.

Proposition 3.1 (Tails of total progeny in two binomial branching processes)**.** *Let $\ell \in \mathbb{N}$. Suppose that $N \in \mathbb{N}$ and $\tilde{N} = \tilde{N}(N)$ satisfy $N \geq \tilde{N}$. Further, assume that $\varepsilon = Np - 1$ is such that $\varepsilon \to 0$ and $\varepsilon \geq N^{-2/3}$; and that $\tilde{\varepsilon} = \tilde{N}p - 1 > 0$ and $|\varepsilon - \tilde{\varepsilon}| \leq \varepsilon/\log N$ as $N \to \infty$. Then, for some constant $C > 0$, as $N \to \infty$,*

$$|\mathbb{P}_{N,p}(F \geq \ell) - \mathbb{P}_{\tilde{N},p}(F \geq \ell)| \leq C\left(|\varepsilon - \tilde{\varepsilon}| + \frac{1}{N\ell^{1/2}} + \frac{1}{\ell^3}\right). \tag{3.1}$$

Proposition 3.2 (Bounds on the total progeny distribution)**.** *Let $N, \ell \in \mathbb{N}$. Suppose that $\varepsilon = Np - 1$ satisfies $\varepsilon \to 0$ and $\varepsilon \geq N^{-2/3}$ as $N \to \infty$. Then, for some constant $C > 0$, as $N \to \infty$,*

$$\mathbb{P}_{N,p}(F \in [\ell, 2\ell]) \leq \frac{C}{\sqrt{\ell}}. \tag{3.2}$$

Proposition 3.3 (Tails of the total progeny near criticality)**.** *Let $N, \ell \in \mathbb{N}$. Suppose that $\varepsilon = Np - 1$ satisfies $\varepsilon \to 0$ and $\varepsilon \geq N^{-2/3}$ as $N \to \infty$. Then, as $N \to \infty$,*

$$\mathbb{P}_{N,p}(F \geq \ell) = 2\varepsilon + O(\varepsilon^2) + O\left(\frac{1}{\sqrt{\ell}}\right). \tag{3.3}$$

We note that with a little care and minor modifications, Propositions 3.1–3.3 could be extended to the case where ε is a positive constant (i.e. strictly above the critical window). In the corresponding statement, (3.3) would have $\zeta_{1+\varepsilon}$ instead of 2ε, where, as before, ζ_λ denotes the non-extinction probability of a Poisson Galton-Watson process with mean family size λ. Our proofs of Propositions 3.1 and 3.2 will make use of the well known *Otter-Dwass formula*, which describes the distribution of the total progeny of a branching process, see [11, 21]. We begin by stating a special case of this formula (due to Otter) for a branching process starting with 1 individual. (The formula was later extended by Dwass to a process starting with r individuals, for arbitrary $r \in \mathbb{N}$, but we do not make use of the extension here.)

Lemma 3.4 (Otter-Dwass formula). *Let $Z_1, Z_2, Z_3, \ldots$ be i.i.d. random variables distributed as Z. Let $\mathbb{P}$ denote the Galton-Watson process measure. For all $l \in \mathbb{N}$,*

$$\mathbb{P}(F = k) = \frac{1}{l}\mathbb{P}(\sum_{i=1}^{l} Z_i = l - 1).$$

We now prove each of Propositions 3.1–3.3 in turn.

Proof of Proposition 3.1. It will be convenient for us to introduce new parameters $\lambda = pN$ and $\tilde{\lambda} = p\tilde{N}$. By assumption $\lambda, \tilde{\lambda} > 1$, so that both branching processes are supercritical. The total progeny size in the $\mathrm{Bi}(\tilde{N}, p)$ process will be denoted by $\tilde{F}$.

By Lemma 3.4, for each $k \in \mathbb{N}$,

$$\mathbb{P}_{N,p}(F = k) = \frac{1}{k}\mathbb{P}(\sum_{i=1}^{k} Z_i = k - 1), \tag{3.4}$$

where the Z_i are i.i.d. $\mathrm{Bi}(N, \lambda/N)$. We now investigate the asymptotics of the formula (3.4) as $N \to \infty$, for large integers $k = k(N)$. Our aim is to obtain estimates for $\mathbb{P}_{N,p}(F = k)$ sharp enough for the errors to be summable.

In outline, our calculation is as follows. First we demonstrate that, if $k \geq C_0 \varepsilon^{-2} \log(1/\varepsilon)$ for a sufficiently large constant C_0, then

$$\mathbb{P}_{N,p}(k \leq F < \infty) \leq k^{-4}. \tag{3.5}$$

Next we show that (3.5) implies an identical upper bound for $\mathbb{P}_{\tilde{N},p}(\tilde{F} = k)$. Subsequently, we prove that there exists a constant $\tilde{C} > 0$ such that, if $k \leq C_0 \varepsilon^{-2} \log(1/\varepsilon)$, then

$$|\mathbb{P}_{N,p}(F = k) - \mathbb{P}_{\tilde{N},p}(\tilde{F} = k)| \leq \frac{\tilde{C}}{k^{3/2}} \exp\left(-\frac{(k-1)\varepsilon^2}{2}\right)$$
$$\times \left|(1 + k\varepsilon)|\varepsilon - \tilde{\varepsilon}| + \frac{k}{N^3} + \frac{\varepsilon}{N} + \frac{1}{kN}\right|, \tag{3.6}$$

We can then sum the errors in (3.6) and (3.5) to show that, for any $\ell \in \mathbb{N}$,

$$|\mathbb{P}_{N,p}(\ell \leq F < \infty) - \mathbb{P}_{\tilde{N},p}(\ell \leq \tilde{F} < \infty)| \leq C\left(|\varepsilon - \tilde{\varepsilon}| + \frac{1}{N\ell^{1/2}} + \frac{1}{\ell^3}\right). \tag{3.7}$$

Finally, since $\lambda > 1$, we need to estimate $\mathbb{P}_{N,p}(F = \infty)$ and $\mathbb{P}_{\tilde{N},p}(\tilde{F} = \infty)$; we shall show that

$$\left|\mathbb{P}_{N,p}(F = \infty) - \mathbb{P}_{\tilde{N},p}(\tilde{F} = \infty)\right| \leq C|\varepsilon - \tilde{\varepsilon}|. \tag{3.8}$$

Combining the last two estimates yields Proposition 3.1. As we shall see below, several steps of our proof will also play a role in proving Propositions 3.2 and 3.3.

Let us make a start on the details. To show (3.5), we note that (3.4) implies

$$\mathbb{P}_{N,p}(F = k) \leq \frac{1}{k}\mathbb{P}(\sum_{i=1}^{k} Z_i \leq k), \tag{3.9}$$

where $\sum_{i=1}^{k} Z_i \sim \text{Bi}(kN, p)$. But if $Z \sim \text{Bi}(n, p)$, then (see for instance [15])

$$\mathbb{P}(X \leq np - t) \leq e^{-\frac{t^2}{2(np+\frac{t}{3})}}. \tag{3.10}$$

Applying (3.10) with $n = kN$, $p = \frac{1+\varepsilon}{N}$ and $t = np - k = k\varepsilon$ (and also using our assumption that $\varepsilon < 1$), we obtain that

$$\mathbb{P}_{N,p}(F = k) \leq \frac{1}{k}e^{-\frac{k\varepsilon^2}{2(1+\frac{4\varepsilon}{3})}} \leq \frac{1}{k}e^{-\frac{k\varepsilon^2}{5}}. \tag{3.11}$$

But if $k \geq C_0\varepsilon^{-2}\log(1/\varepsilon)$ for a sufficiently large $C_0 > 0$ then

$$\frac{1}{k}e^{-\frac{k\varepsilon^2}{4}} \leq k^{-4}, \tag{3.12}$$

so (3.5) follows.

We now show that (3.6) holds for $k \leq C_0\varepsilon^{-2}\log(1/\varepsilon)$. Clearly, (3.4) implies that, for each $k \in \mathbb{N}$,

$$\mathbb{P}_{N,p}(F = k) = \frac{1}{k}\binom{kN}{k-1}\left(\frac{\lambda}{N}\right)^{k-1}\left(1 - \frac{\lambda}{N}\right)^{kN-k+1}. \tag{3.13}$$

By Stirling's formula,

$$(m)_r := m(m-1)\ldots(m-r+1) = m^r \exp\left(-\frac{r^2}{2m} - \frac{r^3}{6m^2} + O\left(\frac{r^4}{m^3}\right)\right).$$

Applying the above approximation with $m = kN$ and $r = k - 1$ (and noting that $\frac{r^4}{m^3} = O(\frac{k}{N^3})$), we arrive at

$$\mathbb{P}_{N,p}(F = k) = \frac{1}{k}\frac{(k\lambda)^{k-1}}{(k-1)!}\exp\left(-\frac{(k-1)^2}{2kN} - \frac{(k-1)^3}{6k^2N^2} + O\left(\frac{k}{N^3}\right)\right)\left(1 - \frac{\lambda}{N}\right)^{kN-k+1}$$

$$= \frac{(k\lambda)^{k-1}}{k!}\exp\left(-\frac{k}{2N} + \frac{1}{N} - \frac{1}{2kN} - \frac{k}{6N^2} + O\left(\frac{1}{N^2} + \frac{k}{N^3}\right)\right)\left(1 - \frac{\lambda}{N}\right)^{kN-k+1}.$$

Observe that

$$\left(1 - \frac{\lambda}{N}\right)^{kN-k+1} = \exp\left((kN - k + 1)\log(1 - \frac{\lambda}{N})\right)$$

$$= \exp\left(-\lambda k + \frac{k\lambda}{N} - \frac{\lambda}{N} - \frac{\lambda^2 k}{2N} + \frac{\lambda^2 k}{2N^2} - \frac{\lambda^3 k}{3N^2} + O(\frac{1}{N^2} + \frac{k}{N^3})\right),$$

so that

$$\mathbb{P}_{N,p}(F = k) = \frac{k^{k-1}e^{-(k-1)}e^{-\lambda}}{k!}\exp((k-1)f(\lambda))\exp\left(-\frac{k(\lambda-1)^2}{2N} + \frac{1-\lambda}{N} - \frac{1}{2kN}\right)$$

$$\times \exp\left(\frac{k}{N^2}g(\lambda) + O\left(\frac{k}{N^3} + \frac{1}{N^2}\right)\right), \tag{3.14}$$

where

$$f(\lambda) = \log\lambda - (\lambda - 1), \qquad g(\lambda) = -\frac{1}{6} + \frac{\lambda^2}{2} - \frac{\lambda^3}{3}. \tag{3.15}$$

The Taylor expansion for $|\lambda - 1|$ small gives

$$f(\lambda) = -\frac{(\lambda - 1)^2}{2} + O(|\lambda - 1|^3). \tag{3.16}$$

But $\lambda - 1 = \varepsilon$ and $k \leq C_0 \varepsilon^{-2} \log(1/\varepsilon)$, where $N^{-2/3} \ll \varepsilon = o(1)$, and so $k|\lambda - 1|^3 = o(1)$, uniformly for all such k. The other error terms can be bounded similarly, and hence, uniformly for $k \leq C_0 \varepsilon^{-2} \log(1/\varepsilon)$,

$$\mathbb{P}_{N,p}(F = k) = (1 + o(1))\frac{k^{k-1}e^{-k}}{k!} \exp\left(-\frac{1}{2}(k-1)(\lambda - 1)^2\right). \tag{3.17}$$

We now compare $\mathbb{P}_{N,p}(F = k)$ to $\mathbb{P}_{\tilde{N},p}(\tilde{F} = k)$ for $k \leq C_0 \varepsilon^{-2} \log(1/\varepsilon)$. Write $\lambda/N = \tilde{\lambda}/\tilde{N}$, where $\tilde{\lambda} = \tilde{N}\lambda/N < \lambda$. Then a calculation similar to the one above shows that

$$\mathbb{P}_{\tilde{N},p}(\tilde{F} = k) = \frac{k^{k-1}e^{-\tilde{\lambda}-(k-1)}}{k!} \exp\left((k-1)f(\tilde{\lambda})\right)$$
$$\times \exp\left(-\frac{k(\tilde{\lambda}-1)^2}{2\tilde{N}} + \frac{1-\tilde{\lambda}}{\tilde{N}} - \frac{1}{2k\tilde{N}} + \frac{k}{\tilde{N}^2}g(\tilde{\lambda}) + O\left(\frac{k}{\tilde{N}^3} + \frac{1}{\tilde{N}^2}\right)\right).$$

By assumption, $\varepsilon = \lambda - 1 > 0$ and $\tilde{\varepsilon} = \tilde{\lambda} - 1 > 0$. Further,

$$f(\lambda) - f(\tilde{\lambda}) = (\tilde{\varepsilon} - \varepsilon)f'(1 + s), \tag{3.18}$$

for some $s \in (\tilde{\varepsilon}, \varepsilon)$, where

$$|f'(1 + s)| = \frac{s}{1 + s} \leq s, \quad s \geq 0, \tag{3.19}$$

and so

$$f(\lambda) - f(\tilde{\lambda}) = O(\varepsilon|\varepsilon - \tilde{\varepsilon}|). \tag{3.20}$$

We deduce that, for $\varepsilon \geq N^{-2/3}$ and all $k \leq C_0 \varepsilon^{-2} \log(1/\varepsilon)$,

$$|\mathbb{P}_{N,p}(F = k) - \mathbb{P}_{\tilde{N},p}(\tilde{F} = k)| = \frac{(k/e)^{k-1}e^{-\lambda}}{k!} \exp((k-1)f(\lambda))\big|\exp(x) - \exp(y)\big|$$

where

$$x = -\frac{k\varepsilon^2}{2N} + \frac{k}{N^2}g(\lambda) + O\left(\frac{k}{N^3} + \frac{\varepsilon}{N} + \frac{1}{kN}\right),$$
$$y = (\varepsilon - \tilde{\varepsilon}) - \frac{k\tilde{\varepsilon}^2}{2\tilde{N}} + \frac{k}{\tilde{N}^2}g(\tilde{\lambda}) + O(k\varepsilon|\varepsilon - \tilde{\varepsilon}|) + O\left(\frac{k}{\tilde{N}^3} + \frac{\tilde{\varepsilon}}{\tilde{N}} + \frac{1}{k\tilde{N}}\right).$$

Since $k \leq C_0 \varepsilon^{-2} \log(1/\varepsilon)$, $N^{-2/3} \ll \varepsilon = o(1)$ and $\tilde{\varepsilon} = o(\varepsilon)$, $x = o(1)$ and also all contributions to y are $o(1)$, except for the term $k\varepsilon|\varepsilon - \tilde{\varepsilon}|$. The latter is bounded for all k in the range if and only if $\log(1/\varepsilon)|\varepsilon - \tilde{\varepsilon}| = O(\varepsilon)$, which follows from our assumption that $|\varepsilon - \tilde{\varepsilon}| \leq \varepsilon/\log N$.

Now, $|x|$ and $|y|$ are bounded, and so for some constant C

$$|\exp(x) - \exp(y)| \leq C|x - y|. \tag{3.21}$$

Since $N^{-2} - \tilde{N}^{-2} = O(N^{-2}|\varepsilon - \tilde{\varepsilon}|)$, the contribution to $|x - y|$ due to the term $g(\lambda)k/N^2 - g(\tilde{\lambda})k/\tilde{N}^2$ is $O(kN^{-2}|\varepsilon - \tilde{\varepsilon}|) = o(k\varepsilon|\varepsilon - \tilde{\varepsilon}|)$, which gives

$$|x - y| \leq C(1 + k\varepsilon)|\varepsilon - \tilde{\varepsilon}| + O\left(\frac{k}{N^3} + \frac{\varepsilon}{N} + \frac{1}{kN}\right). \tag{3.22}$$

Hence for all $k \leq C_0 \varepsilon^{-2} \log(1/\varepsilon^3)$

$$|\mathbb{P}_{N,p}(F = k) - \mathbb{P}_{\tilde{N},p}(\tilde{F} = k)| \leq C \frac{(k/e)^{k-1} e^{-\lambda}}{k!} \exp\left(-\frac{(k-1)\varepsilon^2}{2}\right)$$

$$\times \left((1 + k\varepsilon)|\varepsilon - \tilde{\varepsilon}| + \frac{k}{N^3} + \frac{\varepsilon}{N} + \frac{1}{kN}\right), \qquad (3.23)$$

which yields (3.6).

Summing the estimates (3.5) and (3.6) over $k \geq \ell$,

$$|\mathbb{P}_{N,p}(\ell \leq F < \infty) - \mathbb{P}_{\tilde{N},p}(\ell \leq \tilde{F} < \infty)|$$

$$\leq C \sum_{k \geq \ell} \left((1 + k\varepsilon)|\varepsilon - \tilde{\varepsilon}| + \frac{k}{N^3} + \frac{\varepsilon}{N} + \frac{1}{kN}\right) \frac{\exp(-(k-1)\varepsilon^2/2)}{k^{3/2}} + \sum_{k \geq \ell} k^{-4}.$$

The final contribution is $O(\ell^{-3})$; for the remaining terms observe that

$$\sum_{k \geq \ell} k^{-a} \exp(-(k-1)\varepsilon^2/2) \leq \begin{cases} C\ell^{1-a} & \text{for } a > 1, \\ C\varepsilon^{-2-2a} & \text{for } a < 1, \end{cases} \qquad (3.24)$$

which yields (with a suitably adjusted value of C)

$$|\mathbb{P}_{N,p}(\ell \leq F < \infty) - \mathbb{P}_{\tilde{N},p}(\ell \leq \tilde{F} < \infty)| \leq C\left(|\varepsilon - \tilde{\varepsilon}| + \frac{1}{\varepsilon N^3} + \frac{\varepsilon}{N\ell^{1/2}} + \frac{1}{\ell^{3/2}N} + \frac{1}{\ell^3}\right)$$

$$\leq C\left(|\varepsilon - \tilde{\varepsilon}| + \frac{1}{N\ell^{1/2}} + \frac{1}{\ell^3}\right). \qquad (3.25)$$

To prove (3.8), we need to estimate $|a - \tilde{a}|$, where $a = \mathbb{P}_{N,p}(F < \infty)$ and $\tilde{a} = \mathbb{P}_{\tilde{N},p}(\tilde{F} < \infty)$. The quantities $a, \tilde{a}$ respectively are the smallest positive roots of the equations

$$a = (1 + \frac{\lambda}{N}(a - 1))^N, \qquad \text{and} \qquad \tilde{a} = (1 + \frac{\lambda}{N}(\tilde{a} - 1))^{\tilde{N}}. \qquad (3.26)$$

Using the convexity of probability generating functions and the supercriticality of the branching processes in question, the equations in (3.26) each have precisely one root a and $\tilde{a}$ respectively in the interval $[0, 1)$.

The proof is divided into two main steps. In the first step, we prove that $1 - a = 2\varepsilon + O(\varepsilon^2)$, which also implies that $|a - \tilde{a}| = o(\varepsilon)$ when $|\varepsilon - \tilde{\varepsilon}| = o(\varepsilon)$, so that we may use a Taylor expansion. In the second main step, we prove that $|a - \tilde{a}| \leq C|\varepsilon - \tilde{\varepsilon}|$.

To prove that $1 - a = 2\varepsilon + O(\varepsilon^2)$, we expand the right hand side of (3.26) to obtain

$$a - 1 = \binom{N}{1} \frac{\lambda}{N}(a - 1) + \binom{N}{2} \left(\frac{\lambda}{N}\right)^2 (a - 1)^2 + O(|a - 1|^3)$$

$$1 = \lambda + \frac{N-1}{2N}\lambda^2(a - 1) + O(|a - 1|^2)$$

$$1 = 1 + \varepsilon + \frac{N-1}{2N}(1 + \varepsilon)^2(a - 1) + O(|a - 1|^2),$$

so that

$$(1 - a)(1 + 2\varepsilon + \varepsilon^2 - N^{-1} - 2N^{-1}\varepsilon + O(N^{-1}\varepsilon^2)) = 2\varepsilon + O(|1 - a|^2),$$

and so, again using that $\varepsilon \geq N^{-2/3}$,

$$1 - a = 2\varepsilon + O(\varepsilon^2). \qquad (3.27)$$

To prove that $|a - \tilde{a}| \leq C|\varepsilon - \tilde{\varepsilon}|$, we use that

$$a - \tilde{a} = (1 + \frac{\lambda}{N}(a-1))^N - (1 + \frac{\lambda}{N}(\tilde{a}-1))^{\tilde{N}} = f_{\tilde{N}}(a) - f_{\tilde{N}}(\tilde{a}) + f_{\tilde{N}}(a)\Big((1 + \frac{\lambda}{N}(a-1))^{N-\tilde{N}} - 1\Big), \quad (3.28)$$

where $f_{\tilde{N}}(x) = (1 + \frac{\lambda}{N}(x-1))^{\tilde{N}}$. Note first that

$$(1 + \frac{\lambda}{N}(a-1))^{N-\tilde{N}} - 1 = O\Big(\frac{N - \tilde{N}}{N}(1-a)\Big). \quad (3.29)$$

Further, since $|a - \tilde{a}| = o(|a - 1|)$, we have that $f_{\tilde{N}}(a) \leq 2$. Also,

$$f'_{\tilde{N}}(x) = \frac{\tilde{N}\lambda}{N}\big(1 + \frac{\lambda}{N}(x-1)\big)^{\tilde{N}-1}, \quad (3.30)$$

$$f''_{\tilde{N}}(x) = \frac{\tilde{N}\lambda}{N}\frac{(\tilde{N}-1)\lambda}{N}\big(1 + \frac{\lambda}{N}(x-1)\big)^{\tilde{N}-2}, \quad (3.31)$$

and it is not hard to see that $f''_{\tilde{N}}(x) = O(1)$ uniformly for $x \in [\tilde{a}, a]$. Hence

$$a - \tilde{a} = (a - \tilde{a})f'_{\tilde{N}}(\tilde{a}) + O\Big(\frac{N - \tilde{N}}{N}(1-a)\Big) + O\big((a - \tilde{a})^2\big), \quad (3.32)$$

so that

$$a - \tilde{a} = O\Big(\frac{(N - \tilde{N})(1-a)}{N|1 - f'_{\tilde{N}}(\tilde{a})|}\Big) + O\Big(\frac{(a - \tilde{a})^2}{|1 - f'_{\tilde{N}}(\tilde{a})|}\Big). \quad (3.33)$$

A closer inspection of $f'_{\tilde{N}}(x)$ yields that $f'_{\tilde{N}}(\tilde{a}) - 1 = \varepsilon + o(\varepsilon)$, so that

$$|a - \tilde{a}| \leq O\Big(\frac{|N - \tilde{N}|}{N}\Big) + O\Big(\frac{(a - \tilde{a})^2}{1 - a}\Big) \leq C|\varepsilon - \tilde{\varepsilon}|. \quad (3.34)$$

This completes the proof of Proposition 3.1. ∎

Proof of Proposition 3.2. By (3.17), for all $k \leq C_0 \varepsilon^{-2} \log(1/\varepsilon)$

$$\mathbb{P}_{N,p}(F = k) = (1 + o(1))\frac{(k/e)^{k-1}e^{-\lambda}}{k!} \exp\Big(-\frac{1}{2}(k-1)\varepsilon^2\Big).$$

Also (provided C_0 is large enough) for $k \geq C_0 \varepsilon^{-2} \log(1/\varepsilon)$,

$$\mathbb{P}_{N,p}(F = k) \leq k^{-4}.$$

Summing over $\ell \leq k \leq 2\ell$ we obtain

$$\mathbb{P}_{N,p}(\ell \leq F \leq 2\ell) \leq C \sum_{\ell \leq k \leq 2\ell} \Big(\frac{1}{k^{3/2}}e^{-k\varepsilon^2/2} + k^{-4}\Big) \leq \frac{C}{\ell^{1/2}},$$

where the constant C was adjusted within the final inequality. ∎

Proof of Proposition 3.3. We have

$$\mathbb{P}_{N,p}(F \geq \ell) = 1 - \mathbb{P}_{N,p}(F < \infty) + \mathbb{P}_{N,p}(\ell \leq F < \infty). \quad (3.35)$$

By (3.27), the term $1 - \mathbb{P}_{N,p}(F < \infty)$ is $2\varepsilon + O(\varepsilon^2)$. Calculations similar to those in the proof of Proposition 3.2 show that the final term is bounded by $O(\ell^{-1/2})$, which completes the proof. ∎

4. Component exploration and strategy of proof

We take an initial vertex $\mathbf{v}_0 = (x_0, y_0)$ and *explore* its cluster, $C(\mathbf{v}_0)$, by *exploring* the vertices in that cluster successively one at a time, in a breadth-first order. Exploring a vertex (x, y) means that we consider all the edges (x, j) for $j \neq y$ in the order of increasing j, and decide for each one in turn if it is open with probability p or closed with probability $1 - p$; then we do the same for the edges (i, y) for $i \neq x$ in the order of increasing i.

Let us introduce colours as follows. At time t, all vertices that have not yet been explored and are not yet contained in $C(\mathbf{v}_0)$ are *white*. All unexplored vertices connected to $\mathbf{v}_0$ (that is, included in $C(\mathbf{v}_0)$) at time t are *green*. All explored vertices are *red*. (Thus, in particular, at time 0 all vertices are white except for $\mathbf{v}_0$, which is green.) In fact, we need to modify this exploration process slightly as follows: when exploring a green vertex we only consider those of its edges where the other endpoint of the edge is white. If such an edge is found to be open, then we colour its other endpoint green.

Let $C_t(\mathbf{v}_0)$ be the set of vertices included in the cluster of $\mathbf{v}_0$ by the time t. Let also $G_t(\mathbf{v}_0)$ be the set of green vertices in the cluster at time t. Thus $C_t(\mathbf{v}_0)$ consists of all green and red vertices at time t, and $C_t(\mathbf{v}_0) \setminus G_t(\mathbf{v}_0)$ is the set of red vertices. All the remaining vertices in the graph are white.

Let $T_{\mathbf{v}_0}$ denote the smallest time t when there are no green vertices remaining, that is $T_{\mathbf{v}_0} = \inf\{t : |G_t(\mathbf{v}_0)| = 0\}$. Note that $T_{\mathbf{v}_0} = |C(\mathbf{v}_0)|$, the size of the cluster of vertex $\mathbf{v}_0$. Choose a parameter $\eta = \eta(\varepsilon, V)$ such that $0 < \eta \ll \varepsilon$ and let

$$T = T_{\mathbf{v}_0} \wedge \lceil \eta V \rceil \tag{4.1}$$

be the minimum of $T_{\mathbf{v}_0}$ and $\lceil \eta V \rceil$.

Given an integer $i \in \{1, \ldots, n\}$, let $C_t(\mathbf{v}_0, i)$ be the set of vertices (i, y) included in the cluster at time t. (This is the collection of all the elements of the i-th horizontal line added by time t during the exploration procedure.) Let also $C(\mathbf{v}_0, i)$ be the set of all vertices (i, y) in $C(\mathbf{v}_0)$, that is, the collection of all the elements in the i-th horizontal line contained in $C(\mathbf{v}_0)$. We further denote the number of elements of the i-th horizontal line included in $C(\mathbf{v}_0)$ until time T by $N(\mathbf{v}_0, i) = |C_T(\mathbf{v}_0, i)|$.

Similarly, let $\hat{C}_t(\mathbf{v}_0, i)$ be the set of vertices (x, i) included in the cluster at time t, that is all the i-th vertical line elements added by time t during the exploration procedure.) Let $\hat{C}(\mathbf{v}_0, i)$ be the set of all vertices (x, i) in $C(\mathbf{v}_0)$; and, finally, denote the number of elements of the i-th vertical line included in $C(\mathbf{v}_0)$ until time T by $\hat{N}(\mathbf{v}_0, i) = |\hat{C}_T(\mathbf{v}_0, i)|$.

We write (x_t, y_t) for the vertex that is explored at time t if such a vertex exists, that is, if $t \leq T$. We may identify the set of colours with the set $\{0, 1, 2\}$. The state of the exploration process at time t is the list giving the colour of each vertex, in other words, an n-vector with values in $\{0, 1, 2\}^n$. This process defines a natural filtration $\varphi_0 \subseteq \varphi_1 \subseteq \ldots \subseteq \varphi_T$, where φ_t is the smallest σ-field with respect to which the state at time t is measurable. (Informally, φ_t corresponds to "everything that has occurred until time t".) We note that T is a stopping time with respect to this filtration. We note also that, even on the event $\{T = \lceil \eta V \rceil\}$, it is not necessarily the case that $C_T(\mathbf{v}_0) = \lceil \eta V \rceil$, since the number of new vertices added at each exploration step is a random variable, which can be smaller or greater than 1. We stop our process at time T, and we make the convention that $C_t(\mathbf{v}_0) = C_T(\mathbf{v}_0)$ for all $t \geq T$ (and similarly for all other relevant random variables). This is important when $T = T_{\mathbf{v}_0} < \eta V$, that is, when the process dies out before time ηV.

Following the notation of Section 3, we let F denote the total population size of a Galton-Watson process starting with one individual, where the offspring distribution is $\mathrm{Bi}(\Omega, p)$; and further $\mathbb{P}_{\Omega,p}$ denotes the probability measure corresponding to this branching process. Proposition 2.1 involves upper and lower bounds on $\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell)$ for appropriate choices of ℓ. These bounds are formulated in Lemmas 4.1– 4.3 below.

Lemma 4.1 (Stochastic domination of cluster size by branching process progeny size)**.** *For every* $\ell \in \mathbb{N}$,

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell) \leq \mathbb{P}_{\Omega,p}(F \geq \ell).$$

In particular, for every N_v *such that* $\varepsilon^{-2} \ll N_v \leq \eta V$, *with* $\eta \ll \varepsilon$,

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \eta V) \leq \mathbb{P}_p(|C(\mathbf{v}_0)| \geq N_v) \leq \mathbb{P}_{\Omega,p}(F \geq N_v). \tag{4.2}$$

The proof of Lemma 4.1 can be found in Section 5, and is based on a coupling between our component exploration procedure described above and the $\mathrm{Bi}(\Omega, p)$ Galton-Watson process. This coupling implies a stochastic comparison between $|C(\mathbf{v}_0)|$ and F.

Lemma 4.2 below follows directly from Lemma 4.1 and establishes the upper bound part of Proposition 2.1. It is also used in the proof of Lemma 7.2 in Section 7.

Lemma 4.2. *For every* $\ell \in \mathbb{N}$, *and for* $\varepsilon \geq V^{-1/3}$,

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell) \leq 2\varepsilon + O(\varepsilon^2) + O\Big(\frac{1}{\sqrt{\ell}}\Big).$$

In particular, if $\varepsilon^{-2} \ll N_v \leq \eta V$, *where* $\eta \ll \varepsilon$, *then*

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq N_v) \leq 2\varepsilon(1 + o(1)). \tag{4.3}$$

Proof. By Lemma 4.1, for every $\ell \in \mathbb{N}$,

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell) \leq \mathbb{P}_{\Omega,p}(F \geq \ell).$$

Our choice of $p = p(n)$ implies that $\Omega p = 1 + \varepsilon > 1$, that is, the $\mathrm{Bi}(\Omega, p)$ Galton-Watson process is supercritical. By Proposition 3.3,

$$\mathbb{P}_{\Omega,p}(F \geq N_v) = 2\varepsilon + O(\varepsilon^2) + O\Big(\frac{1}{\sqrt{\ell}}\Big).$$

But $N_v \gg \varepsilon^{-2}$, and so $1/\sqrt{N_v} = o(\varepsilon)$, which completes the proof. $\qquad\square$

Our next lemma establishes a lower bound on $\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell)$, that is, the lower bound part of Proposition 2.1.

Lemma 4.3 (Stochastic domination of cluster size over branching process progeny size)**.** *For every* $\ell \ll \varepsilon V$,

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell) \geq \mathbb{P}_{\Omega',p}(F \geq \ell) + O(V^{-3}),$$

where $\Omega' = \Omega - \frac{5}{2}\max\{\ell n^{-1}, C\log n\}$.
Consequently, if $\varepsilon \gg V^{-1/3}$ *and* $\varepsilon^{-2} \ll N_v \leq \eta V$, *with* $\eta \ll \varepsilon$, *then*

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq N_v) \geq 2\varepsilon(1 + o(1)). \tag{4.4}$$

Lemma 4.3 is proved in Section 6, where we show that the cluster size stochastically dominates the $\mathrm{Bi}(\Omega', p)$ Galton-Watson process.

5. Upper bounds on the cluster size and structure

In this section we establish upper bounds on some quantities related to the cluster size and structure.

5.1. **Upper bound on the cluster tail.** This subsection contains the proof of Lemma 4.1. Recall from (4.1) that $T = T_{\mathbf{v}_0} \wedge \lceil \eta V \rceil$. Since $\mathbb{P}_p(T = \lceil \eta V \rceil) = \mathbb{P}_p(|C(\mathbf{v}_0)| \geq \eta V)$, this lemma yields an upper bound on $\mathbb{P}_p(T = \lceil \eta V \rceil)$, the probability that the exploration process of $\mathbf{v}_0$ does not die out before the time ηV. Our proof makes use of a classical coupling, which has appeared in the literature before in different contexts and under different guises.

Proof of Lemma 4.1. We couple our exploration procedure described in Section 4 with a Galton-Watson branching process such that the family size is $\mathrm{Bi}(\Omega, p)$, and (with probability 1) the population is derived from a single individual. Indeed, it is clear that at each time $t \leq T$ the number of vertices added to $C_t(\mathbf{v}_0)$ as a result of exploring (x_t, y_t), given the σ-field φ_t, is stochastically dominated by a $\mathrm{Bi}(2(n-1), p)$ random variable.

The Galton-Watson process in question may be represented as a rooted tree, where each vertex has $\Omega = 2(n-1)$ descendants, and the root corresponds to the single progenitor of the population. Initially, the root is green and all the other vertices are white. We *explore* the Galton-Watson tree one vertex at a time. Here, exploring a vertex means determining the offspring of the corresponding individual. Whenever we add a new descendant, we colour it green. Once we have explored a vertex completely, we colour it red. Then $C_t(\mathbf{v}_0)$ consists precisely of all the red and green vertices.

Let F_t be the number of individuals that have been born until time t, and let F_t' be the number of green individuals at time t in the Galton-Watson tree. Since the Hamming graph $H(2, n)$ is not a tree, at times the exploration and colouring process described in Section 4 will not be considering all the Ω neighbours of a currently explored vertex. This will occur if some of these neighbours have already been considered and thus already coloured green or red due to being adjacent to another vertex explored at an earlier time.

The above discussion implies that, on the event that $|C_t(\mathbf{v}_0)| \leq F_t$ and $|G_t(\mathbf{v}_0)| \leq F_t'$, given the σ-field φ_t (that is, everything else that may have happened until time t), we can couple the two exploration processes for another step so that $|C_{t+1}(\mathbf{v}_0)| \leq F_{t+1}$ and $|G_{t+1}(\mathbf{v}_0)| \leq F_{t+1}'$. By induction, we thus achieve a step-by-step coupling up until the stopping time T. We deduce that the random variable $|C(\mathbf{v}_0)| = \sup_t |C_t(\mathbf{v}_0)|$ is stochastically at most $F = \sup_t F_t$, the total size of the Galton-Watson population. Hence, for each $\ell \in \mathbb{N}$, we have that

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell) \leq \mathbb{P}_{\Omega, p}(F \geq \ell).$$

This holds in particular with $\ell = N_V$, as claimed. ∎

5.2. **Cluster structure.** In this subsection we give an upper bound on the number of elements of a large cluster that belong to a particular horizontal line. The following proposition is crucial in the proofs of Propositions 2.3 and 4.3.

Proposition 5.1 (Upper bound on the number of elements per line in a large cluster). *Let* $\varepsilon = \varepsilon(n) \geq 0$ *be such that* $\varepsilon = \varepsilon(n) \leq 1/20$ *and choose* $\eta \ll \varepsilon$. *Further, let* $N(\mathbf{v_0}, i)$ *be the number of elements of* $C_T(\mathbf{v_0}) = C_{\eta V}(\mathbf{v_0})$ *that belong to the horizontal line* i. *There exists a positive constant* c_1 *such that for every* $\nu > 0$

$$\mathbb{P}_p \Big(\max_{i=1,\ldots,n} : N(\mathbf{v_0}, i) \geq (1+\nu)\frac{11}{9}\eta n \Big) \leq ne^{-c_1 \nu \eta n}. \tag{5.1}$$

Furthermore, there exist constants $c_2, c_3, c_4 > 0$ such that the following holds:

(1) *Let $n \in \mathbb{N}$ and $\eta = \eta(n)$ be such that $\eta n \geq c_2 \log n$. If n is sufficiently large, then*

$$\mathbb{P}_p\Big(\max_{i=1,\ldots,n} : N(\mathbf{v_0}, i) \geq \frac{5}{4}\eta n \Big) \leq c_4 V^{-3}. \tag{5.2}$$

(2) *Let $n \in \mathbb{N}$ and $\eta = \eta(n)$ be such that $\eta n / \log n < c_2$. If n is sufficiently large, then*

$$\mathbb{P}_p\Big(\max_{i=1,\ldots,n} : N(\mathbf{v_0}, i) \geq c_3 \log n \Big) \leq c_4 V^{-3}. \tag{5.3}$$

Here is an informal outline of the proof. Whenever we explore a vertex *not* on the line i, we add an element of line i with probability p. On the other hand, each vertex belonging to the line i has $n-1$ neighbours on that line. Whenever such a vertex is explored, each one of its neighbours on the line i is included with probability p (unless it is already in the cluster). It follows that the number of new elements on line i resulting from exploring a vertex belonging to that line is stochastically dominated by a $\text{Bi}(n-1, p)$ Galton-Watson process. Since $p = \frac{1+\varepsilon}{2(n-1)}$ and $\varepsilon \leq 1/2 < 1$ for n large enough, we have that $(n-1)p < 1$, so that the Galton-Watson process is subcritical. Hence, using standard concentration of measure techniques, we are able to upper bound the number of elements on line i that make it into a large cluster. We now make this argument precise.

Proof of Proposition 5.1. Let $i \in \{1, \ldots, n\}$ and, for each $t = 1, \ldots, T$, let $S_t(i)$ be the number of times s such that (x_{s-1}, y_{s-1}), (i, y_{s-1}) is open and $x_{s-1} \neq i$. That is, for each time $t \leq T$, $S_t(i)$ is the number of times we enter the horizontal line i until time t. We can write

$$S_t(i) = \sum_{s=1}^{t} Y_s(i),$$

where $Y_t(i)$ is the indicator of the event that the edge between (x_{t-1}, y_{t-1}), (i, y_{t-1}) is open, and $x_{t-1} \neq i$. For each time t,

$$\mathbb{P}_p(Y_t(i) = 1 | \varphi_{t-1}) \leq p,$$

and so known results (see for instance Lemma 2.2 in [19]) imply that $S_t(i)$ is stochastically dominated by a $\text{Bi}(t, p)$ random variable. Consequently, for every $u \geq 0$, the moment generating function $M_{S_t(i)}(u)$ is bounded above by $(1 + p(e^u - 1))^t$.

For $r = 1, 2, \ldots$, let $Z_r(i)$ be the number of vertices (i, x) added as a result of the r-th entry on to horizontal line i. Given that vertex $(i, \tilde{x}_0) \in C(\mathbf{v}_0)$, the number of its neighbours (i, x) added to $C(\mathbf{v}_0)$ during its exploration (if it has occurred by the time ηV) is easily seen to be stochastically dominated by a random variable $\text{Bi}(n-1, p)$. Hence, for each r, $Z_r(i)$ is stochastically dominated by the total progeny in a branching process with offspring distribution $\text{Bi}(n-1, p)$ descending from a single individual. Since $p = (1+\varepsilon)/2(n-1)$ and $\varepsilon < 1/2$, this branching process is subcritical. We deduce that, for $u \geq 0$, the moment generating function $M_{Z_r(i)}(u)$ of $Z_r(i)$ is bounded above by the moment generating function $M_Z(u)$ of an integer-valued, finite random variable Z, whose distribution is given by the Otter-Dwass formula (Lemma 3.4). In other words, for each $N \in \mathbb{N}$,

$$\mathbb{P}_p(Z = N) = \frac{\mathbb{P}(\xi_1 + \ldots + \xi_N = N - 1)}{N}, \tag{5.4}$$

where the ξ_r are i.i.d. $\mathrm{Bi}(n-1,p)$. It follows that

$$M_z(u) = \sum_{N=1}^{\infty} \frac{e^{uN}}{N} \mathbb{P}\big(\mathrm{Bi}(N(n-1),p) = N-1\big)$$

$$= \sum_{N=1}^{\infty} \frac{e^{uN}}{N} \binom{N(n-1)}{N-1} p^{N-1}(1-p)^{N(n-1)-(N-1)}.$$

Our aim is to derive an upper bound for the above expression. Unlike the branching processes considered in Section 3, which were (slightly) supercritical, we are now *subcritical*. Recall that the expected total progeny of a $\mathrm{Bi}(m,p)$ Galton-Watson process is $\frac{1}{1-mp}$; using this fact with $m = n-1$ and $p = \frac{1+\varepsilon}{2(n-1)}$, we see that $\mathbb{E}[Z] = \frac{2}{1-\varepsilon}$.

As $N! \geq (N/e)^N$, we have

$$\frac{1}{N}\binom{N(n-1)}{N-1} \leq \frac{[N(n-1)]^{N-1}}{N!} \leq \frac{(n-1)^{N-1}}{N}e^N, \tag{5.5}$$

which in turn implies that

$$M_z(u) \leq \sum_{N=1}^{\infty} \frac{e^{uN}}{N}(n-1)^{N-1}e^N\left(\frac{1+\varepsilon}{2(n-1)}\right)^{N-1}\left(1-\frac{1+\varepsilon}{2(n-1)}\right)^{N(n-1)-(N-1)}.$$

Since $1 - x \leq e^{-x}$, we see that

$$M_z(u) \leq \frac{2}{1+\varepsilon}\sum_{N=1}^{\infty}\frac{1}{N}\left(\frac{e^{u+1}(1+\varepsilon)e^{-(1+\varepsilon)/2}}{2}\right)^N e^{\frac{(1+\varepsilon)(N-1)}{2(n-1)}}, \tag{5.6}$$

which is finite for $0 \leq u < \frac{1+\varepsilon}{2} - 1 - \log\frac{1+\varepsilon}{2}$ and n large. Further, it is easily seen that for such u and n the contribution due to terms with $N > C\log n$ is negligible, provided that C is a sufficiently large constant.

Clearly, $C_t(\mathbf{v}_0, i)$ is always bounded above by $\sum_{r=1}^{S_t(i)} Z_r(i)$. In particular, $\sum_{r=1}^{S_T(i)} Z_r(i)$ is an upper bound on $C_T(\mathbf{v}_0, i)$, which equals the number of vertices (i, x) included in the cluster of $\mathbf{v}_0$ from time 0 to time ηV.

Let $u_0 = \frac{1}{2}(\frac{1+\varepsilon}{2} - 1 - \log\frac{1+\varepsilon}{2})$. Since Z is non-negative, $M_Z(u) \geq 1$. It follows from the above (using also the bound $1 + x \leq e^x$ for $x \geq 0$) that for $0 \leq u \leq u_0$,

$$M_{N(\mathbf{v},i)}(u) \leq M_{S_{\eta V}(i)}(\log M_Z(u)) \leq (1 + p(M_z(u) - 1))^{\lceil \eta V\rceil} \tag{5.7}$$

$$\leq \exp\left(\eta V p(M_z(u) - 1)\right)$$

$$= (1 + o(1))\exp\left(\frac{1}{2}\eta n(1+\varepsilon)(M_z(u) - 1)\right)$$

$$\leq (1 + o(1))\exp\left(\left[\eta n\frac{1+\varepsilon}{1-\varepsilon}u + \frac{1}{2}\eta n(1+\varepsilon)u^2 M_z''(u_0)\right]\right),$$

where the final inequality comes from a second order Taylor expansion. Also we have used the fact that $\mathbb{E}[Z] = 2/(1-\varepsilon)$, and that the second derivative $M_z''(u)$ is increasing in u for $u \leq u_0$.

Now we run a standard large deviations argument. For all k,

$$\mathbb{P}_p(N(\mathbf{v}, i) \geq k) \leq \frac{M_{N(\mathbf{v}, i)}(u)}{e^{uk}}$$

$$\leq (1 + o(1)) \exp\left(\left(\eta n \frac{1 + \varepsilon}{1 - \varepsilon} - k\right)u + \frac{1}{2}\eta n(1 + \varepsilon)u^2 M_z''(u_0)\right). \qquad (5.8)$$

The expression in (5.8) can be optimised with respect to u in the usual way, and one finds that there exists a constant $c_1 > 0$ such that for all $\nu > 0$,

$$\mathbb{P}_p\left(N(\mathbf{v}, i) \geq (1 + \nu)\eta n \frac{1 + \varepsilon}{1 - \varepsilon}\right) \leq e^{-c_1 \nu \eta n},$$

which yields

$$\mathbb{P}_p\left(\max_{i=1,\ldots,n} N(\mathbf{v}, i) \geq (1 + \nu)\eta n \frac{1 + \varepsilon}{1 - \varepsilon}\right) \leq n e^{-c_1 \nu \eta n}.$$

Since $\varepsilon \leq 1/20$, we have $(1 + \varepsilon)/(1 - \varepsilon) \leq 11/9$; and therefore there exists a constant $\tilde{c}_1$ such that, for $\nu > 0$,

$$\mathbb{P}_p\left(\max_{i=1,\ldots,n} N(\mathbf{v}, i) \geq \frac{11}{9}(1 + \nu)\eta n\right) \leq n e^{-2\tilde{c}_1 \nu \eta n}, \qquad (5.9)$$

which proves the first statement of Proposition 5.1.

For the remainder, first suppose that $\eta n / \log n \geq 176/\tilde{c}_1$; we may take $\nu = 1/44$ in (5.9) to deduce that

$$\mathbb{P}_p\left(\max_{i=1,\ldots,n} N(\mathbf{v}, i) \geq \frac{5}{4}\eta n\right) \leq n^{-6} = V^{-3}.$$

Now assume that $\eta n / \log n < 176/\tilde{c}_1$. Note that $N(\mathbf{v}, i)$ is stochastically dominated by $\sum_{r=1}^{\tilde{S}} Z_r(i)$, where $\tilde{S}$ is the maximum of a $\text{Bi}(\lceil \eta V \rceil, p)$ and a $\text{Bi}(\lceil 176\tilde{c}_1^{-1} n \log n \rceil, p)$, and hence is stochastically dominated by a $\text{Bi}(\lceil \eta V \rceil \vee \lceil 176\tilde{c}_1^{-1} n \log n \rceil, p)$. Since $\eta n / \log n < 176/\tilde{c}_1$, $\lceil \eta V \rceil \geq \lceil 176\tilde{c}_1^{-1} n \log n \rceil$ so that, asymptotically, $\tilde{S}$ is at most $\text{Bi}(\lceil 176\tilde{c}_1^{-1} n \log n \rceil, p)$. We can perform the moment generating function and large deviations calculations as in (5.6)– (5.8) above, to find that

$$\mathbb{P}_p\left(\max_{i=1,\ldots,n} N(\mathbf{v}, i) \geq \frac{220}{\tilde{c}_1} \log n\right) \leq n e^{-8 \log n(1 + o(1))} = o(V^{-3}).$$

Taking $c_2 = 176/\tilde{c}_1$, $c_3 = 220/\tilde{c}_1$ and $c_4 = 1$ completes the proof of Proposition 5.1. ∎

6. Lower bounds on the cluster size and structure

In this section we establish corresponding lower bounds on the cluster size and structure. We first give a proof of Lemma 4.3, which will in particular establish a lower bound on $\mathbb{P}_p(T = \lceil \eta V \rceil)$, that is, a lower bound on $\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \eta V)$. Our argument will rely on a coupling with a suitable lower bounding Galton-Watson process and the estimates of Proposition 5.1.

Proof of Lemma 4.3. As in Sections 4 and 5, $C_t(\mathbf{v}_0, i)$ denotes the set of vertices (i, y) included in the exploration of cluster $C(\mathbf{v}_0)$ until time t; and $\hat{C}_t(\mathbf{v}_0, i)$ denotes the set of vertices (x, i) added to $C(\mathbf{v}_0)$ until time t.

Let $\tilde{c}_1$ be as in Proposition 5.1. Let $m = \frac{5}{4} \max\{\eta n, \frac{176}{\tilde{c}_1} \log n\}$. For each time t, let $\mathcal{E}_t$ be the event that $|C_t(\mathbf{v}_0, i)| \leq m$ and $|\hat{C}_t(\mathbf{v}_0, i)| \leq m$ for all $i = 1, \ldots, n$.

Define $\Omega' = \Omega - 2m = 2(n - 1 - m)$. Then, provided that the event $\mathcal{E}_t$ occurs, conditionally on φ_t (that is, given everything else that may have happened until time t), the number of vertices added to $C_t(\mathbf{v}_0)$ as a result of exploring (x_t, y_t) stochastically dominates a $\mathrm{Bi}(\Omega', p)$ random variable. We note that $\Omega - \Omega' = O(\eta n + \log n) = o(n)$, since $\eta \to 0$ as $n \to \infty$.

We shall couple our exploration process with a Galton-Watson process starting with a single individual, where the offspring distribution is $\mathrm{Bi}(\Omega', p)$. The mean offspring size for this Galton-Watson process is

$$\Omega' p := 1 + \varepsilon' = \left(1 - \frac{m}{n-1}\right)(1 + \varepsilon) = 1 + \varepsilon - O(\eta + n^{-1}\log n) = 1 + \varepsilon(1 + o(1)),$$

where we have used the fact that $\eta \ll \varepsilon$ and $n^{-2/3} \ll \varepsilon$. By Proposition 3.3, its survival probability is $2\varepsilon + O(\eta + n^{-1}\log n + \varepsilon^2) = 2\varepsilon(1 + o(1))$.

We now explore the $\mathrm{Bi}(\Omega', p)$ Galton-Watson tree in the same way that we explored the $\mathrm{Bi}(\Omega, p)$ process tree in the proof of Lemma 4.1. Let F_t be the population size and let F_t' be the set of *green* or *active* individuals at time t. Also, let $F = \sup_t F_t$ be the total population size. By the above, on the event $\mathcal{E}_t$ intersected with the event that $|C_t(\mathbf{v}_0)| \geq F_t$ and $|G_t(\mathbf{v}_0)| \geq F_t'$, given φ_t (that is, given everything else that happened until time t), we can couple the Galton-Watson process with the cluster exploration processes for another step so that $|C_{t+1}(\mathbf{v}_0)| \geq F_{t+1}$ and $|G_{t+1}(\mathbf{v}_0)| \geq F_{t+1}'$.

It follows by induction that for each t the random variable $|C_t(\mathbf{v}_0)|I[\mathcal{E}_t]$ is stochastically at least $F_t I[\mathcal{E}_t]$. Hence, for each k,

$$\mathbb{P}_p(|C_t(\mathbf{v}_0)| \geq k) \geq \mathbb{P}_p\big(\mathcal{E}_t \cap \{|C_t(\mathbf{v}_0)| \geq k\}\big) \geq \mathbb{P}_{\Omega,\Omega',p}\big(\mathcal{E}_t \cap \{F_t \geq k\}\big)$$
$$\geq \mathbb{P}_{\Omega',p}(F_t \geq k) - \mathbb{P}_p(\mathcal{E}_t^c),$$

where $\mathbb{P}_{\Omega,\Omega',p}$ denotes the coupling measure. In the second inequality, we have used the fact that for every pair of events $\mathcal{A}, \mathcal{B}$, we have $\mathbb{P}_p(\mathcal{A} \cap \mathcal{B}) \geq \mathbb{P}_p(\mathcal{A}) - \mathbb{P}_p(\mathcal{B}^c)$.

By Proposition 5.1, $\mathbb{P}_p(\overline{\mathcal{E}}_t) = O(V^{-3})$, and so, for each $t \leq \eta V$, we obtain

$$\mathbb{P}_p(|C_t(\mathbf{v}_0)| \geq k) \geq \mathbb{P}_{\Omega',p}(F_t \geq k) + O(V^{-3}).$$

Similarly, for each time t and non-negative integer k,

$$\mathbb{P}_p(|G_t(\mathbf{v}_0)| \geq k) \geq \mathbb{P}_{\Omega',p}(F_t' \geq k) - \mathbb{P}_p(\mathcal{E}_t^c),$$

and, in particular, for each $t \leq \eta V$,

$$\mathbb{P}_p(T \geq t) = \mathbb{P}_p(|G_t(\mathbf{v}_0)| \geq 1) \geq \mathbb{P}_{\Omega',p}(F_t' \geq 1) + O(V^{-3}).$$

Notice that, for $t \leq \eta V$,

$$\mathbb{P}_{\Omega',p}(F_t' = 0) \leq \mathbb{P}_{\Omega',p}(F_{\eta V}' = 0) \leq \mathbb{P}_{\Omega',p}(F < \infty).$$

In this way we arrive at

$$\mathbb{P}_p(|C_t(\mathbf{v}_0)| \geq k) \geq \mathbb{P}_{\Omega',p}(F_t \geq k) + O(V^{-3})$$
$$\geq \mathbb{P}_{\Omega',p}(F_t \geq k, F_t' > 0) + O(V^{-3})$$
$$= \mathbb{P}_{\Omega',p}(F_t' > 0) - \mathbb{P}_{\Omega',p}(F_t' > 0, F_t < k) + O(V^{-3})$$
$$\geq \mathbb{P}_{\Omega',p}(F_t' > 0) - \mathbb{P}_p(\mathrm{Bi}(t\Omega', p) < k) + O(V^{-3})$$
$$\geq \mathbb{P}_{\Omega',p}(F = \infty) - \mathbb{P}_p(\mathrm{Bi}(t\Omega', p) < k) + O(V^{-3}),$$

since on the event that the process is alive at time t and the event $\mathcal{E}_t$ occurs, we can couple the number of vertices added at all steps until t so that it is at least as large as a sum of t independent binomials $\mathrm{Bi}(\Omega', p)$.

Hence, for every constant $\delta \in (0,1)$,

$$\mathbb{P}_p\Big(|C(\mathbf{v}_0)| \geq (1-\delta)\eta V\Big) \geq \mathbb{P}_p(|C_{\eta V}(\mathbf{v}_0)| \geq (1-\delta)\eta V) \tag{6.1}$$

$$\geq \mathbb{P}_{\Omega',p}(F = \infty) - \mathbb{P}_p\big(\mathrm{Bi}(\eta V\Omega', p) < (1-\delta)\eta V\big) + O(V^{-3})$$

$$= 2\varepsilon + O(\eta + n^{-1}\log n + \varepsilon^2) + e^{-\Omega(\eta V)} + O(V^{-3}).$$

But equally, we could run the exploration process until time $(1+\delta)\eta V$ to obtain a cluster of size ηV **whp**, that is, we could use the above with η replaced by $\eta/(1-\delta)$ to obtain that

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \eta V) \geq 2\varepsilon + O(\eta + n^{-1}\log n + \varepsilon^2) + e^{-\Omega(\eta V)} + O(V^{-3}). \tag{6.2}$$

This establishes Proposition 4.3, and hence also completes the proof of Proposition 2.1. ∎

Let us call a horizontal line *good* if it contains at least $\eta V/(4n) = \eta n/4$ elements in $C(\mathbf{v}_0)$ along that line, and *bad* otherwise. We shall now prove Proposition 2.3, thus establishing a lower bound on the number of good lines.

Proof of Proposition 2.3. As earlier, for a vertex $\mathbf{v}_0$ and $i \in \{1, \ldots, n\}$, the random variable $C_t(\mathbf{v}_0, i)$ denotes the number of elements of the i-th horizontal line contained in $C_t(\mathbf{v}_0)$, the part of $C(\mathbf{v}_0)$ obtained by running the exploration process until time t. Also, $C(\mathbf{v}_0, i)$ is the number of elements of the i-th horizontal line in $C(\mathbf{v}_0)$ and $N(\mathbf{v}_0, i)$ is the number of such elements included in $C(\mathbf{v}_0)$ until time $\lceil \eta V \rceil$.

Let c_2 be as in Proposition 5.1 and choose $C = c_2$. By Proposition 5.1, statement (1), we have

$$\mathbb{P}_p\Big(\max_{\mathbf{v}_0} \max_{i=1,\ldots,n} : N(\mathbf{v}_0, i) \leq \frac{5}{4}\eta n\Big) = O(V^{-3}).$$

We select a vertex $\mathbf{v}_0$. Let $\mathcal{A}_1$ be the event that $\max_{i=1,\ldots,n} : N(\mathbf{v}_0, i) \leq \frac{5}{4}\eta n$, and let $\mathcal{A}_2$ be the event that $|C_{\lceil \eta V \rceil}(\mathbf{v}_0)| \geq \frac{7}{8}\eta V$. Let also $\mathcal{B}$ ("$\mathcal{B}$" for "bad") be the event that fewer than $5n/8$ lines are good for the cluster $C(\mathbf{v}_0)$. We only need to show that

$$\mathbb{P}_p\big(\mathcal{B} \cap \{|C(\mathbf{v}_0)| \geq \eta V\}\big) = o(V^{-1}). \tag{6.3}$$

Indeed, summing over all vertices $\mathbf{v}_0$ we may deduce from (6.3) that **whp** there is no $\mathbf{v}_0$ such that $|C(\mathbf{v}_0)| \geq \eta V$ and fewer than $5n/8$ lines are good for $C(\mathbf{v}_0)$. In order to establish (6.3), we write

$$\mathbb{P}_p\big(\mathcal{B} \cap \{|C(\mathbf{v}_0)| \geq \eta V\}\big) \leq \mathbb{P}_p\big(\mathcal{B} \cap \mathcal{A}_2\big) + \mathbb{P}_p\big(\mathcal{A}_2^c \cap \{|C(\mathbf{v}_0)| \geq \eta V\}\big), \tag{6.4}$$

and bound the two terms on the right-hand side separately.

First we shall see that

$$\mathbb{P}_p(\mathcal{B} \cap \mathcal{A}_2) \leq \mathbb{P}_p(\mathcal{A}_1^c). \tag{6.5}$$

Clearly, $|C(\mathbf{v}_0, i)| \geq N(\mathbf{v}_0, i)$ for every i. Let us write $g_{\mathbf{v}_0}$ and $b_{\mathbf{v}_0}$ respectively for the number of good and bad lines in $C_{\eta V}(\mathbf{v}_0)$.

On the event $\mathcal{A}_1 \cap \mathcal{A}_2$, the explored cluster $C_{\lceil \eta V \rceil}(\mathbf{v}_0)$ at time ηV contains at most $5\eta n/4$ elements of every good line and at the same time has size at least $7\eta V/8$. Hence, using also that $g_{\mathbf{v}} = n - b_{\mathbf{v}_0}$, on $\mathcal{A}_1 \cap \mathcal{A}_2$,

$$\frac{7}{8}\eta V \leq |C_{\eta V}(\mathbf{v}_0)| \leq \frac{5}{4}\eta n g_{\mathbf{v}_0} + \frac{1}{4}\eta n b_{\mathbf{v}_0} = \eta n g_{\mathbf{v}_0} + \frac{1}{4}\eta V,$$

which gives

$$\frac{5}{8}\eta V \leq \eta n g_{\mathbf{v}_0}$$

and hence

$$g_{\mathbf{v}_0} \geq \frac{5}{8}n.$$

In other words, on $\mathcal{A}_1 \cap \mathcal{A}_2$, the number of good lines is at least $5n/8$, which means that

$$\mathbb{P}_p(\mathcal{B} \cap \mathcal{A}_1 \cap \mathcal{A}_2) = 0, \tag{6.6}$$

and so establishes claim (6.5). Then from Proposition 5.1 we see that

$$\mathbb{P}_p(\mathcal{B} \cap \mathcal{A}_2) \leq \mathbb{P}_p(\mathcal{A}_1^c) = O(V^{-3}), \tag{6.7}$$

as required.

Let us now bound $\mathbb{P}_p\big(\mathcal{A}_2^c \cap \{|C(\mathbf{v}_0)| \geq \eta V\}\big)$; we have

$$\mathbb{P}_p\big(\mathcal{A}_2^c \cap \{|C(\mathbf{v}_0)| \geq \eta V\}\big) \leq \mathbb{P}_p\big(\mathcal{A}_1^c\big) + \mathbb{P}_p\big(\mathcal{A}_2^c \cap \mathcal{A}_1 \cap \{|C(\mathbf{v}_0)| \geq \eta V\}\big). \tag{6.8}$$

The first term is $O(V^{-3})$. With regard to the second term, we note that if the final cluster $C(\mathbf{v}_0)$ has size at least $\frac{7}{8}\eta V$, but has not achieved that size until time ηV, then it must be the case that $G_t(\mathbf{v}_0) > 0$ for all $t \leq \eta V$. Then on $\mathcal{A}_1$, the number of added elements is stochastically bounded from below by a $\mathrm{Bi}(\eta V \Omega', p)$ random variable, where $\Omega' = \Omega - \frac{5}{4}\eta n$. By the above, and by the concentration of measure of the binomial distribution (as in e.g. (3.10)),

$$\mathbb{P}_p\big(\mathcal{A}_2^c \cap \mathcal{A}_1 \cap \{|C(\mathbf{v}_0)| \geq \eta V\}\big) \leq \mathbb{P}_p\big(\mathrm{Bi}(\eta V \Omega', p) < \frac{7}{8}\eta V\big) = o(V^{-1}),$$

This completes the proof of Proposition 2.3, since we can replace η by $\frac{8}{7}\eta$, and hence $\frac{7}{8}\eta V$ by ηV. ∎

7. Concentration of measure for the number of vertices in large clusters

This section contains our proof of Proposition 2.2. First, in Section 7.1 we present estimates of the cluster size that arise from branching processes estimates and comparisons established in Sections 3–6. In Section 7.2, we estimate the variance of some random variables related to $Z_{\geq N}$. Finally, in Section 7.3 we complete our proof of Proposition 2.2.

7.1. **Key ingredients.** As before, for a positive integer N and an edge probability p, $\mathbb{P}_{N,p}$ denotes the probability measure corresponding to the Galton-Watson process where the family size is a $\mathrm{Bi}(N,p)$ random variable; also, F is the total progeny.

In this section, we prove Proposition 2.2. First let us outline the strategy of the proof. The goal is to establish concentration of measure for $Z_{\geq \ell}$. This will be carried out by second moment methods, in a slightly unusual way.

For every ℓ, let

$$\bar{Z}_{\geq \ell} = Z_{\geq \ell} - \mathbb{E}_p[Z_{\geq \ell}]. \tag{7.1}$$

We need two positive numbers N and N_V satisfying certain conditions stated below, and such that I defined by

$$N_V = N2^I \tag{7.2}$$

is an integer. We aim to prove that

$$\mathbb{P}_p\Big(|\bar{Z}_{\geq N_V}| \geq \delta\varepsilon V\Big) = o(1). \tag{7.3}$$

We further require a sequence $\{\delta_i\}_{i=1}^{I-1}$ such that each $\delta_i > 0$ and $\sum_{i=1}^{I-1} \delta_i \leq \frac{\delta}{2}$. We can write

$$|\bar{Z}_{\geq N_V}| \leq |\bar{Z}_{\geq 2N}| + \sum_{i=1}^{I-1} |\bar{Z}_{\geq 2^{i+1}N} - \bar{Z}_{\geq 2^i N}|,$$

and so, if $|\bar{Z}_{\geq N_V}| \geq \delta\varepsilon V$, then either $|\bar{Z}_{\geq 2N}| \geq \delta\varepsilon V/2$, or $|\bar{Z}_{\geq 2^{i+1}N} - \bar{Z}_{\geq 2^i N}| \geq \delta_i\varepsilon V$ for some $1 \leq i \leq I-1$. Consequently,

$$\mathbb{P}_p\Big(|\bar{Z}_{\geq N_V}| \geq \delta\varepsilon V\Big) \leq \mathbb{P}_p\big(|\bar{Z}_{\geq 2N}| \geq \delta\varepsilon V/2\big) + \sum_{i=1}^{I-1} \mathbb{P}_p\Big(|\bar{Z}_{\geq 2^{i+1}N} - \bar{Z}_{\geq 2^i N}| \geq \delta_i\varepsilon V\Big), \tag{7.4}$$

and we are going to upper bound each term on the right hand side separately. Our argument relies on estimating the variance of $Z_{\geq N}$ and of the differences $Z_{\geq 2^{i+1}N} - Z_{\geq 2^i N}$. This is accomplished in Section 7.2 – see Lemmas 7.2 and Lemma 7.3. The variance estimates impose various restrictions on N and N_V; in Section 7.3 we show that these are satisfied as long as $\varepsilon^3 V \gg \log n$, which establishes Proposition 2.2. The key to the proof is to choose N, N_V and $\{\delta_i\}_{i=1}^{I-1}$ so as to ensure adequate concentration of measure.

The remainder of this subsection is devoted to establishing a bound on the cluster tail crucial to the arguments in Sections 7.2 and Section 7.3. Recall that $\Omega = 2(n-1)$ and suppose that $\Omega' = \Omega'(n)$ satisfies $\Omega - \Omega' = O(n/\log n)$. Suppose further that $\varepsilon = \varepsilon(n) = \Omega p - 1 \to 0$ such that $V^{-1/3}(\log V)^{1/3} \ll \varepsilon \ll 1$ for $V = n^2$ sufficiently large. Then, by Proposition 3.1, for all $\ell \in \mathbb{N}$, uniformly in n,

$$|\mathbb{P}_{\Omega,p}(F \geq \ell) - \mathbb{P}_{\Omega',p}(F \geq \ell)| \leq C\Big(p|\Omega - \Omega'| + \frac{1}{n\ell^{1/2}} + \frac{1}{\ell^3}\Big). \tag{7.5}$$

We shall use inequality (7.5) in the following lemma to identify the cluster tail distribution more precisely.

Lemma 7.1 (Bound on the cluster tail). *Set $p = p_c + \frac{\varepsilon}{\Omega}$. Let $V^{-1/3}(\log V)^{1/3} \ll \varepsilon \ll 1$, and let $\ell \in \mathbb{N}$ satisfy $\ell \leq V^{2/3}$ and $\ell \leq \varepsilon V/\log n$. Then there exists a constant C such that, for n sufficiently large,*

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \in [\ell, 2\ell]) \leq \frac{C}{\sqrt{\ell}}. \tag{7.6}$$

Proof. By Proposition 4.1,

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell) \leq \mathbb{P}_{\Omega,p}(F \geq \ell).$$

Further, by Lemma 4.3, C can be chosen large enough that

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq 2\ell) \geq \mathbb{P}_{\Omega',p}(F \geq 2\ell) + O(V^{-3}),$$

where $\Omega' = \Omega - \frac{5}{2}\max\{\ell n^{-1}, C\log n\}$. Let $\varepsilon' = \Omega' p - 1$ and note that $|\varepsilon - \varepsilon'| \leq \frac{C}{n^2}(\ell + n\log n)$ (after a suitable adjustment of C). Since $\varepsilon \geq V^{-1/3} = n^{-2/3}$, our assumptions on ℓ imply that $|\varepsilon - \varepsilon'| \leq \varepsilon/\log n$. By Propositions 3.1 and 3.2,

$$\mathbb{P}_{\Omega',p}(F \geq 2\ell) \geq \mathbb{P}_{\Omega,p}(F \geq 2\ell) + O\big(|\varepsilon - \tilde{\varepsilon}| + \frac{1}{n\ell^{1/2}} + \frac{1}{\ell^3}\big)$$

$$\geq \mathbb{P}_{\Omega,p}(F \geq \ell) + O(|\varepsilon - \varepsilon'|) + O(\ell^{-1/2}).$$

It follows that

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell) - \mathbb{P}_p(|C(\mathbf{v}_0)| \geq 2\ell) \leq O(|\varepsilon - \varepsilon'|) + O(\ell^{-1/2}),$$

and we only need to show that $|\varepsilon - \varepsilon'|$ is $O(l^{-1/2})$. This is equivalent to showing that both $\frac{\ell}{n^2}$ and $\frac{\log n}{n}$ are $O(l^{-1/2})$. The condition $\ell \leq V^{2/3}$ is equivalent to $\frac{\ell}{n^2} \leq \ell^{-1/2}$. The bound $\frac{\log n}{n} \leq \ell^{-1/2}$ holds when $\ell \leq n^2/(\log n)^2$; as $\ell \leq V^{2/3} = n^{4/3}$, this is also true for n sufficiently large. ∎

7.2. Variance estimates.

7.2. **Variance estimates.** Lemmas 7.2 and 7.3 below contain variance estimates essential to our proof of Proposition 2.2.

Lemma 7.2 (Variance of the number of vertices in moderate clusters). *Set $p = p_c + \frac{\varepsilon}{\Omega}$. Suppose that $V^{-1/3}(\log V)^{1/3} \ll \varepsilon \ll 1$. Choose N such that $N = o(\sqrt{V})$ and $N = o(\varepsilon^2 V)$. Then*

$$\mathrm{Var}_p(Z_{\geq N}) = o\big((\varepsilon V)^2\big). \tag{7.7}$$

Proof. First note that $\mathrm{Var}_p(Z_{\geq N}) = \mathrm{Var}_p(Z_{<N})$, where

$$Z_{<N} = N - Z_{\geq N} = \sum_{\mathbf{v}} I[|C(\mathbf{v})| < N].$$

We expand $\mathrm{Var}_p(Z_{<N})$ as

$$\mathrm{Var}_p(Z_{<N}) = \sum_{\mathbf{v}_0, \mathbf{v}_1} \big[\mathbb{P}_p\big(|C(\mathbf{v}_0)| < N, |C(\mathbf{v}_1)| < N\big) - \mathbb{P}_p(|C(\mathbf{v}_0)| < N)^2 \big]. \tag{7.8}$$

We separate each term involving distinct $\mathbf{v}_0$ and $\mathbf{v}_1$ into two, according to whether or not $\mathbf{v}_1 \in C(\mathbf{v}_0)$. We can then write

$$\mathrm{Var}_p(Z_{<N}) = S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} + S_{\mathbf{v}_0 \nleftrightarrow \mathbf{v}_1}, \tag{7.9}$$

where $S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1}(N)$, $S_{\mathbf{v}_0 \nleftrightarrow \mathbf{v}_1} = S_{\mathbf{v}_0 \nleftrightarrow \mathbf{v}_1}(N)$ and

$$S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = \sum_{\mathbf{v}_0, \mathbf{v}_1} \mathbb{P}_p(|C(\mathbf{v}_0)| < N, \mathbf{v}_1 \in C(\mathbf{v}_0)), \tag{7.10}$$

$$S_{\mathbf{v}_0 \nleftrightarrow \mathbf{v}_1} = \sum_{\mathbf{v}_0, \mathbf{v}_1} \Big[\mathbb{P}_p(|C(\mathbf{v}_0)| < N, |C(\mathbf{v}_1)| < N, \mathbf{v}_1 \notin C(\mathbf{v}_0)) - \mathbb{P}_p(|C(\mathbf{v}_0)| < N)^2. \Big] \tag{7.11}$$

It is easily seen that

$$S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = V \mathbb{E}_p\big[|C(\mathbf{v}_0)| I[|C(\mathbf{v}_0)| < N]\big], \tag{7.12}$$

and we upper bound

$$\mathbb{E}_p\big[|C(\mathbf{v}_0)| I[|C(\mathbf{v}_0)| < N]\big] = \sum_{l=1}^{N} \mathbb{P}_p(l \leq |C(\mathbf{v}_0)| < N) \leq \sum_{l=1}^{N} \mathbb{P}_p(|C(\mathbf{v}_0)| \geq l)$$

$$\leq C \sum_{l=1}^{N} \big(\varepsilon + \frac{1}{\sqrt{l}}\big), \tag{7.13}$$

where the last inequality follows from Lemma 4.2. It follows that

$$S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = O(V N \varepsilon + V \sqrt{N}) = o(\varepsilon^2 V^2), \tag{7.14}$$

provided $N = o(\varepsilon V)$ and $N = o(\varepsilon^4 V^2)$. When $\varepsilon^3 V \gg 1$ then $\varepsilon V \ll \varepsilon^4 V^2$, so only the first constraint on N is binding, i.e. $S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = o(\varepsilon^2 V^2)$ as long as $N = o(\varepsilon V)$.

To upper bound $S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1}$ note that, by [4, inequality (9.7)],

$$S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1} \leq p \sum_{\{\mathbf{u},\mathbf{v}\}} \mathbb{E}_p \Big[|C(\mathbf{u})||C(\mathbf{v})| I[|C(\mathbf{u})| < N, |C(\mathbf{v})| < N, \mathbf{v} \notin C(\mathbf{u})] \Big], \qquad (7.15)$$

where the summation is over all edges $\{\mathbf{u}, \mathbf{v}\}$ of $H(2, n)$. We can estimate this similarly to $S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1}$ above, and find that

$$S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1} \leq p \sum_{(\mathbf{u},\mathbf{v})} \sum_{l_1,l_2=1}^{N} \mathbb{P}_p \Big(l_1 \leq |C(\mathbf{u})| < N, l_2 \leq |C(\mathbf{v})| < N, \mathbf{v} \notin C(\mathbf{u}) \Big)$$

$$\leq p \sum_{(\mathbf{u},\mathbf{v})} \sum_{l_1,l_2=1}^{N} \mathbb{P}_p \Big(|C(\mathbf{u})| \geq l_1, |C(\mathbf{v})| \geq l_2, \mathbf{v} \notin C(\mathbf{u}) \Big). \qquad (7.16)$$

Since $\mathbf{v} \notin C(\mathbf{u})$, $|C(\mathbf{u})|$ and $|C(\mathbf{v})|$ are each *independently* of one another stochastically dominated by the total progeny of a $\mathrm{Bi}(\Omega, p)$ Galton-Watson process. (To see this in more detail, think of first constructing the cluster of $\mathbf{u}$, and subsequently construct the cluster of $\mathbf{v}$ in the smaller graph with $|C(\mathbf{u})$ removed.) Using Lemma 4.2, we then see that, since Ωp is bounded above as $n \to \infty$,

$$S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1} \leq C\Omega p V \sum_{l_1,l_2=1}^{N} \Big(\varepsilon + \frac{1}{\sqrt{l_1}} \Big) \Big(\varepsilon + \frac{1}{\sqrt{l_2}} \Big)$$

$$\leq CV \big(\varepsilon N + \sqrt{N} \big)^2 \leq O(V \varepsilon^2 N^2 + VN). \qquad (7.17)$$

Thus, as long as $N = o(\sqrt{V})$ and $N = o(\varepsilon^2 V)$,

$$S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1} = o\big((\varepsilon V)^2 \big), \qquad (7.18)$$

which completes the proof. $\blacksquare$

Lemma 7.3 (Variance of the number of vertices in intermediate clusters). *Set $p = p_c + \frac{\varepsilon}{\Omega}$. Assume that $V^{-1/3}(\log V)^{1/3} \ll \varepsilon \ll 1$, and take $N \in \mathbb{N}$ such that $N \leq V^{2/3}$ and $N \leq \varepsilon V / \log n$. Then*

$$\mathrm{Var}_p(Z_{\geq N} - Z_{\geq 2N}) \leq \frac{CV^2}{\sqrt{N}} \Big(\frac{\log n}{n} \bigvee \frac{N}{V} \bigvee \frac{1}{N^3} \Big). \qquad (7.19)$$

Proof. We have

$$\mathrm{Var}_p(Z_{\geq N} - Z_{\geq 2N}) = \sum_{\mathbf{v}_0, \mathbf{v}_1} \mathbb{P}_p(|C(\mathbf{v}_0)| \in (N, 2N], |C(\mathbf{v}_1)| \in (N, 2N]) - \mathbb{P}_p(|C(\mathbf{v}_0)| \in (N, 2N])^2.$$

Once again we split the sum according to whether or not $\mathbf{v}_1 \in C(\mathbf{v}_0)$, obtaining

$$\mathrm{Var}_p(Z_{\geq N} - Z_{\geq 2N}) = S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} + S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1}, \qquad (7.20)$$

where

$$S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = \sum_{\mathbf{v}_0, \mathbf{v}_1} \mathbb{P}_p \Big(|C(\mathbf{v}_0)| \in (N, 2N], \mathbf{v}_1 \in C(\mathbf{v}_0) \Big), \qquad (7.21)$$

$$S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1} = \sum_{\mathbf{v}_0, \mathbf{v}_1} \Big[\mathbb{P}_p \Big(|C(\mathbf{v}_0)|, |C(\mathbf{v}_1)| \in (N, 2N], \mathbf{v}_1 \notin C(\mathbf{v}_0) \Big) - \mathbb{P}_p \Big(|C(\mathbf{v}_0)| \in (N, 2N] \Big)^2 \Big]. \quad (7.22)$$

Just as in the proof of Lemma 7.2,

$$S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = V \mathbb{E}_p \big[|C(\mathbf{v}_0)| I[|C(\mathbf{v}_0)| < N] \big] \leq CV(N\varepsilon + \sqrt{N}). \tag{7.23}$$

But

$$V\sqrt{N} \leq \frac{V^2}{\sqrt{N}} \Big(\frac{\log n}{n} \bigvee \frac{N}{V} \Big), \tag{7.24}$$

and so $S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1}$ is bounded by the right hand side of (7.19).

Dealing with $S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1}$ requires more effort. Define

$$p_{\mathbf{v}_0,\mathbf{v}_1} := \mathbb{P}_p \Big(|C(\mathbf{v}_0)| \in (N, 2N], |C(\mathbf{v}_1)| \in (N, 2N], \mathbf{v}_1 \notin C(\mathbf{v}_0) \Big) - \mathbb{P}_p \big(|C(\mathbf{v}_0)| \in (N, 2N] \big)^2, \tag{7.25}$$

so that

$$S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1} = \sum_{\mathbf{v}_0,\mathbf{v}_1} p_{\mathbf{v}_0,\mathbf{v}_1}.$$

Now rewrite

$$\frac{p_{\mathbf{v}_0,\mathbf{v}_1}}{\mathbb{P}_p(|C(\mathbf{v}_0)| \in (N, 2N])} = \mathbb{P}_p \Big(|C(\mathbf{v}_1)| \in (N, 2N], \mathbf{v}_1 \notin C(\mathbf{v}_0) \big| |C(\mathbf{v}_0)| \in (N, 2N] \Big)$$
$$- \mathbb{P}_p \big(|C(\mathbf{v}_1)| \in (N, 2N] \big).$$

Recall that $N(\mathbf{v}, i)$ is the number of elements in the i-th horizontal line included in the cluster until time ηn^2. The proof of Proposition 5.1 implies that there is some constant $C > 0$ such that **wvhp** every $\mathbf{v}$ such that $|C(\mathbf{v})| \in (N, 2N]$ satisfies

$$N(\mathbf{v}, i) \leq C \Big[\log n \vee \frac{N}{n} \Big]. \tag{7.26}$$

(To see this, think of running the exploration process with stopping time T as in (4.1), where η is defined by $2N = \lceil \eta V \rceil$ (so in particular $\eta \ll \varepsilon$). Since $|C(\mathbf{v})| \in (N, 2N]$, we have $C(\mathbf{v}) = C_T(\mathbf{v})$.)

Letting

$$\Omega' = \Omega - 2C \Big[\log n \vee \frac{N}{n} \Big], \tag{7.27}$$

we can lower bound

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \in (N, 2N]) \geq \mathbb{P}_{\Omega',p}(F \geq N) - \mathbb{P}_{\Omega,p}(F \geq 2N) + O(V^{-3}). \tag{7.28}$$

Further, we can upper bound

$$\mathbb{P}_p(|C(\mathbf{v}_1)| \in (N, 2N], \mathbf{v}_1 \notin C(\mathbf{v}_0) \big| |C(\mathbf{v}_0)| \in (N, 2N])$$
$$= \mathbb{P}_p(|C(\mathbf{v}_1)| \geq N, \mathbf{v}_1 \notin C(\mathbf{v}_0) \big| |C(\mathbf{v}_0)| \in (N, 2N])$$
$$- \mathbb{P}_p(|C(\mathbf{v}_1)| \geq 2N, \mathbf{v}_1 \notin C(\mathbf{v}_0) \big| |C(\mathbf{v}_0)| \in (N, 2N])$$
$$\leq \mathbb{P}_{\Omega,p}(F \geq N) - \mathbb{P}_{\Omega',p}(F \geq 2N) + O(V^{-3}). \tag{7.29}$$

(Once again, to see this, think of first exploring the cluster of $\mathbf{v}_0$ and, after that, the cluster of $\mathbf{v}_1$ in $H(2, n)$ with the cluster of $\mathbf{v}_0$ removed.)

Since $N \leq V^{2/3}$ and $N \leq \varepsilon V / \log n$, we can use Lemma 7.1 to bound $\mathbb{P}_p \big(|C(\mathbf{v}_0)| \in (N, 2N] \big)$ and obtain

$$p_{\mathbf{v}_0,\mathbf{v}_1} \leq \frac{C}{\sqrt{N}} \Big(\mathbb{P}_{\Omega,p}(F \geq N) - \mathbb{P}_{\Omega',p}(F \geq N) + \mathbb{P}_{\Omega,p}(F \geq 2N) - \mathbb{P}_{\Omega',p}(F \geq 2N) \Big) + O(V^{-3}). \tag{7.30}$$

By (7.5), with $\varepsilon = p\Omega - 1$ and $\varepsilon' = p\Omega' - 1$, for every $\ell \in \mathbb{N}$,

$$\mathbb{P}_{\Omega,p}(F \geq \ell) - \mathbb{P}_{\Omega',p}(F \geq \ell) \leq C\Big(|\varepsilon - \varepsilon'| + \frac{1}{n\ell^{1/2}} + \frac{1}{\ell^3}\Big). \tag{7.31}$$

Note that, by (7.27),

$$|\varepsilon - \varepsilon'| = C\Big[\frac{\log n}{n} \bigvee \frac{N}{n^2}\Big], \tag{7.32}$$

so that we always have $\frac{1}{n\ell^{1/2}} = O(|\varepsilon - \varepsilon'|)$.
Consequently (with the value of C adjusted between inequalities), for all vertex pairs $\mathbf{v}_0, \mathbf{v}_1$,

$$p_{\mathbf{v}_0,\mathbf{v}_1} \leq \frac{C}{\sqrt{N}}\Big(|\varepsilon - \varepsilon'| + \frac{1}{N^3}\Big). \tag{7.33}$$

Summing over $\mathbf{v}_0, \mathbf{v}_1$,

$$S_{\mathbf{v}_0 \nleftrightarrow \mathbf{v}_1} = \sum_{\mathbf{v}_0,\mathbf{v}_1} p_{\mathbf{v}_0,\mathbf{v}_1} \leq \frac{CV^2}{\sqrt{N}}\Big(|\varepsilon - \varepsilon'| \vee \frac{1}{N^3}\Big). \tag{7.34}$$

Since $\varepsilon - \varepsilon' = O\Big(\frac{\log n}{n} \vee \frac{N}{n^2}\Big)$, the result follows. ∎

7.3. **Proof of Proposition 2.2.** We are now ready to complete the proof of Proposition 2.2.
We will make essential use of Lemmas 7.2 and 7.3.
Let $\delta > 0$, and, for $i = 1, \ldots, I - 1$, let

$$\delta_i = \frac{\delta}{4\zeta(2)[i \wedge (I - i)]^2}. \tag{7.35}$$

The reasons for our choice for $\{\delta_i\}_{i=1}^{I-1}$ will become apparent shortly. For now let us note that

$$\sum_{i=1}^{I-1} \delta_i \leq \delta/2. \tag{7.36}$$

Recall the definition of $\bar{Z}_{\geq \ell}$ from (7.1) (the centred number of vertices in clusters of size at least ℓ) and the decomposition in (7.4). We will prove that the right hand side of (7.4) is $o(1)$ for suitable N and N_V; the conditions that N and N_V must satisfy are as follows:

$$N = o(n), \qquad N = o(\varepsilon^2 V), \qquad N \gg \frac{(\log n)^2}{n^2\varepsilon^4}, \qquad N \geq \Big(\frac{n}{\log n}\Big)^{1/3}. \tag{7.37}$$

$$N_V \leq \frac{\varepsilon V}{\log n}, \qquad N_V \leq V^{2/3}. \tag{7.38}$$

Further Proposition 2.2 requires that $N_V \gg \varepsilon^{-2}$. We first prove that the concentration bound in Proposition 2.2 holds provided such N and N_V do exist; we then verify that suitable choices of N and N_V can indeed be made so as to meet all the constraints, including the one that $N_V \gg \varepsilon^{-2}$.
By (7.37), N satisfies the hypotheses of Lemma 7.2; hence, using the Chebyshev inequality,

$$\mathbb{P}_p\big(|\bar{Z}_{\geq 2N}| \geq \delta\varepsilon V/2\big) \leq \frac{4\mathrm{Var}_p(Z_{\geq N})}{(\delta\varepsilon V)^2} = o(1). \tag{7.39}$$

Since $N_i \leq N_V$, our assumptions imply that $N_i \leq \frac{\varepsilon V}{\log n}$ and $N_i \leq V^{2/3}$ for each i. Therefore, applying Lemma 7.3 (with $N_i = 2^{i+1}N$) and the Chebyshev inequality, we obtain

$$\mathbb{P}_p\Big(|\bar{Z}_{\geq 2^{i+1}N} - \bar{Z}_{\geq 2^i N}| \geq \delta_i \varepsilon V\Big) \leq (\delta_i \varepsilon V)^{-2} \mathrm{Var}_p\big(Z_{\geq 2^{i+1}N} - Z_{\geq 2^i N}\big)$$

$$\leq (\delta_i \varepsilon V)^{-2}\Big[\frac{CV^2}{\sqrt{N_i}}\Big(\frac{\log n}{n} \bigvee \frac{N_i}{V} \bigvee \frac{1}{N_i^3}\Big)\Big]. \tag{7.40}$$

It follows that under our assumptions

$$\mathbb{P}_p\big(|\bar{Z}_{\geq N_V}| \geq \delta\varepsilon V\big) \leq o(1) + \sum_{i=1}^{I-1} \frac{\frac{CV^2}{\sqrt{N_i}}\Big(\frac{\log n}{n} \bigvee \frac{N_i}{V} \bigvee \frac{1}{N_i^3}\Big)}{(\delta_i \varepsilon V)^2}. \tag{7.41}$$

Each term here is given by

$$\frac{CV^2}{\sqrt{N_i}}\frac{\Big(\frac{\log n}{n} \bigvee \frac{N_i}{V} \bigvee \frac{1}{N_i^3}\Big)}{(\delta_i \varepsilon V)^2} = \frac{C}{\sqrt{N_i}}\frac{\Big(\frac{\log n}{n} \bigvee \frac{N_i}{V} \bigvee \frac{1}{N_i^3}\Big)}{\delta_i^2 \varepsilon^2}. \tag{7.42}$$

By the last assumption in (7.37), for all i, $\frac{\log n}{n} \geq \frac{1}{N^3} \geq \frac{1}{N_i^3}$, so that the last term is never equal to the maximum. It follows that we need to upper bound

$$\mathbb{P}_p\big(|\bar{Z}_{\geq N_V}| \geq \delta\varepsilon V\big) \leq o(1) + \sum_{i=1}^{I-1} \frac{\frac{CV^2}{\sqrt{N_i}}\Big(\frac{\log n}{n} \bigvee \frac{N_i}{V}\Big)}{(\delta_i \varepsilon V)^2}. \tag{7.43}$$

Letting m be the smallest i such that

$$\frac{\log n}{n} \leq \frac{N_i}{V}, \tag{7.44}$$

we can write

$$\sum_{i=1}^{I-1} \frac{\frac{CV^2}{\sqrt{N_i}}\Big(\frac{\log n}{n} \bigvee \frac{N_i}{V}\Big)}{(\delta_i \varepsilon V)^2} = \sum_{i=1}^{m} \frac{C}{\sqrt{N_i}}\frac{\log n}{n\delta_i^2 \varepsilon^2} + \sum_{i=m+1}^{I-1} C\frac{\sqrt{N_i}}{\delta_i^2 \varepsilon^2 V}. \tag{7.45}$$

Using our definition of δ_i in (7.35), we can upper bound $\sum_{i=m+1}^{I-1} \frac{\sqrt{N_i}}{\delta_i^2}$ by

$$\sum_{i=m+1}^{I-1} \frac{\sqrt{N_i}}{\delta_i^2} \leq \frac{16\zeta(2)^2}{\delta^2} 2^{I/2}\sqrt{N} \sum_{i=1}^{I-1} 2^{(i-I)/2}(I-i)^2$$

$$\leq \frac{16\zeta(2)^2}{\delta^2} 2^{I/2}\sqrt{N} \sum_{k=1}^{\infty} k^2 2^{-k/2} \leq C 2^{I/2}\sqrt{N} = C\sqrt{N_V}. \tag{7.46}$$

Hence the second sum in (7.45) is at most

$$\sum_{i=m+1}^{I-1} C\frac{\sqrt{N_i}}{\delta_i^2 \varepsilon^2 V} \leq C\frac{\sqrt{N_V}}{\varepsilon^2 V}. \tag{7.47}$$

We want the right hand side of (7.47) to be $o(1)$, which forces

$$N_V = N_I \ll \varepsilon^4 V^2. \tag{7.48}$$

The bound in (7.48) holds, since $N_V = o(\varepsilon V)$ thanks to the assumption that $\varepsilon^3 V \geq 1$ and the first constraint in (7.38).

On the other hand, the first sum in (7.45) can be upper bounded by

$$\sum_{i=1}^{m} \frac{C}{\sqrt{N_i}} \frac{\log n}{n\delta_i^2 \varepsilon^2} \leq \frac{C\log n}{n\sqrt{N}\varepsilon^2 \delta^2} = o(1), \tag{7.49}$$

since $N \gg \frac{(\log n)^2}{n^2 \varepsilon^4}$ by the third bound in (7.37). This proves the required concentration bound, assuming it is possible to find N and N_v satisfying (7.37) and (7.38).

Finally, we demonstrate that, as long as $\varepsilon^3 V \gg \log n$ and for n sufficiently large, suitable N and N_v *do* exist. One can find a number N satisfying the constraints in (7.37) as long as

$$n \gg \frac{(\log n)^2}{n^2 \varepsilon^4}, \qquad \varepsilon^2 V \gg \frac{(\log n)^2}{n^2 \varepsilon^4}, \qquad n \gg \left(\frac{n}{\log n}\right)^{1/3}, \qquad \varepsilon^2 V \gg \left(\frac{n}{\log n}\right)^{1/3}. \tag{7.50}$$

The third asymptotic bound is obvious. The second constraint is equivalent to $\varepsilon^3 V \gg \log n$, which holds by assumption. This assumption also implies the validity of the first asymptotic relation, which is equivalent to $\varepsilon^4 \gg \frac{(\log n)^2}{n^3}$. Further, the assumption $\varepsilon^3 V \gg \log n$ implies that $\varepsilon^2 V \gg n^{2/3} \gg \left(\frac{n}{\log n}\right)^{1/3}$, and so the fourth condition is true. Hence, an appropriate choice of N is possible.

As for N_v, note that by construction $N_v \geq N$. Therefore, to ensure that the conditions (7.38) can be met we need

$$\frac{\varepsilon V}{\log n} \gg \frac{(\log n)^2}{n^2 \varepsilon^4}, \qquad V^{2/3} \gg \frac{(\log n)^2}{n^2 \varepsilon^4}, \qquad \frac{\varepsilon V}{\log n} \geq \left(\frac{n}{\log n}\right)^{1/3}, \qquad V^{2/3} \geq \left(\frac{n}{\log n}\right)^{1/3}. \tag{7.51}$$

It is easy to see that the first three bounds are all implied by the assumption $\varepsilon \gg n^{-2/3}$; the fourth bound is obvious. Hence one can choose N_v satisfying (7.38). Moreover, since $\varepsilon^3 V \gg \log n$, we have $\varepsilon V / \log n \gg \varepsilon^{-2}$ and $V^{2/3} \gg \varepsilon^{-2}$, and so N_v can also satisfy the requirement $N_v \gg \varepsilon^{-2}$. This completes the proof of Proposition 2.2 and thus of Theorem 1.1. We note that Proposition 2.2 is the key reason why are unable to go all the way down to the critical window, i.e., have to assume that $\varepsilon \gg V^{-1/3}(\log V)^{1/3}$ in Theorem 1.1. It would be of interest to extend the proof all the way down to the critical window, that is to all $\varepsilon \gg V^{-1/3}$.

References

[1] N. Alon and J. Spencer. The Probabilistic Method, 2nd Edition. John Wiley and Sons, New York (2000).

[2] K.B. Athreya and P.E. Ney, Branching Processes, Springer, Berlin, 1972.

[3] A.D. Barbour, L. Holst, S. Janson, Poisson Approximation, OUP, Oxford, 1992.

[4] C. Borgs, J.T. Chayes, R. van der Hofstad, G. Slade and J. Spencer, Random subgraphs of finite graphs: I. The scaling window under the triangle condition. *Random Struct. Alg.* **27** (2005) 137-184.

[5] C. Borgs, J.T. Chayes, R. van der Hofstad, G. Slade and J. Spencer, Random subgraphs of finite graphs: II.The lace expansion and the triangle condition. *Ann. Probab.* **33** (2005) 1886–1944.

[6] C. Borgs, J.T. Chayes, R. van der Hofstad, G. Slade and J. Spencer, Random subgraphs of finite graphs: III. The phase transition on the n-cube. To appear in *Combinatorica*.

[7] C. Borgs, J. T. Chayes, H. Kesten and J. Spencer. Uniform boundedness of critical crossing probabilities implies hyperscaling. *Random Struct. Alg.*, **15**:368–413 (1999).

[8] C. Borgs, J. T. Chayes, H. Kesten and J. Spencer. The birth of the infinite cluster: finite-size scaling in percolation. *Commun. Math. Phys.*, **224**:153–204 (2001).

[9] L.S. Chandran and C.R. Subramanian, A spectral lower bound for the treewidth of a graph and its consequences, preprint, available at www.mpi-sb.mpg.de/ sunil/applyspectree.ps

[10] L. Devroye, Branching Processes and Their Applications in the Analysis of Tree Structures and Tree Algorithms, in Probabilistic Methods for Algorithmic Discrete Mathematics, ed. M.Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed, 249–314, Springer-Verlag, Berlin, 1998.

[11] M. Dwass, The total progeny in a branching process, *J. Appl. Probab.* **6** (1969), 682–686.

[12] M. Heydenreich and R. van der Hofstad. Random graph asymptotics on high-dimensional tori. To appear in *Commun. Math. Phys.*

[13] R. van der Hofstad and G. Slade. Expansion in n^{-1} for percolation critical values on the n-cube and $\mathbb{Z}^n$: the first three terms. *Combin. Probab. Comput.* **15** (2006), 695–713.

[14] S. Janson, Cycles and unicyclic components in random graphs, *Combin. Probab. Comput.* **12** (2003), 27–52.

[15] S. Janson, On concentration of probability, Contemporary Combinatorics, ed. B. Bollobás, Bolyai Soc. Math. Stud. **10** (2002), János Bolyai Mathematical Society, Budapest, 289–301.

[16] S. Janson, D.E. Knuth, T. Łuczak & B. Pittel, The birth of the giant component, *Random Struct. Alg.* **3** (1993), 233–358.

[17] S. Janson, T. Łuczak & A. Ruciński, *Random Graphs*, Wiley, New York, 2000.

[18] V.F. Kolchin, Moments of degeneration of a branching process and height of a random tree, Mathematical Notes of the Academy of Sciences of the USSR **6** (1978), 954–961.

[19] M.J. Luczak, C. McDiarmid and E. Upfal, On-line routing of random calls in networks, *Probab. Theor. Relat. Fields* **125** (2003), 457–482.

[20] C. McDiarmid, Concentration, in Probabilistic Methods for Algorithmic Discrete Mathematics, ed. M.Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed, 195–248, Springer-Verlag, Berlin, 1998.

[21] R. Otter, The multiplicative process, *Ann. Math. Statist.* **20** (1949) 206–224.

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, EINDHOVEN UNIVERSITY OF TECHNOLOGY, 5600 MB EINDHOVEN, THE NETHERLANDS.
E-mail address: rhofstad@win.tue.nl
URL: http://www.win.tue.nl/∼rhofstad

DEPARTMENT OF MATHEMATICS, LONDON SCHOOL OF ECONOMICS, HOUGHTON STREET, LONDON WC2A 2AE, UNITED KINGDOM
E-mail address: m.j.luczak@lse.ac.uk
URL: http://www.lse.ac.uk/people/m.j.luczak@lse.ac.uk/