The diameter of sparse random graphs

Oliver Riordan* Nicholas Wormald†

5th September 2010

Abstract

In this paper we study the diameter of the random graph $G(n,p)$, i.e., the largest finite distance between two vertices, for a wide range of functions $p = p(n)$. For $p = \lambda/n$ with $\lambda > 1$ constant we give a simple proof of an essentially best possible result, with an $O_{\mathrm{p}}(1)$ additive correction term. Using similar techniques, we establish two-point concentration in the case that $np \to \infty$. For $p = (1+\varepsilon)/n$ with $\varepsilon \to 0$, we obtain a corresponding result that applies all the way down to the scaling window of the phase transition, with an $O_{\mathrm{p}}(1/\varepsilon)$ additive correction term whose (appropriately scaled) limiting distribution we describe. Combined with earlier results, our new results complete the determination of the diameter of the random graph $G(n,p)$ to an accuracy of the order of its standard deviation (or better), for all functions $p = p(n)$. Throughout we use branching process methods, rather than the more common approach of separate analysis of the 2-core and the trees attached to it.

1 Introduction and main results

Throughout, we write $\mathrm{diam}(G)$ for the *diameter* of a graph G, meaning the largest graph distance $d(x,y)$ between two vertices x and y in the same component of G:

$$\mathrm{diam}(G) = \max\{d(x,y) : x,y \in V(G),\ d(x,y) < \infty\},$$

where, as usual, $V(G)$ denotes the vertex set of G. In this paper we shall study the diameter of the random graph $G(n,p)$ with vertex set $[n] = \{1,2,\ldots,n\}$, where each possible edge is present with probability $p = p(n)$, independently of the others. For certain functions $p = p(n)$, tight bounds on the diameter of $G(n,p)$ are known; our main aim is to prove such bounds for all remaining functions. In particular, in the special case $p = \lambda/n$ with $\lambda > 1$ constant we shall determine the diameter up to an additive error term that is bounded in probability, where earlier results achieved only $o(\log n)$. A secondary aim is to present a particularly simple

*Mathematical Institute, University of Oxford, 24–29 St Giles', Oxford OX1 3LB, UK. E-mail: riordan@maths.ox.ac.uk.

†Department of Combinatorics and Optimization, University of Waterloo, Waterloo ON, Canada. E-mail: nwormald@uwaterloo.ca. Supported by the Canada Research Chairs Program and NSERC.

proof in this case. All our results apply just as well to $G(n, m)$; in the range of parameters we consider there is essentially no difference between the models. More precisely, although the results for one model do not obviously transfer to the other, the proofs for $G(n, m)$ are essentially the same.

We treat three ranges of $p = p(n)$ separately: $p = \lambda/n$ with $\lambda > 1$ constant, $np \to \infty$ but with an upper bound on the growth rate that extends well into the range covered by classical results, and finally $p = (1 + \varepsilon)/n$ with $\varepsilon(n) \to 0$ but $\varepsilon^3 n \to \infty$. In each case, our analysis investigates the neighbourhoods of vertices, and has three components or phases: 'early growth' — we study the distribution of the number of vertices at distance t from a given vertex v when t is small; 'regular growth' in the middle — we show that the number of vertices at distance t is very likely to grow regularly once the neighbourhoods have become 'moderately large'; 'meeting up' — we show that the distance between two vertices is almost determined by the times their respective neighbourhoods take to become 'large'. This is eventually translated into a result on the diameter.

Our overall plan is made possible by the very accurate information we obtain on the first phase (early growth). The main approach for this is to compare the neighbourhoods of a vertex of $G(n, \lambda/n)$ with the standard Poisson Galton–Watson branching process $\mathfrak{X}_\lambda = (X_t)_{t\geq 0}$; this starts with a single particle in generation 0, and each particle in generation t has a Poisson $\mathrm{Po}(\lambda)$ number of children in the next generation, independently of the other particles and of the history.

A particle in the process $\mathfrak{X}_\lambda$ *survives* if it has descendants in all later generations; the process *survives* if the initial particle survives. If $\lambda > 1$, then the survival probability $s = \mathbb{P}(\forall t : |X_t| > 0)$ is the unique positive solution to

$$1 - s = e^{-\lambda s}. \tag{1.1}$$

Since particles in generation 1 survive independently of each other, the number of such particles that survive has a $\mathrm{Po}(s\lambda)$ distribution, the number that die has a $\mathrm{Po}((1-s)\lambda)$ distribution, and these numbers are independent. It follows that conditioning on the process dying, we obtain again a Poisson Galton–Watson process $\mathfrak{X}_{\lambda_\star} = (X_t^-)_{t\geq 0}$, with the 'dual' parameter

$$\lambda_\star = \lambda(1 - s), \tag{1.2}$$

which may also be characterized as the solution $\lambda_\star < 1$ to

$$\lambda_\star e^{-\lambda_\star} = \lambda e^{-\lambda}. \tag{1.3}$$

This parameter is crucial to understanding the diameter of $G(n, \lambda/n)$. For this and other basic branching process results, see, for example, Athreya and Ney [3].

Throughout the paper we use standard notation for probabilistic asymptotics as in [28]. In particular, $X_n = o_{\mathrm{p}}(f(n))$ means $X_n/f(n)$ converges to 0 in probability, and $X_n = O_{\mathrm{p}}(f(n))$ means $X_n/f(n)$ is bounded in probability.

Our first aim is to give a proof of a tight estimate for the diameter of $G(n, \lambda/n)$ when $\lambda > 1$ is constant as $n \to \infty$ that is simpler than our result for the general case, and also compares favourably with the existing proofs of much weaker bounds for the more general models discussed below.

Theorem 1.1. *Let $\lambda > 1$ be fixed, and let $\lambda_\star < 1$ satisfy $\lambda_\star e^{-\lambda_\star} = \lambda e^{-\lambda}$. Then*

$$\operatorname{diam}(G(n,\lambda/n)) = \frac{\log n}{\log \lambda} + 2\frac{\log n}{\log(1/\lambda_\star)} + O_{\mathrm{p}}(1). \tag{1.4}$$

As usual, we say that an event holds *with high probability*, or *whp*, if its probability tends to 1 as $n \to \infty$. Theorem 1.1 simply says that, for any $K = K(n) \to \infty$, the diameter is whp within K of the sum of the first two terms on the right of (1.4).

The proof of Theorem 1.1 is fairly simple, and will be given in Section 2.

Turning to the case $\lambda = \lambda(n) \to \infty$, we obtain the following result, proved in Section 3 using essentially the same method, although there are various additional complications.

Theorem 1.2. *Let $\lambda = \lambda(n)$ satisfy $\lambda \to \infty$ and $\lambda \le n^{1/1000}$, and let $\lambda_\star < 1$ satisfy $\lambda_\star e^{-\lambda_\star} = \lambda e^{-\lambda}$. Then $\operatorname{diam}(G(n,\lambda/n))$ is two-point concentrated: there exists a function $f(n,\lambda)$ satisfying*

$$f(n,\lambda) = \frac{\log n}{\log \lambda} + 2\frac{\log n}{\log(1/\lambda_\star)} + O(1)$$

such that whp $\operatorname{diam}(G(n,\lambda/n)) \in \{f(n,\lambda), f(n,\lambda)+1\}$. Furthermore, for any $\varepsilon > 0$ and any function λ such that, for large n, neither $\log n/\log(1/\lambda_\star)$ nor $\log n/\log\lambda$ is within ε of an integer, we have

$$\operatorname{diam}(G(n,\lambda/n)) = \left\lceil \frac{\log n}{\log \lambda} \right\rceil + 2\left\lfloor \frac{\log n}{\log(1/\lambda_\star)} \right\rfloor + 1 \tag{1.5}$$

whp.

Bruce Reed has independently announced a related result, in joint work with Nikolaos Fountoulakis; the details are still to appear. We believe that the methods used are quite different.

The main interest of Theorem 1.2 is when λ tends to infinity fairly slowly; if λ grows significantly faster than $\log n$, then the situation is much simpler, and much more precise results are known. Indeed, when $\lambda/(\log n)^3 \to \infty$, Bollobás [6] showed concentration of the diameter on at most two values, and found the asymptotic probability of each value. In the light of this result we would lose nothing by assuming that $\lambda \le (\log n)^4$, say; however, the bound $\lambda \le n^{1/1000}$ turns out to be enough for our arguments.

The bulk of the paper is devoted to the case of expected degree tending to 1, where we prove the following result.

Theorem 1.3. *Let $\varepsilon = \varepsilon(n)$ satisfy $0 < \varepsilon < 1/10$ and $\varepsilon^3 n \to \infty$. Set $\lambda = \lambda(n) = 1 + \varepsilon$, and let $\lambda_\star < 1$ satisfy $\lambda_\star e^{-\lambda_\star} = \lambda e^{-\lambda}$. Then*

$$\operatorname{diam}(G(n,\lambda/n)) = \frac{\log(\varepsilon^3 n)}{\log \lambda} + 2\frac{\log(\varepsilon^3 n)}{\log(1/\lambda_\star)} + O_{\mathrm{p}}(1/\varepsilon). \tag{1.6}$$

Our method in fact gives a description of the limiting distribution of the final correction term (after rescaling); see Theorem 5.1.

A weaker form of Theorem 1.3 has been obtained independently by Ding, Kim, Lubetzky and Peres [19, 20]; see the Remark below.

In the rest of this section we briefly discuss the results above and their relationship to earlier work.

Theorem 1.1 is best possible in the following sense: it is not hard to see that the diameter cannot be concentrated on a set of values with bounded size as $n \to \infty$. Indeed, given any (labelled) graph G with diameter d and at least two isolated vertices, let G' be constructed from G by taking a path P of length d joining two vertices at maximal distance in G, and adding an edge joining each end of P to an isolated vertex. Each graph G' constructed in this way contains a unique pair of vertices at maximal distance $d+2$, and G may be recovered uniquely from G' by deleting the (unique) edges incident with these vertices. Restricting our attention to graphs G with $\Theta(n)$ isolated vertices, the relation (G, G') is thus 1 to $\Theta(n^2)$. Since the probability of G' in the model $G(n,p)$, $p = \lambda/n$, is equal to the probability of G multiplied by $p^2/(1-p)^2 = \Theta(1/n^2)$, it follows easily that for any d we have

$$\mathbb{P}\big(\mathrm{diam}(G(n,\lambda/n)) = d+2\big) \geq \Theta(1)\mathbb{P}\big(\mathrm{diam}(G(n,\lambda/n)) = d\big) - o(1);$$

the $o(1)$ term comes from the possibility that $G(n,\lambda/n)$ has fewer than $\Theta(n)$ isolated vertices. It follows that $\mathrm{diam}(G(n,\lambda/n))$ cannot be concentrated on a finite set of values. In fact, our methods allow us to obtain the limiting distribution of the $O_{\mathrm{p}}(1)$ correction term in (1.4), although this is rather complicated to describe; we return to this briefly in Section 5.

A much weaker form of Theorem 1.1, with a $o(\log n)$ correction term, is a special case of a result of Fernholz and Ramachandran [23] for random graphs with a given degree sequence, and also of a result of Bollobás, Janson and Riordan [11, Section 14.2] for inhomogeneous random graphs with a finite number of vertex types. We shall follow the ideas of [11] to some extent, although the present simpler context allows us to take things much further, obtaining a much more precise result. Earlier, Chung and Lu [16] also studied $\mathrm{diam}(G(n,\lambda/n))$, $\lambda > 1$ constant, but their results were not strong enough to give the correct asymptotic form. Indeed, they conjectured that, under suitable conditions, the diameter is approximately $\log n/\log\lambda$, as one might initially expect.

For the subcritical case, which is much simpler, Łuczak [32] proved very precise results: he showed, for example, that if $\varepsilon \to 0$ and $\varepsilon^3 n \to \infty$, then the subcritical random graph $G = G(n,(1-\varepsilon)/n)$ satisfies

$$\mathrm{diam}(G) = \frac{\log(2\varepsilon^3 n) + O_{\mathrm{p}}(1)}{-\log(1-\varepsilon)}; \tag{1.7}$$

see his Theorem 11(iii), and note that the exponent 2 instead of 3 appearing there is a typographical error. (He also proved a simple formula for the limiting distribution of the $O_{\mathrm{p}}(1)$ term – the probability that it exceeds a constant ρ tends to $1 - \exp(-e^{-\rho})$ as $n \to \infty$; the limiting distribution in the present supercritical case turns out to be much more complicated.) Łuczak's results are effectively the last word on the subcritical case, which we shall not discuss further.

Returning to constant $\lambda > 1$, the lack of concentration on a finite number of points contrasts with the case of random d-regular graphs studied by Bollobás and Fernandez de la Vega [10], who established concentration on a small set of values in this case. Sanwalani and

Wormald [38] have recently shown two-point concentration. (More precisely, they prove one-point concentration for almost all n, and for the remaining n find the probabilities of the two likely values within $o(1)$.) Note that the diameter in this case is simply $\log(n \log n)/\log(d-1) + O_{\mathrm{p}}(1)$ for $d \geq 3$; as we shall see, the behaviour of the two models for this question is very different. Usually, $G(n, \lambda/n)$ is much simpler to study than a random regular graph, but here there are additional complications corresponding to the $2\log n/\log(1/\lambda_\star)$ term in (1.4).

Let us briefly mention a few related results for other random graph models. Perhaps the earliest results in this area are those of Burtin [14, 15] and Bollobás [6]. Turning to results determining the asymptotic diameter when the average degree is constant, one of the first is the result of Bollobás and Fernandez de la Vega [10] for d-regular random graphs mentioned above; another is that of Bollobás and Chung [9], finding the asymptotic diameter of a cycle plus a random matching, which is again logarithmic. Later it was shown by 'small subgraph conditioning' (see [39]) that for such graphs any whp statements are essentially the same as for the uniform model of random 3-regular graphs. The same goes for a variety of other random regular graphs constructed by superposing random regular graphs of various types. For a rather different model, namely a precise version of the Barabási–Albert 'growth with preferential attachment' model, Bollobás and Riordan [12] obtained a (slightly) *sub*logarithmic diameter, contradicting the logarithmic diameter suggested by Barabási, Albert and Jeong [2, 4] (on the basis of computer experiments) and Newman, Strogatz and Watts [35] (on the basis of heuristics).

More recently, related results, often concerning the 'typical' distance between vertices, rather than the diameter, have been proved by many people, for various models. A few examples are the results of Chung and Lu [17, 18], and van den Esker, van der Hofstad, Hooghiemstra, van Mieghem and Znamenski [22, 25, 26]; for a discussion of related work see [25], for example.

The formula (1.4) is easy to understand intuitively: typically, the size of the d-neighbourhood of a vertex (the set of vertices at distance d) grows by a factor of λ at each step (i.e., as d is increased by one). Starting from two typical vertices, taking $\log(\sqrt{n})/\log\lambda$ steps from each, the neighbourhoods reach size about $\sqrt{n}$; at around this point the neighbourhoods are likely to overlap, so the typical distance between vertices is $\log n/\log\lambda$. The second term in (1.4) comes from exceptional vertices whose neighbourhoods take some time to start expanding, or, equivalently, from the few very longest trees attached to (typical vertices of) the *2-core* of $G(n, \lambda/n)$, the maximal subgraph with no vertices of degree 0 or 1. It is well known that the trees hanging off the 2-core of $G(n, \lambda/n)$ have roughly the distribution of the branching process $\mathfrak{X}_{\lambda_\star}$; hence, some of these trees will have height roughly $\log n/\log(1/\lambda_\star)$, and it turns out that the diameter arises by considering two trees of (almost) maximal height attached to vertices in the 2-core at (almost) typical distance.

Although we shall use the 2-core viewpoint later, its use has an intrinsic difficulty caused by the significant variation in the distances between vertices in the 2-core. One can view the variation in the distance between two random vertices of $G = G(n, \lambda/n)$ as coming from three sources: (i) variation in the distances to the 2-core, (ii) variation in the times the neighbourhoods in the 2-core take to start expanding, and (iii) variation in the time the neighbourhoods of the two vertices take to join up once they have reached a certain size. An advantage of our approach is that it seamlessly integrates (i) and (ii), by looking simply at neighbourhood growth in the whole graph G. Taking this viewpoint, the dual parameter $\lambda_\star$ arises as follows: let $X_t^+ \subset X_t$ be the set of particles of $\mathfrak{X}_\lambda$ that survive (have descendants in all future generations). Then X_0^+ contains the initial particle with probability s, and is empty otherwise. Moreover, conditioning on a particle being in X_t^+ is exactly the same as conditioning on at least one its children surviving, so the number of surviving children then has the distribution $Z = Z_\lambda$ of a $\mathrm{Po}(s\lambda)$ random variable conditioned to be at least 1. Hence, (X_t^+) is again a Galton–Watson branching process, but now with offspring distribution Z, and X_0^+ either empty or, with probability s, consisting of a single particle. Note that

$$\mathbb{P}(Z = 1) = \frac{s\lambda e^{-s\lambda}}{1 - e^{-s\lambda}} = \frac{s\lambda(1-s)}{s} = \lambda_\star, \tag{1.8}$$

using (1.1) and (1.2). Hence, the probability that X_t^+ consists of a single particle, given that the whole process survives, is exactly $\lambda_\star^t$. Roughly speaking, this event corresponds to the branching process staying 'thin' for t generations, i.e., the neighbourhood growth process taking time t to 'get going'. In the next section we shall prove a more precise version of this statement.

Turning to Theorem 1.3, the form of the diameter here differs from what one might expect in the presence of the factors ε^3 inside the logarithms in the numerators. Very loosely speaking, these factors turn out to be related to the fact that the branching process survives with probability $\Theta(\varepsilon)$, and then usually has size $\Theta(1/\varepsilon)$ larger than its unconditional expected size, as well as to the fact that, roughly speaking, it takes on the order of $1/\varepsilon$ generations for anything much to happen; we shall return to this at various points. An alternative way of thinking about these factors is that the 'interesting' structure of $G(n,p)$ is captured by the *kernel*, the graph obtained from the 2-core by suppressing vertices of degree 2. The results of Łuczak [31] or alternatively Pittel and Wormald [36] imply in particular that the number of vertices in the kernel is asymptotically $8\varepsilon^3 n/6$.

Remark. In the first draft of this paper, we obtained a slightly weaker form of Theorem 1.3, giving the same conclusion but requiring an additional assumption, that $\varepsilon^3 n$ grows at least as fast as an explicit *extremely* slowly growing function of n (essentially $\log^* n$, i.e., the minimum k such that the kth iterated logarithm of n is less than 1). This is a less restrictive assumption than what is common in related contexts, that $\varepsilon^3 n$ is at least some power of $\log n$. Since then, Ding, Kim, Lubetzky and Peres [19, 20] have obtained a form of Theorem 1.3 with a larger error term (a multiplicative factor of $1 + o(1)$), valid whenever $\varepsilon^3 n \to \infty$ and $\varepsilon \to 0$; under these assumptions, $\log \lambda \sim \varepsilon$ and $\log(1/\lambda_\star) \sim \varepsilon$, so the diameter is $(3 + o_{\mathrm{p}}(1)) \log(\varepsilon^3 n)/\varepsilon$. Their approach, based around the 2-core and kernel, is very different to ours. Seeing this paper stimulated us to remove the unnecessary restriction on ε; it turned out that one simple observation (Lemma 4.28 below) was the main missing ingredient. Using this lemma, it became possible to simplify some of our original arguments and, with a little further technical work, to extend them to the entire weakly supercritical range.

We remark also that Łuczak and Seierstad [33] have obtained a 'process version' of Theorem 1.3. This gives much weaker bounds on the diameter, differing by a constant factor, but can be applied to the random graph process to show (roughly speaking) that whp these bounds hold simultaneously for the entire range of densities considered in Theorem 1.3.

In Theorem 1.3, the condition $\varepsilon \le 1/10$ is imposed simply for convenience; this may be weakened to $\varepsilon = O(1)$ without problems. However, for $\varepsilon = O(1)$ bounded away from zero one can instead apply Theorem 1.1: it is not hard to check that the constants implicit in the correction term vary smoothly with λ, and so are bounded over any compact set of $\varepsilon > 0$. For this reason, in proving Theorem 1.3 we may assume that $\varepsilon \to 0$ as $n \to \infty$; we shall do this whenever it is convenient.

On the other hand, the condition $\varepsilon n^{1/3} \to \infty$ is almost certainly necessary for our method to give nontrivial information. If $\varepsilon n^{1/3}$ is bounded, then we are inside the 'window' of the phase transition, so $G(n, (1+\varepsilon)/n)$ is qualitatively similar in behaviour to $G(n, 1/n)$, and the behaviour of the diameter is much more complicated than outside the window. For one thing, there is no longer a unique 'giant' component that is much larger than all other components. Also, the 2-core of each non-tree component contains only a bounded number of cycles; to study the distribution of the diameter accurately, one needs to study the distribution of the lengths of these cycles, which is very different from the situation with supercritical graphs. Nachmias and Peres [34] showed that inside the window the diameter of the largest component is $O_{\mathrm{p}}(n^{1/3})$, with a corresponding lower bound; more recently, Addario-Berry, Broutin and Goldschmidt [1] have established convergence in distribution of the rescaled diameter, and given a (rather complicated) description of the limit in terms of continuum random trees.

Finally, as in Theorem 1.1, the $O_{\mathrm{p}}(1/\varepsilon)$ correction term in Theorem 1.3 is in some sense best possible. As noted earlier, our method gives a description of the limiting distribution of this correction term; see Section 5.

In summary, the results of Łuczak [32] (below the critical window), Addario-Berry, Broutin and Goldschmidt [1] (inside the window), Theorem 1.3 (above but average degree tending to 1), Theorem 1.1 (constant average degree), Theorem 1.2 (average degree tending to infinity slowly) and Bollobás [6] (average degree tending to infinity quickly) together establish tight bounds on the diameter of $G(n, p)$ throughout the entire range of the parameters.

2 The case $p = \lambda/n$, $\lambda > 1$ constant

In this section we shall prove Theorem 1.1. We start by recalling a basic fact about branching processes.

From standard branching process results (see, for example, Athreya and Ney [3]), the martingale $|X_t|/\lambda^t$ converges almost surely to a random variable $Y = Y_\lambda$ whose distribution (which depends on λ) is continuous except for mass $1-s$ at 0, with strictly positive density on $\mathbb{R}^+$. Furthermore, $Y = 0$ coincides (except possibly on a set of measure zero) with the event that the branching process dies out. Since almost sure convergence implies convergence in probability, a trivial consequence of this is that, for $\lambda > 1$ and $0 < c_1 < c_2$ all fixed,

$$\inf_t \mathbb{P}\big(c_1\lambda^t \le |X_t| \le c_2\lambda^t\big) > 0, \tag{2.1}$$

where the infimum is over all $t \ge 1$ such that the interval $[c_1\lambda^t, c_2\lambda^t]$ contains an integer. The following result indicates that unusually small populations in a given generation are typically due (at least, with a significant probability) to a branching process that stays essentially nonbranching (with only small 'side branches') until a point where it branches at a typical rate.

Lemma 2.1. *Let $\lambda > 1$ be fixed. There are constants $c, C > 0$ such that for every $\omega \geq 2$ and $t \geq 1$ we have*

$$c \min\{\lambda_\star^{t-t_1}, 1\} \leq \mathbb{P}\big(0 < |X_t| < \omega\big) \leq C\lambda_\star^{t-t_1}, \tag{2.2}$$

where $t_1 = \lfloor \log \omega / \log \lambda \rfloor$.

Proof. The lemma is essentially a statement about the asymptotics of Y near 0; this statement follows, for example, from a result of Harris [24]. However, translating back to a statement about X_t rather than Y would introduce an extra error term, corresponding to the probability that X_t/λ^t still differs from Y by more than a constant factor when X_t first exceeds ω, so we shall give a direct proof.

We start by proving the upper bound. Conditioned on $\mathfrak{X}_\lambda = (X_t)_{t\geq 0}$ dying out, an event of probability $1-s$, this process has the distribution of the subcritical process $\mathfrak{X}_{\lambda_\star} = (X_t^-)_{t\geq 0}$. Hence,

$$\mathbb{P}\big(|X_t| > 0,\, \exists t' : X_{t'} = \emptyset\big) = (1-s)\mathbb{P}(|X_t^-| > 0) \leq (1-s)\,\mathbb{E}(|X_t^-|) = (1-s)\lambda_\star^t.$$

Let $p_t = \mathbb{P}(|X_t| > 0)$. Then

$$p_t = s + \mathbb{P}\big(|X_t| > 0,\, \exists t' : X_{t'} = \emptyset\big) = s + O(\lambda_\star^t). \tag{2.3}$$

Let us note for later that the implicit constant is independent of λ; indeed, it may be taken to be 1.

We may partition X_1, the set of children of the initial particle, into two sets: the set S consisting of those that have descendants $t-1$ generations later (i.e., in X_t), and the set $X_1 \setminus S$ of those that do not. Since the probability that a particle in X_1 has one or more descendants in X_t is p_{t-1}, the size of S has a Poisson distribution with mean λp_{t-1}. Let us condition on $|X_t| > 0$. Then the conditional distribution of $|S|$ is that of a Poisson distribution with mean λp_{t-1} conditioned on being at least 1, and we have

$$\mathbb{P}\big(|S| = 1 \;\big|\; |X_t| > 0\big) = \frac{\lambda p_{t-1} e^{-\lambda p_{t-1}}}{1 - e^{-\lambda p_{t-1}}} = \frac{\lambda s e^{-\lambda s}}{1 - e^{-\lambda s}}(1 + O(\lambda\lambda_\star^{t-1})) = \lambda_\star + O(\lambda\lambda_\star^t),$$

using (2.3) and (1.8). Note for later use that the implicit constant is independent of λ provided $\lambda > 1$ and λ is bounded away from 1.

Let $r_t = \mathbb{P}(|X_t| < \omega \mid |X_t| > 0)$. If $|X_t| < \omega$, then every particle in S has fewer than ω descendants in X_t. Hence,

$$\begin{aligned} r_t &\leq \mathbb{P}\big(|S| = 1 \;\big|\; |X_t| > 0\big) r_{t-1} + \mathbb{P}\big(|S| > 1 \;\big|\; |X_t| > 0\big) r_{t-1}^2 \\ &= (\lambda_\star + O(\lambda\lambda_\star^t)) r_{t-1} + (1 - \lambda_\star + O(\lambda\lambda_\star^t)) r_{t-1}^2 \\ &= r_{t-1}(\lambda_\star + (1-\lambda_\star) r_{t-1}) + O(\lambda\lambda_\star^t r_{t-1}). \end{aligned} \tag{2.4}$$

Setting $r'_t = r_t/\lambda_\star^t$ and recalling that λ is constant, we thus have

$$r'_t \le r'_{t-1} + \frac{1-\lambda_\star}{\lambda_\star} r_{t-1} r'_{t-1} + O(r_{t-1}). \tag{2.5}$$

Using only the trivial inequality $\mathbb{P}(0 < |X_t| < \omega_1) \le \mathbb{P}(0 < |X_t| < \omega)$ for $\omega_1 < \omega$, the upper bound in (2.2) for ω at least some constant ω_0 implies the same bound, with a different constant, for all $\omega \ge 2$. Thus we may assume that ω is at least some large constant ω_0, and hence that t_1 is large. We may also assume $t \ge t_1$. By (2.1) we have $\mathbb{P}(|X_{t_1}| > \omega) \ge c_0$ for some constant $c_0 > 0$. Hence $r_{t_1} \le 1 - c_0$ is bounded away from 1. Choosing ω_0 large enough, so $\lambda_\star^t \le \lambda_\star^{t_1}$ is small, the error term in (2.4) can be assumed arbitrarily small relative to r_{t-1}. Using (2.4), and noting that for $t > t_1$ we have $\lambda_\star + (1-\lambda_\star) r_{t-1} < \lambda_\star + (1-\lambda_\star)(1-c_0) < 1$, it then follows that r_t decreases exponentially as t increases from t_1, i.e., that there is a $c_1 > 0$ (depending only on λ, not on ω) such that $r_{t_1+t} \le e^{-c_1 t}$. Hence, $\sum_{t \ge t_1} r_t$ is bounded (independently of ω). Using (2.5), it follows that there is a constant C_0 such that for $t \ge t_1$ we have $r'_t \le C_0(r'_{t_1} + 1)$. In other words,

$$r_t \le C_0 \lambda_\star^{t-t_1} r_{t_1} + C_0 \lambda_\star^t \le C_0(1+\lambda_\star^{t_1})\lambda_\star^{t-t_1} \le 2C_0 \lambda_\star^{t-t_1} = O(\lambda_\star^{t-t_1}).$$

Since

$$\mathbb{P}\big(0 < |X_t| < \omega\big) \le \mathbb{P}\big(|X_t| < \omega \;\big|\; |X_t| > 0\big) = r_t,$$

this completes the proof of the upper bound.

Turning to the lower bound, this is essentially trivial if $t \le t_1$: in this case, $\mathbb{P}(0 < |X_t| < \omega)$ is bounded away from 0 by (2.1). We may thus assume that $t > t_1$. We shall prove the lower bound by considering the following much more specific event E, the event that $|X^+_{t-t_1}| = 1$, that the unique particle v of $X^+_{t-t_1}$ has between 1 and $\omega - 1$ descendants in X_t, and that no other particles of X_{t-t_1} have descendants in X_t. Clearly, if E holds then $0 < |X_t| < \omega$.

Recalling that $\mathfrak{X}^+ = (X^+_t)$ is the set of particles whose descendants survive forever, any such particle always has at least one child by definition, and, by (1.8), has exactly one child with probability $\lambda_\star$. Thus

$$\mathbb{P}(|X^+_{t-t_1}| = 1) = s\lambda_\star^{t-t_1}. \tag{2.6}$$

Given that $|X^+_{t-t_1}| = 1$, the number N_v of descendants in X_t of the unique particle v in $X^+_{t-t_1}$ has the distribution of $|X_{t_1}|$ conditioned on the whole process surviving. From (2.1), the (unconditional) probability that $|X_{t_1}|$ is between $\omega/2$ and $\omega - 1$, say, is bounded away from zero, and the conditional probability that $\mathfrak{X}_\lambda$ survives given this event is at least s. Thus

$$\begin{aligned} \mathbb{P}\big(N_v < \omega \;\big|\; |X^+_{t-t_1}| = 1\big) &= \mathbb{P}\big(|X_{t_1}| < \omega \;\big|\; \forall t : |X_t| > 0\big) \\ &\ge \mathbb{P}\big(|X_{t_1}| < \omega,\, \forall t : |X_t| > 0\big) \ge c_2, \end{aligned}$$

for some positive constant c_2.

It remains to exclude descendants in X_t of other particles in X_{t-t_1}. By definition, these particles do not survive. We may construct $\mathfrak{X}_\lambda$ as follows: first construct $\mathfrak{X}^+ = (X^+_t)_{t \ge 0}$. Then add in the particles that die: for each particle in each set X^+_r, we must add an independent copy of $\mathfrak{X}_{\lambda_\star}$ rooted at this particle.

Given that $|X^+_{t-t_1}| = 1$, we have $|X^+_r| = 1$ for all $r \le t - t_1$. The probability that the copy of $\mathfrak{X}_{\lambda_\star}$ started at time r survives to time t is $\mathbb{P}(|X^-_{t-r}| > 0) \le \lambda_\star^{t-r}$. Since the different copies are independent, the probability that all die before time t is at least $\prod_{r\le t-t_1}(1-\lambda_\star^{t-r})$. Now $\lambda_\star < 1$, so $\sum_{r\le t-t_1}\lambda_\star^{t-r} = O(\lambda_\star^{t_1}) = O(1)$, and $\prod_{r\le t-t_1}(1-\lambda_\star^{t-r}) \ge c_3$, for some $c_3 > 0$ depending only on λ. Hence,

$$\mathbb{P}\big(0 < |X_t| < \omega\big) \ge s\lambda_\star^{t-t_1}c_2c_3 = \Omega(\lambda_\star^{t-t_1}),$$

completing the proof of the lemma. □

The above lemma tells us virtually all we need to know about the branching process for the 'early growth' part of the proof of Theorem 1.1. The next ingredient for this phase is a lemma connecting the growth of neighbourhoods in the graph to the branching process. The branching process model is most relevant if the growing neighbourhood of a vertex remains a tree. To be sure, almost all vertices do not lie on or near a short cycle. However, we cannot simply ignore the exceptional vertices, since a result about the diameter makes a statement about *all* vertices, not just almost all. So we must be a little careful.

We deal with the problem of non-tree neighbourhoods as follows. Given a vertex x of a graph G, let $\Gamma_t(x)$ be the set of vertices at graph distance t from x. Let $G_{\le t}(x)$ be the subgraph of G induced by $\bigcup_{t'\le t}\Gamma_{t'}(x)$, regarded as a rooted graph with root x. We shall explore the neighbourhoods $\Gamma_t(x)$ in the following essentially standard way. Fix once and for all an order on $V(G)$. Having found $\Gamma_t(x)$ (starting with $t = 0$), go through the vertices of $\Gamma_t(x)$ one by one in the predetermined order. For each vertex v we expose all edges from v to vertices not yet reached in the exploration; this means we test each potential edge *to an as yet unreached vertex* for its presence; any edges detected are called 'uncovered.' If we uncover an edge vw, we add w to $\Gamma_{t+1}(x)$. Of course this process correctly identifies the sets $\Gamma_t(x)$. However, it only uncovers certain edges: let $G^0_{\le t}(x)$ denote the graph formed by the edges uncovered in our tests exploring up to $\Gamma_t(x)$. Then $G^0_{\le t}(x)$ is a tree: it is a spanning tree in the graph $G_{\le t}(x)$.

In the following results, $X_{\le t}$ denotes the union of generations 0 to t of the branching process $\mathfrak{X}_\lambda$, regarded as a rooted tree with root the initial particle, and $\cong$ denotes isomorphism of rooted trees.

Lemma 2.2. *Let $\lambda > 0$ be fixed. For any rooted tree T with $|T| \le n/2$ we have*

$$\mathbb{P}\big(G^0_{\le t}(x) \cong T\big) = e^{O(|T|^2/n)}\mathbb{P}\big(X_{\le t} \cong T\big)$$

and

$$\mathbb{P}\big(G_{\le t}(x) \cong T\big) = e^{O(|T|^2/n)}\mathbb{P}\big(X_{\le t} \cong T\big),$$

where the implicit constants depend only on λ.

Proof. This is well known and easy to prove. The first statement follows from the natural step-by-step coupling between $G^0_{\le t}(x)$ and the branching process, where each step investigates the children (of a vertex or a particle, respectively). Suppose we have reached $r-a$ vertices in total so far. Then the probabilities of finding a vertices in the next step are $p_1 = \binom{n-r+a}{a}(\lambda/n)^a(1-\lambda/n)^{n-r}$ and $p_2 = e^{-\lambda}\lambda^a/a!$ in the two models. The ratio of these probabilities is

$$p_1/p_2 = (n-r+a)_{(a)}n^{-a}(1-\lambda/n)^{-r}e^{\lambda}(1-\lambda/n)^n = e^{O(ar/n+r\lambda/n+\lambda^2/n)},$$

where $x_{(a)} = x(x-1)\cdots(x-a+1)$. The sum of ar or r over all vertices in the tree is trivially at most $|T|^2$, so it follows that

$$\frac{\mathbb{P}\big(G^0_{\le t}(x) \cong T\big)}{\mathbb{P}\big(X_{\le t} \cong T\big)} = \exp\big(O(|T|^2/n + \lambda|T|^2/n + \lambda^2|T|/n)\big). \tag{2.7}$$

Since λ is fixed, this proves the first statement.

If $G_{\le t}(x) \cong T$, then $G_{\le t}(x)$ is a tree, so $G^0_{\le t}(x) = G_{\le t}(x)$. Hence $G_{\le t}(x) \cong T$ implies $G^0_{\le t}(x) \cong T$. Given that $G^0_{\le t} \cong T$, the probability that none of the untested edges between the $|T|$ vertices found is also present is again $e^{O(|T|^2/n)}$. So the second statement follows from the first. □

Using Lemmas 2.1 and 2.2, we can study the initial rate of growth of the neighbourhoods of the vertices of $G(n, \lambda/n)$. The first step is to show that these neighbourhoods cannot stay small but non-empty for too long. The basic picture is that after about

$$t_1 = \lfloor \log\omega / \log\lambda \rfloor \tag{2.8}$$

steps, we expect a typical vertex neighbourhood to expand to size approximately ω. It is very unlikely that there are any vertices in the graph whose neighbourhoods expand to some reasonable size, say around $\log n$, and then fail to expand to size ω in roughly the expected time from that point. However, some unusual vertices take up to

$$t_0 = \lfloor \log n / \log(1/\lambda_\star) \rfloor \tag{2.9}$$

steps before their neighbourhoods expand significantly, and so take this many more steps than usual to reach size roughly ω.

We argue this more precisely as follows.

Set $\omega = (\log n)^6$, say, and define t_0 and t_1 as above. Let $K = K(n)$ tend to infinity slowly (for instance, slower than $\log\log n$).

For each vertex x, let $B_1(x)$ be the 'bad' event that $1 \le |\Gamma_{t'}(x)| < \omega$ holds for all $0 \le t' \le t = t_0 + t_1 + K$. The event $B_1(x)$ is a disjoint union of events of the form $G^0_{\le t} \cong T$, where each tree T has size at most $t\omega = o(\sqrt{n})$. Also, the corresponding union of the events $X_{\le t} \cong T$ is the event that $0 < |X_{t'}| < \omega$ holds for all $t' \le t$. Hence, by Lemma 2.2,

$$\mathbb{P}(B_1(x)) \sim \mathbb{P}\big(\forall t' \le t : 0 < |X_{t'}| < \omega\big) \le \mathbb{P}\big(0 < |X_t| < \omega\big) = O(\lambda_\star^{t_0+K}), \tag{2.10}$$

where the last step is from Lemma 2.1.

Let B_1 be the event that $B_1(x)$ holds for some x. Then

$$\mathbb{P}(B_1) \le n\mathbb{P}(B_1(x)) = O(n\lambda_\star^{t_0+K}) = O(\lambda_\star^K) = o(1). \tag{2.11}$$

We now move on to the 'regular growth' part of the proof. That is, our next aim is to show that once the neighbourhoods of a vertex x reach size ω, with very high probability they then grow at a predictable rate until they reach size comparable with n. We shall use the following convenient form of the Chernoff bounds on the binomial distribution; see [27], for example.

Lemma 2.3. *Let Y have a binomial distribution with parameters n and p. If $0 \le \delta \le 1$ then*

$$\mathbb{P}\big(|Y - np| \ge \delta np\big) \le 2e^{-\delta^2 np/3}.$$

□

Let $0 < \delta < 1/1000$ be an arbitrary (small) constant. Let us say that a vertex x has *regular large neighbourhoods* if one of the following holds: either $|\Gamma_t(x)| < \omega$ for all t, or, setting $t^- = \min\{t : |\Gamma_t(x)| \ge \omega\}$ and $t^+ = t^- + \log(n^{3/4}/\omega)/\log\lambda$, we have

$$(1-\delta)\lambda^{t-t^-+1}|\Gamma_{t^--1}(x)| \le |\Gamma_t(x)| \le (1+\delta)\lambda^{t-t^-+1}|\Gamma_{t^--1}(x)|$$

for $t^- \le t \le t^+$. In other words, the neighbourhoods grow by almost exactly a factor of λ at each step from just before the first time they reach size ω until they reach size around $n^{3/4}$. Note that since we start from the last 'small' neighbourhood $\Gamma_{t^--1}(x)$, the growth condition above certainly implies that

$$\frac{1-\delta}{1+\delta} \;\le\; \frac{|\Gamma_t(x)|}{\omega\lambda^{t-t^-}} \;\le\; \lambda(1+\delta) \tag{2.12}$$

holds for $t^- \le t \le t^+$.

Let $B_2(x)$ be the 'bad' event that a given vertex x of $G(n, \lambda/n)$ fails to have regular large neighbourhoods, and $B_2 = \bigcup_x B_2(x)$ the global bad event that not all vertices have regular large neighbourhoods.

Lemma 2.4. *For each fixed vertex x of $G(n, \lambda/n)$ we have $\mathbb{P}(B_2(x)) = o(n^{-1})$. Thus $\mathbb{P}(B_2) = o(1)$.*

Proof. This is well known (c.f. Janson, Łuczak and Ruciński [28, Section 5.2]), and essentially trivial from the Chernoff bounds (or Hoeffding's inequality); we nevertheless give the details. We explore the successive neighbourhoods of x in $G(n, \lambda/n)$ in the usual way, writing a_t for $|\Gamma_t(x)|$. Conditional on $a_0, a_1, \ldots, a_t$, the distribution of a_{t+1} is binomial with parameters $n - m$ and $p = 1 - (1 - \lambda/n)^{a_t}$, where $m = \sum_{t' \le t} a_{t'}$ and p is the probability that one of the undiscovered vertices is adjacent to at least one member of $\Gamma_t(x)$. Assuming that $m = O(n^{3/4})$, say, we have $\mathbb{E}(a_{t+1} \mid a_0, \ldots, a_t) = \lambda a_t(1 + O(n^{-1/4}))$. It then follows from Lemma 2.3 that, conditional on $a_0, \ldots, a_t$, if $a_t \ge \omega/(100\lambda)$ then we have

$$\mathbb{P}\left(\left|\frac{a_{t+1}}{a_t} - \lambda\right| \ge \frac{1}{(\log n)^2}\right) = e^{-\Omega((\log n)^{-4}a_t)} = o(n^{-100}),$$

using $\omega = (\log n)^6$. Similarly, if $a_t < \omega/(100\lambda)$ then $\mathbb{P}(a_{t+1} \ge \omega) \le n^{-100}$.

Let t^- be the first t with $a_t \geq \omega$, if such a t exists. We have already shown above that the probability that $0 < a_t < \omega$ holds for all t up to $t_0 + t_1 + K = O(\log n)$ is $o(n^{-1})$, so with probability $1 - o(n^{-1})$ either t^- is undefined, in which case there is nothing to prove, or $t^- = O(\log n)$, in which case we have so far uncovered $O(t^- \omega) = o(n^{3/4})$ vertices. From the estimates above, with very high probability $a_{t^- -1} \geq \omega/(100\lambda)$, and, from this point on, the ratios a_{t+1}/a_t are within a factor $1 + O((\log n)^{-2})$ of λ until a_t first exceeds $n^{3/4}$. It follows that x has regular large neighbourhoods with probability $1 - o(n^{-1})$, as claimed. □

Note that we took ω as large as $(\log n)^6$ just to simplify the estimates. If we are a little more careful, a large constant times $\log n$ will in fact do: significant deviations in the ratio a_{t+1}/a_t are only likely near the beginning, so we can bound these ratios above and below by sequences approaching 1 geometrically with high enough probability.

We now move onto the third phase of the proof, where we consider the meeting up of neighbourhoods of different vertices and hence the distance between them. This still involves a careful look at the early development of neighbourhoods, since, from the second phase of the proof, we know that vertices with large close neighbourhoods will have large distant neighbourhoods. We treat the upper and lower bounds in the Theorem 1.1 separately.

2.1 Upper bound

As above, set $\omega = (\log n)^6$, say, and let $K = K(n)$ tend to infinity slowly.

For $x \in V(G)$ let $t_\omega(x) = \min\{t : |\Gamma_t(x)| \geq \omega\}$, if this minimum exists; otherwise $t_\omega(x)$ is undefined. Note that if the event B_1 defined above does not hold, then whenever $t_\omega(x)$ is defined, we have $t_\omega(x) \leq t_0 + t_1 + K$.

Set

$$t_2 = \lfloor \log(n/\omega^2)/\log \lambda \rfloor,$$

and, for $x, y \in V(G)$, let $E_{x,y,i,j}$ be the event that $t_\omega(x) = t_0 + t_1 - i$, $t_\omega(y) = t_0 + t_1 - j$, and $d(x, y) \geq t_\omega(x) + i + t_\omega(y) + j + t_2 + 3K + c_0$ all hold, where $c_0 > 2$ is some constant. Our next aim is to bound the probability of the event $E_{x,y,i,j}$ for given vertices x and y and given $i, j \geq -K$.

Recall that B_2 is the event that not all vertices have regular large neighbourhoods. We claim that there is some $c > 0$ (depending only on λ) such that

$$\begin{aligned} \mathbb{P}(E_{x,y,i,j} \setminus B_2) &\leq \lambda_\star^{t_0-i}\lambda_\star^{t_0-j}e^{-c\lambda^{3K+i+j}} + o(n^{-100}) \\ &= O\big(n^{-2}\lambda_\star^{-i-j}e^{-c\lambda^{3K+i+j}}\big) + o(n^{-100}). \end{aligned} \tag{2.13}$$

First, arguing as in the proof of (2.10), using Lemma 2.1 and a version of Lemma 2.2 where we start with two vertices and compare with two copies of the branching process, we see that

$$\mathbb{P}\big(t_\omega(x) = t_0 + t_1 - i,\, t_\omega(y) = t_0 + t_1 - j,\, d(x, y) > t_\omega(x) + t_\omega(y)\big) = O(\lambda_\star^{t_0-i}\lambda_\star^{t_0-j}).$$

Exploring the neighbourhoods of x and y in the obvious way, suppose we find that $t_\omega(x) = t_0 + t_1 - i$, $t_\omega(y) = t_0 + t_1 - j$, and $d(x, y) > t_\omega(x) + t_\omega(y)$, i.e., our explorations have not yet met. Set $\ell = \lfloor (t_2 + 3K + i + j)/2 \rfloor$. Suppose for the moment that $\ell \leq \ell_0 = \log(n^{3/4}/\omega)/\log \lambda$.

Continuing the exploration of the two neighbourhoods a further ℓ steps in each case, we may assume that with respect to the neighbourhoods of x and y that have been revealed so far, the regular large neighbourhood condition has not yet been violated. (If it has been, the event B_2 must hold, and we are bounding the probability of an event contained in the complement of B_2.) Then

$$\min\{|\Gamma_{t_\omega(x)+\ell}(x)|,\, |\Gamma_{t_\omega(y)+\ell}(y)|\} \geq 0.99\omega\lambda^\ell = \Omega\big(\sqrt{n\lambda^{3K+i+j}}\big).$$

It may be that $d(x,y) \leq t_\omega(x)+\ell+t_\omega(y)+\ell$, in which case we are done. Otherwise, the edges between $\Gamma_{t_\omega(x)+\ell}(x)$ and $\Gamma_{t_\omega(y)+\ell}(y)$ have not yet been tested, so the chance that no such edge is present is

$$(1-\lambda/n)^{|\Gamma_{t_\omega(x)+\ell}(x)||\Gamma_{t_\omega(y)+\ell}(y)|} \leq e^{-(\lambda/n)\Omega(n\lambda^{3K+i+j})} \leq e^{-c\lambda^{3K+i+j}},$$

for some constant $c > 0$. Multiplying by the $O(\lambda_\star^{t_0-i}\lambda_\star^{t_0-j})$ bound obtained above gives (2.13) in this case.

If $\ell > \ell_0$, the argument is similar; this time, assuming B_2 does not hold only allows us to control the sizes of the neighbourhoods for $\ell_0 < \ell$ steps beyond $t_\omega(x)$ and $t_\omega(y)$. But by this time they reach size at least $n^{3/4}/2$, and the probability that they do not join is at most $e^{-(\lambda/n)n^{3/2}/4} = o(n^{-100})$.

Let B be the event that

$$\begin{aligned} \mathrm{diam}(G) \geq 2t_0+2t_1+t_2+3K+10 \quad &\geq \quad 2\frac{\log\omega}{\log\lambda} + 2\frac{\log n}{\log(1/\lambda_\star)} + \frac{\log n - 2\log\omega}{\log\lambda} + 3K \\ &= \quad \frac{\log n}{\log\lambda} + 2\frac{\log n}{\log(1/\lambda_\star)} + 3K. \end{aligned}$$

Our aim is to prove that with $K \to \infty$ arbitrarily slowly, we have $\mathbb{P}(B) = o(1)$; in order to do so, it suffices to show that $\mathbb{P}(B \setminus (B_1 \cup B_2)) = o(1)$.

Suppose that B holds but $B_1 \cup B_2$ does not, and let x and y be vertices at maximum distance. Since B_1 does not hold, and $d(x,y)$ is so large, exploring successive neighbourhoods of x and y, these neighbourhoods both reach size at least ω before they meet. Hence $E_{x,y,i,j}$ holds for some i and j. Since B_1 does not hold, $E_{x,y,i,j}$ can only hold if $i,j \geq -K$. Hence, using (2.13),

$$\begin{aligned} \mathbb{P}(B \setminus (B_1 \cup B_2)) \leq \sum_{i,j\geq -K}\ \sum_{x,y\in V(G)} \mathbb{P}(E_{x,y,i,j} \setminus B_2) \\ \leq o(n^{-90}) + n^2 \sum_{i,j\geq -K} n^{-2} O\left(\lambda_\star^{-i-j} e^{-c\lambda^{3K+i+j}}\right), \end{aligned}$$

which is $o(1)$ since

$$\sum_{r\geq -2K} (r+2K+1)\lambda_\star^{-r} e^{-c\lambda^{3K+r}} = O(e^{-c\lambda^K}) = o(1).$$

This completes the proof of the upper bound.

Remark. Note that one cannot prove the upper bound directly by the first moment method; a separate argument excluding very long thin neighbourhoods (bounding the probability of B_1) is needed. Indeed, it is not too hard to show that the estimates above are essentially tight. Thus, if for some $r \geq K$ there happens to be a vertex x with $t_\omega(x) = t_0 + t_1 + r$, say, an event of probability around $\lambda_\star^r$, then x will be at distance roughly $d = 2t_0 + 2t_1 + t_2 + K$ from many of the roughly $(1/\lambda_\star)^{r-K}$ vertices y with $t_\omega(y) = t_0 + t_1 - r + K$. Since there are $\Theta(\log n)$ possible values of r, the expected number of pairs of vertices at distance d will tend to infinity if $K \to \infty$ slowly enough.

2.2 Lower bound

The idea of the lower bound is simple. Let S be the set of vertices x with $t_\omega(x) \geq t_0 + t_1 - K$. Then, from the arguments in the previous section, the expected size of S is roughly $(1/\lambda_\star)^K$, which tends to infinity. We would like to show that $|S|$ is large with high probability using the second moment method. Since two vertices in S are likely to be far apart, the result will follow. There are two problems. A minor one is that the events that different vertices lie in S are not that close to independent: vertices in S will usually be located in trees attached to the 2-core, and S roughly corresponds to the set of vertices at least a certain distance from the 2-core. Although most trees attached to the 2-core will contain no such vertices, it turns out that, on average, each tree contributing one or more such vertices contributes some constant number larger than 1, so $|S|$ is not well approximated by a Poisson distribution. A more serious, related, problem is that to find vertices at large distance we need to find vertices in S whose short-range neighbourhoods do not overlap, i.e., vertices coming from different trees. We solve both these problems by looking for vertices $x \in S$ satisfying an additional condition, the *strong wedge condition*, that usually corresponds to x being the unique vertex in its tree at maximal distance from the 2-core.

Note that as we are now looking for a lower bound on the diameter, we do not need to consider all promising pairs of vertices for our candidate vertices at large distance. We may thus impose additional conditions as convenient, and our result will still be sharp enough as long as these conditions are likely enough to be satisfied. One such condition is that the neighbourhoods are trees up to a suitable distance.

Let $x \in V(G)$, and suppose that $G_{\leq t}(x)$ is a tree for some $t > 0$. The *weak/strong wedge condition* holds from x to $x_2 \in \Gamma_t(x)$ if for every $z \neq x$ in the graph $G_{\leq t}(x)$, the distance from z to the closest vertex y on the unique path from x to x_2 is at most/strictly less than the distance from x to y. Note that either condition implies that the degree of x in G must be 1. In this section we shall always work with the strong wedge condition; the weak wedge condition will play a role in Section 4.

Let t_K denote $t_0 - K$, where $K = K(n) \to \infty$ arbitrarily slowly, in particular with $K \leq \log \log n$, and let W_x^0 be the event that $G_{\leq t_K}(x)$ is a tree with the following properties: there is a unique vertex x_2 at distance t_K from x, and the strong wedge condition holds from x to x_2. Let W_x be the event that W_x^0 holds and $G_{\leq t_K}(x)$ contains fewer than $\omega/2$ vertices, where $\omega = (\log n)^6$ as before.

Note for later that if W_x (or W_x^0) holds, then the tree $G_{\leq t_K}(x)$ consists of an x-x_2 path P_x of length t_K with a (possibly empty) set of trees attached to each interior vertex, the height of each tree being strictly less than the distance to the nearest endvertex of P_x. (Thus W_x^0 is a sort of 'diamond' condition. We will use a precise version of this terminology in the next section.) It follows that the diameter of $G_{\leq t_K}(x)$ is t_K, and that x and x_2 are the unique pair of vertices of $G_{\leq t_K}(x)$ at this distance.

Let W^0 and W be the branching process events corresponding to W_x^0 and W_x, so W^0 is the branching process version of our diamond condition. The event that W holds is a disjoint union of events that $X_{\leq t_K}$ is one of certain trees with at most $\omega/2 = o(n^{1/2})$ vertices, so by Lemma 2.2 we have $\mathbb{P}(W_x) \sim \mathbb{P}(W)$.

Once the branching process reaches size $(\log n)^4$, it is very unlikely ever to shrink down to size 1, and in fact the probability that W^0 holds but one of the first $t_K = t_0 - K$ generations has size at least $(\log n)^4$ is $o(n^{-100})$. (This follows from the proof of Lemma 2.4, but is much simpler.) Assuming this does not happen, the sum of sizes of the first $t_0 - K$ generations is at most $t_0(\log n)^4 = O(\log^5 n)$. It follows that $\mathbb{P}(W^0 \setminus W) = o(n^{-100})$, so

$$\mathbb{P}(W) = \mathbb{P}(W^0) + o(n^{-100}). \tag{2.14}$$

To calculate $\mathbb{P}(W^0)$, consider the event W', that W^0 holds and the unique particle in generation $t_0 - K$ survives. Note that

$$\mathbb{P}(W') = s\mathbb{P}(W^0). \tag{2.15}$$

If W' holds, then $|X_t^+| = 1$ for $t = t_0 - K$ and hence for $t = 0, 1, \ldots, t_0 - K$, an event of probability $s\lambda_\star^{t_0-K}$. Conversely, constructing $\mathfrak{X}_\lambda$ as before by starting from $\mathfrak{X}^+$ and adding in independent copies of $\mathfrak{X}_{\lambda_\star} = (X_r^-)$ started at each particle, W' holds if and only if $|X_{t_0-K}^+| = 1$ and, for $0 \leq t < t_0 - K$, the copy of $\mathfrak{X}_{\lambda_\star}$ started at the unique particle of X_t^+ dies within $\min\{\max\{t, 1\}, t_0 - K - t\}$ generations: dying within $t_0 - K - t$ generations ensures that $|X_{t_0-K}| = |X_{t_0-K}^+| = 1$, and, for $t > 0$, dying within t generations ensures that the strong wedge condition holds. Let $d_t = \mathbb{P}(|X_t^-| = 0)$ be the probability that the subcritical process $\mathfrak{X}_{\lambda_\star}$ dies within t generations. Then we have

$$\begin{aligned}\mathbb{P}(W') = s\lambda_\star^{t_0-K} d_1 \prod_{t=1}^{t_0-K-1} d_{\min\{t, t_0-K-t\}} \\ = s\lambda_\star^{t_0-K} d_1 d_1^2 d_2^2 d_3^2 \cdots d_{\lfloor (t_0-K)/2 \rfloor - 1}^2 d_{\lfloor (t_0-K)/2 \rfloor}^\theta,\end{aligned} \tag{2.16}$$

where the exponent θ of the last factor is 1 or 2 depending on the parity of $t_0 - K$. As we shall see, the later factors in the product are essentially irrelevant. Indeed,

$$1 - d_t = \mathbb{P}(|X_t^-| > 0) \leq \mathbb{E}(|X_t^-|) = \lambda_\star^t, \tag{2.17}$$

so $1 - \lambda_\star^t \leq d_t \leq 1$, and $-\log d_t = O(\lambda_\star^t)$. Since $\sum_t \lambda_\star^t$ is convergent, we thus have $\prod_t d_t = \Theta(1)$, so $\mathbb{P}(W') = \Theta(\lambda_\star^{t_0-K})$ and, using (2.14) and (2.15),

$$\mathbb{P}(W) = \mathbb{P}(W^0) + o(n^{-100}) = s^{-1}\mathbb{P}(W') + o(n^{-100}) = \Theta(\lambda_\star^{t_0-K}).$$

Since $\lambda_\star^{t_0}$ is of order $1/n$ and $K \to \infty$, it follows that $n\mathbb{P}(W) \to \infty$, and hence that $n\mathbb{P}(W_x) \to \infty$.

Recalling that $t_1 = \lfloor \log \omega / \log \lambda \rfloor$, set $t = t_K = t_0 - K$, let W_x^+ be the event that W_x holds, $|\Gamma_{t+t_1}(x)| \ge \omega$, and $|V(G_{\le t+t_1}(x))| < \omega^2$. If W_x holds, then exploring the neighbourhoods of x to distance t we have by definition reached at most $\omega/2$ vertices. Using (2.1), it is easy to show that $\mathbb{P}(W_x^+ \mid W_x) = \Theta(1)$, so $\mathbb{P}(W_x^+) = \Theta(\mathbb{P}(W_x))$.

Let N be the number of vertices x for which W_x^+ holds, so $\mathbb{E}(N) = n\mathbb{P}(W_x^+) = \Theta(n\mathbb{P}(W_x)) \to \infty$. We shall use the second moment method to show that N is concentrated about its mean. The argument is slightly more complicated than one might expect (or hope for); while one can give simpler arguments that are very plausible, we have so far failed to turn such an argument into a rigorous proof. In fact, the argument we do present deals with all issues of possible dependence with very little calculation.

Suppose that x and y are distinct vertices and that W_x^+ and W_y^+ both hold. Then W_x and W_y also hold. Our immediate aim is to show that the subgraphs $G_{\le t}(x)$ and $G_{\le t}(y)$ must be edge disjoint, i.e., they can meet only if $x_2 = y_2$, and then only at this one vertex. We shall write $W_x \star W_y$ for the event that W_x and W_y hold, and $G_{\le t}(x)$ and $G_{\le t}(y)$ are edge disjoint. In other words $W_x \star W_y = W_x \cap W_y \cap \{d(x,y) \ge 2t\}$. We define $W_x^+ \star W_y^+$ similarly, so $W_x^+ \star W_y^+ = W_x^+ \cap W_y^+ \cap \{d(x,y) \ge 2t + 2t_1\}$. (One must be careful here: when W_x holds, $G_{\le t}(x)$ is *not* a certificate for this event in the sense of the van den Berg–Kesten box product [5], i.e., specifying that this particular subgraph is present as an induced subgraph does not guarantee that W_x holds. To guarantee W_x, one must also certify that various edges are absent, from $G_{\le t-1}$ to vertices outside $G_{\le t}$; such certificates for W_x and W_y can never be disjoint, so we cannot simply apply Reimer's Theorem [37] to bound $\mathbb{P}(W_x \star W_y)$.)

Still assuming that x and y are distinct vertices such that W_x^+ and W_y^+ both hold, suppose first that y lies strictly inside $G_{\le t}(x)$, i.e., that $y \in V(G_{\le t}(x)) \setminus \{x_2\}$. As noted earlier, since W_x holds, $G_{\le t}(x)$ has diameter t, and this diameter is realized uniquely by x and x_2. Thus the vertex y_2, which is at distance t from y, must lie outside $G_{\le t}(x)$. But then the unique y-y_2 path P_y passes through x_2. Considering the vertex z where P_y first meets P_x, the strong wedge condition for x gives $d(y,z) < d(x,z)$. But the strong wedge condition for y gives $d(x,z) < d(y,z)$, a contradiction.

We may thus assume that y lies outside $V(G_{\le t}(x)) \setminus \{x_2\}$. Suppose now that y_2 also lies outside this set, and that $y_2 \ne x_2$. Since x_2 is a cutvertex, it follows that all of P_y is outside $V(G_{\le t}(x)) \setminus \{x_2\}$. If $G_{\le t}(x)$ and $G_{\le t}(y)$ meet, then, since x_2 is a cutvertex, x_2 must be a vertex of $G_{\le t}(y)$. Furthermore, since $G_{\le t}(y)$ consists of y_2 plus a component of $G \setminus \{y_2\}$, all of $G_{\le t}(x)$ lies in $G_{\le t}(y) \setminus \{y_2\}$. In particular $x \in G_{\le t}(y) \setminus \{y_2\}$ and we obtain a contradiction as above.

If $x_2 = y_2$, then each of $G_{\le t}(x)$ and $G_{\le t}(y)$ is formed by x_2 together with a tree component of $G - x_2$. Since each of x and y is the unique vertex at maximal distance from $x_2 = y_2$ within its tree, and $x \ne y$, these components are different, and so disjoint, so $W_x \star W_y$ holds.

We may thus assume that, if $W_x^+ \cap W_y^+$ holds but $W_x \star W_y$ fails, then y lies outside $V(G_{\le t}(x)) \setminus \{x_2\}$ but y_2 is inside. It is easy to check that in this case $G_{\le t}(x) \cup G_{\le t}(y)$ forms a component of G (and actually y_2 must lie on the path from x to x_2). Since W_x^+ holds, this component (the component of G containing x) has size at least $|\Gamma_{t+t_1}(x)| \ge \omega$; however, $W_x \cap W_y$ also holds, so it has size less than $2\omega/2$, a contradiction.

We have just shown that if $W_x^+ \cap W_y^+$ holds, then so does $W_x \star W_y$. It follows that either

$W_x^+ \star W_y^+$ holds, or $d(x_2, y_2) \le 2t_1$, implying $d(x, y) \le 2(t_0 - K + t_1)$. Thus, for the second moment,

$$\begin{aligned}
\mathbb{E}\, N^2 - \mathbb{E}\, N &= \sum_x \sum_{y \ne x} \mathbb{P}(W_x^+ \cap W_y^+) \\
&= \sum_x \sum_{y \ne x} \mathbb{P}(W_x^+ \cap W_y^+ \cap (W_x \star W_y)) \\
&\le \sum_x \sum_{y \ne x} \mathbb{P}(W_x^+ \star W_y^+) + \mathbb{P}\big(W_x \star W_y \cap \{d(x, y) \le 2(t_0 - K + t_1)\}\big).
\end{aligned}$$

For the first term, we have $\mathbb{P}(W_x^+ \star W_y^+) \sim \mathbb{P}(W_x^+)\mathbb{P}(W_y^+)$, since testing the event W_x^+ uses up at most ω^2 vertices, which does not affect the probability of W_y^+ significantly. (Alternatively, as before we may use Lemma 2.2 and a version of this lemma where we start at two vertices.)

To handle the second term, we use the following inequality, which we shall prove in a moment:

$$\mathbb{P}\big(W_x \star W_y \cap \{d(x, y) \le 2(t_0 - K) + t_3 - K'\}\big) = o(n^{-2}), \tag{2.18}$$

where $K' = 3K \log(1/\lambda_\star)/\log \lambda$ and $t_3 = \log n/\log \lambda$. Assuming this, using the fact that $2t_1 \le t_3 - K'$ for large n if K tends to infinity sufficiently slowly, we have $\mathbb{E}\, N^2 \le \mathbb{E}\, N + (1 + o(1))(\mathbb{E}\, N)^2 + n^2 o(n^{-2})$. Since $\mathbb{E}\, N \to \infty$, it follows that $\mathbb{E}\, N^2 \sim (\mathbb{E}\, N)^2$, so by Chebyshev's inequality N is concentrated about its mean, and in particular, $N \ge 2$ whp.

Set $d = 2(t_0 - K) + t_3 - K'$, so $d = \log n/\log \lambda + 2\log n/\log(1/\lambda_\star) - O(K)$. With K tending to infinity arbitrarily slowly, our aim in this subsection is to prove that $\mathrm{diam}(G) \ge d$ holds whp.

Let M be the number of pairs of distinct vertices x, y for which $W_x \star W_y \cap \{d(x, y) \le d\}$ holds. Using (2.18) again, we have $\mathbb{E}\, M = o(1)$, so $M = 0$ whp. Thus, whp, we have $N \ge 2$ and $M = 0$. Then there are distinct vertices x, y for which W_x^+ and W_y^+ hold. As shown above, it then follows that $W_x \star W_y$ holds. Since $M = 0$, we have $d(x, y) > d$. From the classical results of Erdős and Rényi [21], there is some constant $A > 0$ such that whp exactly one component of G, the 'giant' component, contains more than $A \log n$ vertices. Since (for n large) W_x^+ implies that x is in a component with at least $\omega > A \log n$ vertices, whp any pair x, y satisfying the conditions above lies in the giant component, so $d < d(x, y) < \infty$, and $\mathrm{diam}(G) > d$, as required.

It remains only to prove (2.18). To do so, we explore the neighbourhoods of a given pair x, y of vertices as usual, to test whether $W_x \star W_y$ holds. If so, the possible edges between the remaining vertices, including x_2 and y_2, have not yet been tested, so each is present with its original unconditional probability. Hence, given $W_x \star W_y$, summing over all possible paths we see that the probability that $d(x_2, y_2) \le \ell$ is at most

$$\sum_{k \le \ell} n^{k-1}(\lambda/n)^k = \sum_{k \le \ell} \lambda^k/n = O(\lambda^\ell/n)$$

and

$$\begin{aligned}\mathbb{P}\big(W_x \star W_y \cap \{d(x,y) \le 2(t_0-K)+t_3-K'\}\big) &= \mathbb{P}(W_x)\mathbb{P}(W_y)O(\lambda^{t_3-K'}/n)\\ &= O\big(\lambda_\star^{t_0-K}\lambda_\star^{t_0-K}\lambda^{-K'}\big)\\ &= O(1/n^2)(1/\lambda_\star)^{2K}\lambda^{-K'}\\ &= O(1/n^2)(1/\lambda_\star)^{2K-3K} = o(n^{-2}),\end{aligned}$$

as required.

Combining the lower bound on the diameter we have just proved, and the upper bound proved in Subsection 2.1, we obtain Theorem 1.1.

3 Average degree tending to infinity

In this section we shall prove Theorem 1.2. Throughout, when we consider $G(n, \lambda/n)$ we assume that $\lambda = \lambda(n) \to \infty$ with $\lambda \le n^{1/1000}$. For convenience, we always assume that λ is larger than some absolute constant λ_0, chosen so that the various statements 'provided λ is large enough' in what follows hold for $\lambda \ge \lambda_0$. With λ tending to infinity, some aspects of the proof become easier than the λ constant case, whilst some become more difficult.

We retain the same basic plan of attack as for the case of λ constant. One of the main problems is that we cannot simply work with the time that the neighbourhoods of a vertex take to reach a certain size ω, since the first neighbourhood larger than this may have size anywhere from ω to around $\lambda\omega$; this difference is too big for our later arguments. Instead we will look at the size of the neighbourhoods at a specific time. We could consider sizes in certain ranges, but it turns out that we can simply consider individual sizes, bounding the probability that a certain neighbourhood has exactly a certain size r. Roughly speaking, as in the previous section, the probability that the neighbourhoods of a vertex take a generations longer than usual to reach (or exceed) some given size turns out to be around $\lambda_\star^a$, where $\lambda_\star < 1$ is the dual branching process parameter, defined by $\lambda_\star e^{-\lambda_\star} = \lambda e^{-\lambda}$. This event corresponds to the (later) neighbourhoods being a factor of λ^a smaller than usual. So we study for *real* parameters a the probability that the neighbourhoods are λ^a smaller than usual, expressing this probability as a power of $\lambda_\star$.

Throughout this section it will be useful to bear in mind the asymptotic formula

$$\lambda_\star = \lambda e^{-\lambda} + O(\lambda^2 e^{-2\lambda}), \tag{3.1}$$

which follows easily from $\lambda_\star e^{-\lambda_\star} = \lambda e^{-\lambda}$ and $\lambda_\star < 1$. Note in particular that $\lambda_\star$ is asymptotically smaller than any constant negative power of λ.

3.1 Branching process preliminaries

We first give some lemmas describing the growth behaviour of the branching process.

Lemma 3.1. *Suppose $\lambda \ge 10$ and $0 < \delta \le 1/2$. Given that $|X_r| = k \ge 1$, with probability at least $1 - e^{-c\delta^2\lambda k}$ we have $|X_t|/(\lambda^{t-r}k) \in [1-\delta, 1+\delta]$ for all $t \ge r$, where $c > 0$ is an absolute constant.*

Proof. We may assume without loss of generality that $r = 0$. For $t \geq 0$, let $\rho_t = |X_{t+1}|/(\lambda|X_t|)$, and let E_t be the event that $|\rho_t - 1| > \delta/3^{t+1}$; it suffices to prove that $\mathbb{P}(\bigcup_t E_t) \leq e^{-c\delta^2\lambda k}$. Let F_t be the event that E_t holds but no E_s holds, $s < t$, so $\mathbb{P}(\bigcup_t E_t) = \sum \mathbb{P}(F_t)$. If no E_s holds for $s < t$, then $|X_t| \geq k\lambda^t \prod_{s<t}(1 - \delta/3^{s+1}) \geq k\lambda^t/10$. Turning to $|X_{t+1}|$, conditional on $|X_t|$, by Lemma 2.3 the probability that ρ_t lies outside $[1 - \delta/3^{t+1}, 1 + \delta/3^{t+1}]$ is at most $2\exp(-c_0\delta^2 9^{-t}\lambda|X_t|)$, for some $c_0 > 0$. Hence $\mathbb{P}(F_t) \leq 2\exp(-c_0\delta^2 9^{-t}\lambda^{t+1}k/10)$, and the result follows by summing this rapidly decreasing sequence. □

For $0 \leq a < 1$ define $g(a) = g(\lambda, a)$ by $\lambda_\star^{g(a)} = \mathbb{P}(Z \leq \lambda^{1-a})$, where Z has a Poisson distribution with mean λ. Thus $\lambda_\star^{g(a)}$ is the probability that Z is smaller than its mean by a factor of λ^a or more. Note that $g(a)$ is (weakly) increasing in a. Also, as $\lambda \to \infty$ we have $\mathbb{P}(Z \leq \lambda) \to 1/2$ and $\lambda_\star \to 0$, so $g(0) = o(1)$. A simple calculation shows that $g(a) = 1 - o(1)$ for any fixed $0 < a < 1$. Also, using (3.1) we have $\mathbb{P}(Z \leq 1) = (1+\lambda)e^{-\lambda} = \lambda_\star e^{-\lambda_\star}(1+1/\lambda) > \lambda_\star$ for large enough λ, and so

$$0 \leq g(a) < 1 \tag{3.2}$$

for all $0 \leq a < 1$.

Extend g to the real line by defining $g(x) = \lfloor x \rfloor + g(x - \lfloor x \rfloor)$; this gives an increasing function which, from (3.2), satisfies

$$\lfloor x \rfloor \leq g(x) \leq \lfloor x \rfloor + 1 \tag{3.3}$$

for all x. It is straightforward to check that for any constant $b \geq 3$, say, if n is large enough then

$$\lambda_\star^{g(a-\log b/\log\lambda)} \geq \lambda^{b/4}\lambda_\star^{g(a)} \tag{3.4}$$

holds for all a. Indeed, if $m \leq a - \log b/\log\lambda$, $a < m+1$ for some integer m, then (3.4) decodes to a statement of the form $\mathbb{P}(Z \leq bk) \geq \lambda^{b/4}\mathbb{P}(Z \leq k)$, where $1 \leq k \leq \lambda/b$; the inequality is easily verified by considering, for example, the ranges $k \geq \lambda/(10b)$, $\sqrt{\lambda} \leq k \leq \lambda/(10b)$, and $1 \leq k \leq \sqrt{\lambda}$. On the other hand, if $a - \log b/\log\lambda < m \leq a$ then it decodes to $\lambda_\star^{-1}\mathbb{P}(Z \leq kb/\lambda) \geq \lambda^{b/4}\mathbb{P}(Z \leq k)$, with $k < \lambda$ and $bk > \lambda$; this is easily verified by considering the cases $k \geq 0.9\lambda$ and $k < 0.9\lambda$, say.

We next give an analogue of the upper bound in Lemma 2.1; note that we do not round t_1 to an integer.

Lemma 3.2. *Suppose that $\omega \geq \lambda$ and that $t \geq 0$ is an integer. Then for λ at least some absolute constant, setting $t_1 = \log\omega/\log\lambda$ we have*

$$\mathbb{P}\big(0 < |X_t| < \omega/2\big) \leq 3\lambda_\star^{g(t-t_1)}.$$

Proof. Note first that if $t < t_1$, then $g(t - t_1) \leq 0$ by (3.3), so the result holds trivially. We may thus assume that $t \geq t_1$, so $t \geq \lceil t_1 \rceil$.

Case 1: $t \geq \lceil t_1 \rceil + 1$.

Similar to the proof of Lemma 2.1, set $r_t = \mathbb{P}(|X_t| < \omega/2 \mid |X_t| > 0)$. Then it suffices to show that $r_t \leq 3\lambda_\star^{g(t-t_1)}$. We shall show in a moment that if $t = \lceil t_1 \rceil + 1$, then

$$\mathbb{P}\big(0 < |X_t| < \omega/2\big) \leq 1.1\lambda_\star^{g(t-t_1)}. \tag{3.5}$$

Suppose for the moment that this holds. Then by monotonicity of g and the fact that $g(1) \geq 1$, and since $\mathbb{P}(|X_t| > 0) \sim 1$, for such t we have $r_t \leq 1.2\lambda_\star$ if λ is at least some (absolute) constant.

As noted in the proof of Lemma 2.1, the implicit constant in all $O(\cdot)$ notation leading to (2.4) may be taken to be absolute when $\lambda > 1$ is bounded away from 1, so this bound applies with λ growing as a function of n. In particular, from (2.4), we have for arbitrary $t \geq 1$

$$r_t \leq r_{t-1}(\lambda_\star + r_{t-1} + O(\lambda\lambda_\star^t)) = r_{t-1}(\lambda_\star + r_{t-1} + o(\lambda_\star^{t-1})), \tag{3.6}$$

using $\lambda\lambda_\star = o(1)$ for the last step, which follows from (3.1).

We may iterate (3.6), with $\lambda_\star$ sufficiently small in the following (as λ can be assumed large). Beginning with $t = \lceil t_1 \rceil + 1$, when $r_t \leq 1.2\lambda_\star$ from (3.5) and hence $r_{t+1} \leq 2.7\lambda_\star^2$ from (3.6), we see that that r_t decreases extremely rapidly: $r_t \leq 3\lambda_\star r_{t-1}$ for $t > \lceil t_1 \rceil + 1$. Feeding the resulting bound $r_{\lceil t_1 \rceil + k} \leq (3\lambda_\star)^k$, $k > 1$, back into (3.6), it follows that for $t > \lceil t_1 \rceil + 1$ we have $r_t \leq r_{t-1}\lambda_\star(1 + \varepsilon_t)$ where the first error term $\varepsilon_{\lceil t_1 \rceil + 2}$ is at most 1.3 and later ones decrease extremely rapidly. Since $\prod_t (1 + \varepsilon_t) \leq 2.4$ for λ large enough, the result for Case 1 now follows from (3.5).

It remains to prove (3.5). Assuming now that $t = \lceil t_1 \rceil + 1$, put $a = t - t_1 - 1$, so that $0 \leq a < 1$. We claim that

$$\mathbb{P}\big(0 < |X_2| \leq \lambda^{1-a}\big) \sim \lambda_\star^{1+g(a)} \tag{3.7}$$

and that

$$\mathbb{P}\left(|X_2| > \lambda^{1-a},\, |X_t| < \omega/2\right) = o(\lambda_\star^{1+g(a)}). \tag{3.8}$$

Since $1 + g(a) = g(t - t_1)$, these imply (3.5).

Note that $\mathbb{P}(|X_1| = 1) = \lambda e^{-\lambda} \sim \lambda_\star$, and the probability that subsequently $|X_2| \leq \lambda^{1-a}$ is $\lambda_\star^{g(a)}$ by definition of g. Thus,

$$\mathbb{P}\left(|X_1| = 1,\, |X_2| \leq \lambda^{1-a}\right) \sim \lambda_\star^{1+g(a)}.$$

On the other hand, conditioning on $|X_1| = k \geq 2$, the conditional distribution of $|X_2|$ is Poisson $\mathrm{Po}(k\lambda)$. Since $a \geq 0$, we may assume that $\lambda^{1-a} \leq k\lambda/2$, and it follows that there is an absolute constant $c_2 > 0$ such that for all $k \geq 2$, $\mathbb{P}(|X_2| \leq \lambda^{1-a} \mid |X_1| = k) < e^{-c_2 k\lambda}$. Fixing $k_0 > 3/c_2$, we have

$$\begin{aligned}\mathbb{P}\big(|X_1| > k_0,\, |X_2| \leq \lambda^{1-a}\big) &\leq \mathbb{P}\left(|X_2| \leq \lambda^{1-a} \;\middle|\; |X_1| > k_0\right) \\ &\leq e^{-3\lambda} = o(\lambda_\star^2) = o\big(\lambda_\star^{1+g(a)}\big).\end{aligned}$$

Turning to $2 \leq k \leq k_0$, we have $\mathbb{P}(|X_1| = k) = O(\lambda^{k-1}\lambda_\star)$. Suppose firstly that $g(a) < c_2$ as defined above. Then $\mathbb{P}\left((2 \leq |X_1| \leq k_0) \wedge |X_2| \leq \lambda^{1-a}\right) < O(\lambda^{k_0-1}\lambda_\star)e^{-2c_2\lambda} = \lambda_\star^{1+2c_2+o(1)} = o(\lambda_\star^{1+g(a)})$. So we may assume that $g(a) \geq c_2$. Then, noting that for $|X_2| \leq \lambda^{1-a}$ to hold each particle in X_1 must have at most λ^{1-a} children, we have

$$\mathbb{P}\big(|X_1| = k,\, |X_2| \leq \lambda^{1-a}\big) = O(\lambda^{k-1}\lambda_\star\lambda_\star^{kg(a)}) = o(\lambda_\star^{1+g(a)}),$$

since $\lambda\lambda_\star^{g(a)} = \lambda_\star^{g(a)-o(1)} = o(1)$. Putting the pieces together, we have established (3.7).

The proof of (3.8) is similar. Condition on $|X_2| = k$, where $k > \lambda^{1-a}$. In the event that $|X_t| < \omega/2$, the average number of descendants in X_t of a particle in X_2 is less than $\omega/(2\lambda^{1-a})$. However, we know that any one such particle expects $\lambda^{t-2} = \lambda^{t_1+a-1} = \omega/\lambda^{1-a}$ such descendants, and applying Lemma 3.1, we see that $\mathbb{P}(|X_t| < \omega/2 \mid |X_2| = k) \leq e^{-c_3 k\lambda}$ for some $c_3 > 0$. Arguing as for the proof of (3.7), there exists k_1 such that $\mathbb{P}(|X_t| < \omega/2 \mid |X_2| > k_1) = o\big(\lambda_\star^{1+g(a)}\big)$.

We are left with showing (3.8) in the case that $\lambda^{1-a} < |X_2| \leq k_1$, which requires $|X_2| \geq 2$ since $a < 1$. It is easy to see that $\mathbb{P}(|X_2| \leq k_1) = \Theta(\lambda^{k_1-1}\lambda_\star^2) = \lambda_\star^{2-o(1)}$. Conditional upon this, for $|X_t| < \omega/2$ to hold at least one particle in X_2 must have at most half its expected number of descendants in X_t. By Lemma 3.1 and the union bound, the conditional probability of this is at most $k_1 e^{-c_3\lambda} = o(\lambda_\star^{c_3/2})$. Hence

$$\mathbb{P}(|X_2| \leq k_1)\,\mathbb{P}\big(|X_t| < \omega/2 \;\big|\; \lambda^{1-a} < |X_2| \leq k_1\big) = o(\lambda_\star^{2-o(1)+c_3/2}) = o(\lambda_\star^2)$$

and we have (3.8) since $1 + g(a) \leq 2$.

Case 2: $t = \lceil t_1 \rceil$.

In this case, setting $a = t - t_1 \in [0, 1)$, we have $\mathbb{P}(0 < |X_1| \leq \lambda^{1-a}) < \lambda_\star^{g(a)}$ by definition of g. Using this in place of (3.7), it suffices to show that $\mathbb{P}\big(|X_1| > \lambda^{1-a},\, |X_t| < \omega/2\big) = o(\lambda_\star^{g(a)})$; the proof is identical to that of (3.8), apart from the notation. □

We next turn to the analogue of the lower bound in Lemma 2.1; as there, we bound the probability of a rather specific event involving extra conditions that will be needed in our lower bound on the diameter.

We say that the branching process (X_t) satisfies the *diamond condition to generation r* if $|X_1| = 1$, there is a unique particle x_r in X_r, and the chain $x_0x_1\cdots x_r$ of ancestors of x_r is such that any 'side branches' starting from x_i die within $\min\{i, r-i\}$ further generations. For $r = 0$ we interpret the diamond condition to hold vacuously.

Lemma 3.3. *Let $t' \geq 0$ be an integer, and $0 \leq a < 1$ a real number. Let F_0 be the event that $|X_{t'}| = 1$ and the diamond condition holds to generation t', and let F_1 be the event that $|X_{t'+1}| \leq \lambda^{1-a}$. Then as $\lambda \to \infty$ we have*

$$\mathbb{P}(F_0 \cap F_1) \sim \lambda_\star^{g(t'+a)},$$

uniformly in t' and a. Furthermore, provided λ is at least some absolute constant, then for any $\omega \geq \lambda$ and $t \geq t_1 = \log\omega/\log\lambda$ there is a ρ with $\omega/3 \leq \rho \leq 2\omega$ such that

$$\mathbb{P}\big(F_0 \cap F_1 \cap \{|X_t| = \rho\}\big) \geq \lambda_\star^{g(t-t_1)}/(3\lambda\omega),$$

where F_0 and F_1 are defined as above with t' and a the integer and fractional parts of $t - t_1$, respectively.

Note that t_1 is *not* rounded to an integer. Essentially, the lemma says that the probability that $\mathfrak{X}_\lambda$ survives but (after some time) is a factor λ^x smaller than it should be is around $\lambda_\star^{g(x)}$. The second statement shows that there is some specific size in a suitable range such that the probability of hitting exactly this size is not much smaller.

Proof. The event F_0 is exactly the event W^0 referred to in (2.14), but with $t_0 - K$ replaced by t'. Using (2.15) to translate (2.16) back in terms of W^0, we have $\lambda_\star^{t'} \geq \mathbb{P}(F_0) \geq \lambda_\star^{t'} d_1 d_1^2 d_2^2 d_3^2 \cdots$, where $d_t = \mathbb{P}(|X_t^-| = 0)$ is at least $1 - \lambda_\star^t$ from (2.17). Since $\lambda_\star \to 0$, it follows that $\mathbb{P}(F_0) \sim \lambda_\star^{t'}$.

Conditioning on F_0 says nothing about the descendants of the unique particle z in $X_{t'}$, so if Z is Poisson with mean λ then

$$\mathbb{P}(F_1 \mid F_0) = \mathbb{P}(Z \leq \lambda^{1-a}) = \lambda_\star^{g(a)},$$

where the last step is the definition of $g(a)$. Since $\lambda_\star^{t'} \lambda_\star^{g(a)} = \lambda_\star^{g(t'+a)}$, this proves the first statement.

Turning to the second statement, suppose that $\omega \geq \lambda$ and $t \geq t_1 = \log \omega / \log \lambda$. Let $t' = \lfloor t - t_1 \rfloor$ and $a = t - t_1 - t'$. Let F_1' be the event that $|X_{t'+1}| = \lfloor \lambda^{1-a} \rfloor$. Noting that $x = \lfloor \lambda^{1-a} \rfloor$ is the most likely value x of Z with $x \leq \lambda^{1-a}$, arguing as above we have $\mathbb{P}(F_1' \mid F_0) \geq \lambda_\star^{g(a)}/\lambda$, and hence $\mathbb{P}(F_0 \cap F_1') \geq \lambda_\star^{g(t'+a)}/(2\lambda)$, provided λ is large enough.

Noting that $t - t' \geq t_1 \geq 1$, let F_2 be the event that the ratio $|X_t|/(\lambda^{t-t'-1}|X_{t'+1}|)$ is between $9/10$ and $11/10$. Then by Lemma 3.1 we have $\mathbb{P}(F_2 \mid F_0 \cap F_1') \to 1$, so

$$\mathbb{P}(F_0 \cap F_1' \cap F_2) \geq \lambda_\star^{g(t'+a)}/(3\lambda).$$

Noting that $\lambda^{1-a}\lambda^{t-t'-1} = \lambda^{t-(t'+a)} = \lambda^{t-(t-t_1)} = \lambda^{t_1} = \omega$, and that $\lfloor \lambda^{1-a} \rfloor \geq \lambda^{1-a}/2$, if $F_0 \cap F_1' \cap F_2$ holds then so does the event $E_\rho = F_0 \cap F_1 \cap \{|X_t| = \rho\}$ for some ρ between $9\omega/20$ and $11\omega/10$. So there is some ρ in this range for which $\mathbb{P}(E_\rho) \geq \lambda_\star^{g(t-t_1)}/(3\lambda\omega)$, as required. □

We also need an analogue of Lemma 2.2 without the assumption that λ is fixed.

Lemma 3.4. *Let $\lambda = \lambda(n)$ satisfy $\lambda \leq n^{1/10}$. Then the estimates*

$$\mathbb{P}\big(G_{\leq t}(x) \cong T\big) \sim \mathbb{P}\big(G^0_{\leq t}(x) \cong T\big) \sim \mathbb{P}\big(X_{\leq t} \cong T\big)$$

hold uniformly over rooted trees T with $|T| \leq n^{2/5}$, where t is the height of T.

Proof. The proof is essentially identical to that of Lemma 2.2. Indeed, the estimate (2.7) is valid assuming only that $|T|$, $\lambda \leq n/2$, say; under our present assumptions this estimate is $\exp(O(n^{-1/5} + n^{-1/10} + n^{-2/5})) = 1 + o(1)$. As before, the result for $G_{\leq t}(x)$ follows, now noting that the expected number of untested edges present is $O(\lambda|T|^2/n) = o(1)$. □

3.2 Neighbourhoods in the graph and how they meet

Our immediate plan is to examine those vertices for which the breadth first search procedure takes an unusually long time to reach a 'large' number of vertices. For convenience we choose 'large' to mean around λ^{10}; since we assume $\lambda < n^{1/1000}$, say, λ^{10} is much less than $n^{1/4}$. We do not attempt to optimise the power of n giving the upper bound on λ. We first work towards a lemma that gives asymptotically the probability that two neighbourhoods of size at least $\lambda^9/4$ have a certain distance between them. This will be needed in particular later when we make variance calculations in using the second moment method.

As in Section 2, set

$$t_0 = \lfloor \log n / \log(1/\lambda_\star) \rfloor.$$

For $r \geq 1$, let S_r be the set of vertices x in the random graph with $|\Gamma_{t_0+10}(x)| = r$.

Lemmas 3.3 and 3.2, in conjunction with Lemma 3.4, give some information on the expected size of S_r, or, more precisely, on the size of unions of such sets over r in suitable ranges, though (as will be apparent in the argument below) the upper and lower bounds given by the lemmas can differ by a factor of λ or more.

We first consider the branching process. For $r \geq \lambda$, setting $\omega = 3r > 2r$ in Lemma 3.2 gives

$$\mathbb{P}(0 < |X_t| \leq r) < 3\lambda_\star^{g(t-\log(3r)/\log\lambda)}. \tag{3.9}$$

Although we shall not use it, let us note that in the other direction, with α constant and λ large enough, applying Lemma 3.3 with $\omega = \frac{1}{2}\alpha r$ and then Lemma 3.1 gives

$$\mathbb{P}(0 < |X_t| < \alpha r) \geq \frac{1}{2}\lambda_\star^{g(t-\log(\alpha r/2)/\log\lambda)}, \tag{3.10}$$

provided the argument of g is greater than 0.

We will transfer the bounds above to the random graph using Lemma 3.4, which shows that the corresponding random graph and branching process events have asymptotically the same probability, provided there are not too many vertices close to x, so that the trees used in applying Lemma 3.4 are not too large. First, define $\Gamma_{\leq i}(x) = \bigcup_{j=0}^{i} \Gamma_j(x)$.

Let B_1 be the ('bad') set of vertices x such that $|\Gamma_{\leq t_0+10}(x)| > n^{1/4}$. From (3.1) we have $\log(1/\lambda_\star) \sim \lambda$, which is much larger than $\log\lambda$, so $\lambda^{t_0+10} = n^{o(1)}\lambda^{10} \leq n^{1/8}$ if n is large enough. For fixed $k \geq 1$, the number of unlabelled rooted trees of height t with exactly k (non-root) leaves, all at distance t from the root, can be estimated by adding paths to leaves one at a time, giving the crude upper bound $O(1)(t+1)^{k-1}$. It is thus easily seen that for fixed k we have $\mathbb{E}\,|\Gamma_{\leq t}(x)|^k \leq O(1)(t+1)^{k-1}\lambda^{tk}$. With $t = t_0 + 10 = O(\log n)$ and $k = 20$, this gives $\mathbb{E}\,|\Gamma_{\leq t_0+10}(x)|^{20} \leq (\log n)^{O(1)} n^{2.5} = o(n^3)$. Thus Markov's inequality gives

$$\mathbb{E}\,|B_1| \leq n\mathbb{P}\big(|\Gamma_{\leq t_0+10}(x)|^{20} \geq n^5\big) = o(n^{-1}). \tag{3.11}$$

A similar calculation shows that

$$\mathbb{P}(B_1^0) = o(n^{-2}), \tag{3.12}$$

where B_1^0 is the branching process event corresponding to B_1.

Define $\widetilde{\mu}_r$ to be the expected number of vertices x in $S_r \setminus B_1$. Applying Lemma 3.4 to each relevant tree, which has at most $n^{1/4}$ vertices by definition, and summing over x, we have $\widetilde{\mu}_r \leq (1+o(1))n\mathbb{P}(|X_{t_0+10}| = r)$, so by (3.9) we have

$$\begin{aligned} \widetilde{\mu}_r &< n(3+o(1))\lambda_\star^{g(t_0+10-\log(3r)/\log\lambda)} \\ &< (3+o(1))\lambda_\star^{8-\log(3r)/\log\lambda} \end{aligned} \tag{3.13}$$

using (3.3).

We can similarly see easily that the union of the sets $S_r \setminus B_1$ over all $r < \lambda^9/4$ is whp empty: setting $\omega = \lambda^9/2$ in Lemma 3.2 gives

$$\begin{aligned}\mathbb{P}\big(0 < |X_{t_0+10}| < \lambda^9/4\big) &\le 3\lambda_\star^{g(t_0+1+\log 2/\log\lambda)} \\ &\le \frac{3}{n}\lambda_\star^{g(\log 2/\log\lambda)}\end{aligned}$$

which is $3n^{-1}\mathbb{P}(\mathrm{Po}(\lambda) \le \lambda/2) = o(1/n)$. Hence, using Lemma 3.4 again, $\sum_{r<\lambda^9/4}\widetilde{\mu}_r = o(1)$. Since $\mathbb{E}\,|B_1| = o(1)$, it follows that

$$\bigcup_{1\le r\le\lambda^9/4} S_r = \emptyset \text{ whp.} \tag{3.14}$$

Thus, we are interested in S_r for $r \ge \lambda^9/4$.

An annoying feature of the present situation is that with some small probability, the size of $\Gamma_i(x)$ can 'misbehave' for $i > t_0 + 10$. Although there are whp no vertices for which this happens to a significant extent, we need to treat these vertices separately. Define $\ell(r) = \max\{0, \lceil\log\big(2(\log^6 n)/r\big)/\log\lambda\rceil\}$, so $\ell(r) \ge 0$ is minimal subject to $r\lambda^{\ell(r)} \ge 2\log^6 n$. Let B_2 be the set of 'bad' vertices x with the property that $|\Gamma_{t_0+10}(x)| \ge \lambda^9/4$, and the 'ratio error'

$$\frac{|\Gamma_{t_0+10+i}(x)|}{\lambda^i|\Gamma_{t_0+10}(x)|} - 1 \tag{3.15}$$

has absolute value at least λ^{-2} for some $0 \le i \le \ell(r)$. We write $V_0 = V \setminus (B_1 \cup B_2)$ for the set of 'good' vertices.

Lemma 3.5. *(a)* $\mathbb{E}\,|B_2| = o(1)$*.*

(b) Conditional on two vertices x and y being in $S_{r_1} \cap V_0$ and $S_{r_2} \cap V_0$ respectively, where $\lambda^9/4 \le r_i \le n^{1/4}$ ($i = 1$ and 2), and additionally conditional on $d(x,y) > 2t_0 + 20 + \ell(r_1) + \ell(r_2)$, we have

$$\mathbb{P}\left(d(x,y) > 2t_0 + 20 + k\right) = \exp\left(-\frac{r_1r_2}{n}\big(1 + O(\lambda^{-2})\big)\sum_{i=1}^{k}\lambda^i\right) + o(n^{-3})$$

for all $k > \ell(r_1) + \ell(r_2)$, where the constant implicit in the $O(\cdot)$ terms is uniform over all such r_1, r_2 and k.

Proof. As in the proof of Lemma 2.4, we explore the successive neighbourhoods of a vertex. If a_t denotes $|\Gamma_t(x)|$, then conditional on the part of the graph explored up to this point, and assuming that it contains at most $n^{2/3}$ vertices, a_{t+1} is distributed as binomial with parameters $n - O(n^{2/3})$ and $p = (\lambda a_t/n)(1 + O(\lambda a_t/n))$. The mean is $\lambda a_t(1 + O(n^{-1/4}))$, so by Lemma 2.3 (a Chernoff bound) we have

$$\mathbb{P}(|a_{t+1} - \lambda a_t| \le (\lambda a_t)^{3/4}) = 1 - e^{-\Omega(\sqrt{\lambda a_t})}. \tag{3.16}$$

To prove (a), in view of (3.11) we only need to show that $\mathbb{E}\,|B_2 \setminus B_1| = o(1)$. First explore the successive neighbourhoods of any vertex x up to $\Gamma_{t_0+10}(x)$. If the cardinality of this set, a_{t_0+10}, is less than $\lambda^9/4$ or greater than $2\log^6 n$, or if $|\Gamma_{\le t_0+10}(x)| > n^{1/4}$, then x is certainly not in $B_2 \setminus B_1$. Condition on the exploration so far assuming that none of these events hold, and that $a_{t_0+10} = r$, so $\lambda^9/4 \le r \le 2\log^6 n$. Next, continue exploring a further $\ell(r)$ steps. Provided the event in the left side of (3.16) holds at each exploration step, the 'relative error' $|a_{t+1}/\lambda a_t - 1|$ is at most $(\lambda a_t)^{-1/4}$. In this case,

$$\left|\frac{a_{t_0+10+i}}{\lambda^i a_{t_0+10}} - 1\right| < 2(\lambda a_{t_0+10})^{-1/4},$$

which is less than λ^{-2} since $a_{t_0+10} \ge \lambda^9/4$. This implies $x \notin B_2$. On the other hand, the probability that the event in the left side of (3.16) fails to hold for at least one of the relevant t is at most $e^{-\Omega\left(\sqrt{\lambda r}\right)}$, which is $\lambda_\star^{\Omega\left(\sqrt{r/\lambda}\right)}$ since $\lambda_\star > e^{-\lambda}$. The expected number of vertices $x \notin B_1$ with $a_{t_0+10} = r$ is $O(\lambda_\star^{8-\log(3r)/\log\lambda})$ by (3.13). Multiplying these bounds together and summing over $r \ge \lambda^9/4$ gives $o(1)$, that is, $\mathbb{E}\,|B_2 \setminus B_1| = o(1)$, as required.

We turn to (b). Let $\lambda^9/4 \le r_i \le n^{1/4}$ ($i = 1$ and 2). Take any vertices x and y, and explore the successive neighbourhoods of each up to distance $t_0+10+\ell(r_1)$ and $t_0+10+\ell(r_2)$ respectively. At this point, it is revealed whether these neighbourhoods are all disjoint, which is equivalent to $d(x,y) > 2t_0 + 20 + \ell(r_1) + \ell(r_2)$, and also (recalling that $V_0 = V \setminus (B_1 \cup B_2)$) whether $x \in V_0$ and $y \in V_0$. Condition on the event that all three of these hold. It follows from $x \notin B_2$ that $|\Gamma_{t_0+10+i}(x)| = r_1(1 + O(\lambda^{-2}))\lambda^i$ for $0 \le i \le \ell(r_1)$, and similarly for y.

We next explore the further neighbourhoods of x and y, each time choosing the smaller of the two for further exposure, until one of them has reached cardinality at least $n^{3/5}$, or until they meet, whichever happens first. Note that for all r we have $r\lambda^{\ell(r)} \ge 2\log^6 n$ by the definition of ℓ. Since $x \notin B_2$, using the 'error ratio' property in the definition of B_2 (see (3.15)) it follows that $|\Gamma_{t_0+10+\ell(r_1)}(x)| \ge \log^6 n$, and similarly for y. So, by applying (3.16) and conditioning on non-failure at each step, we conclude that with probability at least $1 - o(n^{-3})$, at each step

$$|\Gamma_{t_0+10+k}(x)| = r_1\lambda^k(1 + O(\lambda^{-2})),$$

and similarly for y. So we may assume this is the case each time. From this, when the sum of the two distances is $2t_0 + 20 + k - 1$, the product of the sizes of the neighbourhoods is $r_1 r_2 \lambda^{k-1}(1 + O(\lambda^{-2}))$, and hence the probability of not joining in the next step is

$$\exp\big(-r_1 r_2 \lambda^{k-1}(1 + O(\lambda^{-2}))\lambda/n\big).$$

The result follows, as long as the probability that they do not meet by the time that one of the neighbourhoods has reached size $n^{3/5}$ is bounded above by $o(n^{-3})$. This must be the case since on the previous step, the neighbourhood that was extended must have had size at least $n^{3/5}/\lambda(1 + o(1))$, so the product of sizes on the previous step must have been at least $n^{6/5}/\lambda^2(1 + o(1))$ which is at least $n^{11/10}$ as $\lambda < n^{1/1000}$. Thus the probability of not joining on the last step was at most $\exp(-\lambda n^{1/10}(1 + o(1)) = o(n^{-3})$. □

We now turn to the proof of Theorem 1.2.

Proof of Theorem 1.2. Recall that $\lambda = \lambda(n)$ is some given function of n satisfying $\lambda \to \infty$ and $\lambda \le n^{1/1000}$. All limits are as $n \to \infty$, or, equivalently, as $\lambda \to \infty$. As usual, all inequalities we claim are required to hold only if n (or λ) is sufficiently large.

Our first aim is to estimate the probability of the event conditioned on in Lemma 3.5(b). Let $\widehat{P}_r$ denote the probability that a given vertex is in $V_0 \cap S_r$, and $\widehat{P}_{r_1,r_2}$ the probability that a given pair of distinct vertices x and y satisfy $x \in V_0 \cap S_{r_1}$, $y \in V_0 \cap S_{r_2}$, and $d(x,y) > 2t_0 + 20 + \ell(r_1) + \ell(r_2)$. Note that $x \in V_0 \cap S_{r_1}$ iff the set of vertices at distance at most $t_0 + 10 + \ell(r_1)$ from x forms one of a specific set of graphs with less than $n^{1/4} + O(\lambda(\log n)^6) = o(n^{1/3})$ vertices, and $\widehat{P}_{r_1,r_2}$ counts configurations in which the explorations from x and y are disjoint. Since each exploration 'uses up' $o(n^{1/3})$ vertices, it is easy to see (for example using a version of Lemma 3.4 starting with two vertices) that

$$\widehat{P}_{r_1,r_2} \sim \widehat{P}_{r_1}\widehat{P}_{r_2}. \tag{3.17}$$

For any r, let $\widehat{\mu}_r$ denote $n\widehat{P}_r$, the expected size of $V_0 \cap S_r$; recall that $\widetilde{\mu}_r = \mathbb{E}\,|S_r \setminus B_1|$, so $\widehat{\mu}_r = \widetilde{\mu}_r + o(1)$ by Lemma 3.5(a). Also, for integer $k \ge 1$ define $\widehat{\mu}(r_1, r_2, k)$ to be the expected number of ordered pairs (x, y) of vertices with $x \in V_0 \cap S_{r_1}$, $y \in V_0 \cap S_{r_2}$, and $d(x,y) > 2(t_0 + 10) + k$. Since $V_0 = V$ whp, the number of such pairs essentially determines the diameter. From the above observations and Lemma 3.5(b),

$$\widehat{\mu}(r_1, r_2, k) \sim \widehat{\mu}_{r_1}\widehat{\mu}_{r_2}\left(\exp\Big(-r_1r_2(1+O(\lambda^{-2}))\sum_{i=1}^{k}\lambda^i/n\Big) + o(n^{-3})\right) \tag{3.18}$$

provided $k > \ell(r_1) + \ell(r_2)$ and r_1 and r_2 satisfy the constraints of Lemma 3.5. Note that we shall consider values of k that are at least $\log n/\log\lambda - 30$, which is larger than $2\ell(r)$ for any $r > 0$.

Define $\mu_r = n\,\mathbb{P}(|X_{t_0+10}| = r)$, which we shall analyse using Lemmas 3.3 and 3.2. We claim that

$$\widehat{\mu}_r \le \widetilde{\mu}_r \le \mu_r(1 + o(1)); \tag{3.19}$$

indeed, the first inequality holds by definition. If $r > n^{1/4}$ then $\widetilde{\mu}_r = 0$; otherwise the second inequality follows from Lemma 3.4, summing over the possible neighbourhoods of x. In the other direction, although we shall not these bounds, note that for $r \le n^{1/4}$ we have

$$\widehat{\mu}_r \ge \mu_r(1 + o(1)) + o(1),$$

since $\widehat{\mu}_r \ge \widetilde{\mu}_r + o(1)$ by Lemma 3.5(a), and $\widetilde{\mu}_r \ge \mu_r(1+o(1)) + o(1)$ from Lemma 3.4, together with (3.12).

We next show that vertices in sets $V_0 \cap S_r$ with $r > \lambda^{13.1}$ will not determine the diameter of the graph, for the reason that they join too quickly to all vertices under consideration: we claim that whp all such vertices have distance at most $\log n/\log\lambda + 2t_0 - 0.05$ from all other vertices; we will see later that whp the diameter is greater than this. To establish this claim, without loss of generality consider only $r_1 \ge \lambda^{13.1}$ and $r_2 \ge \lambda^9/4$. Note that the conditions on the r_i in Lemma 3.5(b) are so restrictive because it aims for a fairly accurate asymptotic estimate. In this case we only need to observe that if $x \in V_0 \cap S_r$ for $r = r_1$ or r_2, by definition of B_2, $|\Gamma_{t_0+10+i}(x)| \sim \lambda^i r$ until the neighbourhoods reach size at least $(\log n)^6$ (which they may do at $i = 0$), and for larger neighbourhoods up to size $n^{2/3}$, (3.16) provides the same relation with probability at least $1 - e^{-\Omega(\log^3 n)} = 1 - o(n^{-5})$. Summing over all $O(n^2)$ pairs of vertices x and y gives

$$\widehat{\mu}(r_1, r_2, k) \leq \widehat{\mu}_{r_1}\widehat{\mu}_{r_2} \exp\Big(- (1 + o(1)) r_1 r_2 \lambda^k / n\Big) + o(n^{-3}), \tag{3.20}$$

which is similar to (3.18) but does not have the same restrictions on r_1 and r_2. For $k = \lfloor \log n / \log \lambda - 20.05 \rfloor$ we have $r_1 r_2 \lambda^k / n > (r_1/\lambda^{13.1})(4r_2/\lambda^9)\lambda^{1.04}$. Now (3.19) and (3.13), together with $\lambda_\star = e^{-\lambda + o(\lambda)}$ (see (3.1)), give

$$\widehat{\mu}_r = O(1) \exp\big((1 + o(1))\lambda(\log(3r)/\log\lambda - 8)\big).$$

Summing the resulting bound on $\widehat{\mu}(r_1, r_2, k)$ over all $r_1 \geq \lambda^{13.1}$ and $r_2 \geq \lambda^9/4$ gives $o(1)$, as required to establish the claim. (The key observation is that when r_1 and r_2 take their minimum values, we have $\widehat{\mu}_{r_1}\widehat{\mu}_{r_2} = \exp(O(\lambda))$, while the exponential factor in (3.20) is at most $\exp(-\lambda^{1.04})$. When r_1 and r_2 increase, so does $\widehat{\mu}_{r_1}\widehat{\mu}_{r_2}$, but the exponential factor decreases more than fast enough to compensate.)

Recalling (3.14), let R be the set of indices r, $\lambda^9/4 \leq r \leq \lambda^{13.1}$, for which $\widehat{\mu}_r > \lambda^{-14}$. Then, by the union bound, the expected number of vertices in all sets $V_0 \cap S_r$ with r in this range but not in R is $o(1)$, i.e., there are whp no such vertices. Since $V_0 = V$ whp, using the observation above about sets S_r with $r > \lambda^{13.1}$ and (3.14), we have shown that

$$\operatorname{diam}(G) = \max_{(r_1, r_2) \in R^2} \max\{d(x, y) : x \in V_0 \cap S_{r_1},\ y \in V_0 \cap S_{r_2}\} \text{ whp.} \tag{3.21}$$

It only remains to examine r_1 and r_2 in R. Note that if $r \in R$ then $r \leq \lambda^{13.1}$, so from (3.13) and (3.19) we have

$$\widehat{\mu}_r < \lambda_\star^{-6} < e^{6\lambda}. \tag{3.22}$$

Let $k_0(r_1, r_2)$ denote the maximum k such that $\widehat{\mu}(r_1, r_2, k) > \lambda^{-27}$. (This number k_0 depends on n.) Then $\widehat{\mu}(r_1, r_2, k_0(r_1, r_2) + 1) \leq \lambda^{-27}$. Let $k_{\max}$ be the maximum value of k_0 over all pairs (r_1, r_2) in R^2. From (3.20), (3.22) and the definition of R, it is easy to check that $k_{\max} = \log n / \log \lambda + O(1)$. Setting $f(n, \lambda) = 2(t_0 + 10) + k_{\max}$, to prove the first part of Theorem 1.2 we shall show that the diameter is whp either $f(n, \lambda)$ or $f(n, \lambda) + 1$. Since $|R^2| = O(\lambda^{26.2})$, by the union bound, the expected number of pairs of vertices x and y counted in (3.21) at distance greater than $f(n, \lambda) + 1$ is $o(1)$. Thus $\operatorname{diam}(G) \leq f(n, \lambda) + 1$ holds whp.

To see that the diameter is whp at least $f(n, \lambda) = 2(t_0 + 10) + k_{\max}$ we shall look for vertices at this distance in suitable sets S_{r_i}. Choose (r_1, r_2) in R^2 with $k_0(r_1, r_2) = k_{\max}$. Note that $\widehat{\mu}(r_1, r_2, k_{\max}) > \lambda^{-27}$. That is, from (3.18),

$$\widehat{\mu}_{r_1}\widehat{\mu}_{r_2} \exp\Big(- r_1 r_2 (1 + O(\lambda^{-2})) \sum_{i=1}^{k_{\max}} \lambda^i / n\Big) + o(\widehat{\mu}_{r_1}\widehat{\mu}_{r_2}/n^3) > (1 + o(1))\lambda^{-27}.$$

By definition $\widehat{\mu}_{r_i} \leq n$, and $n^{-1} = o(\lambda^{-27})$, so

$$\widehat{\mu}_{r_1}\widehat{\mu}_{r_2} \exp\Big(- r_1 r_2 (1 + O(\lambda^{-2})) \sum_{i=1}^{k_{\max}} \lambda^i / n\Big) > \lambda^{-27}(1 + o(1)).$$

Using (3.22) for $r = r_1$ and $r = r_2$, it follows that

$$\exp\Big(-r_1 r_2 (1 + O(\lambda^{-2})) \sum_{i=1}^{k_{\max}} \lambda^i/n \Big) > \lambda^{-28} e^{-12\lambda} > e^{-13\lambda},$$

if n is large enough. Taking logs and stopping the sum one step earlier, this gives

$$-r_1 r_2 (1 + O(\lambda^{-2})) \sum_{i=1}^{k_{\max}-1} \lambda^i/n > -13. \tag{3.23}$$

Hence, by Lemma 3.5(b), vertices x and y whose $(t_0 + 10)$-neighbourhoods have sizes r_1 and r_2 respectively have a significant (at least $e^{-13} + o(1)$) probability of being at distance at least $2t_0 + 20 + k_{\max}$. Although by design we expect a large number of pairs of such vertices x and y, it is still possible that the expected number of possibilities for either x or y goes to 0! Our strategy is to consider vertices with $|\Gamma_{t_0+10}(\cdot)|$ around $2000r_i$, say, and show that this gives us many vertices x and y to work with. We also impose certain extra conditions on their neighbourhoods needed later.

For $i = 1, 2$, since r_i is in R, we have $\widehat{\mu}_{r_i} > \lambda^{-14}$. Now (3.19) shows that $\mathbb{P}(|X_{t_0+10}| = r) = \mu_{r_i}/n > (1 + o(1))\lambda^{-14}/n$. By (3.9) it follows that

$$\lambda_\star^{g(t_0+10-\log(3r_i)/\log\lambda)} > (1/3 + o(1))\lambda^{-14}/n. \tag{3.24}$$

Let $\omega_i = 1000r_i \le \lambda^{14}$. By (3.4) and (3.24) we have

$$\lambda_\star^{g(t_0+10-\log\omega_i/\log\lambda)} > (1/3 + o(1))\lambda^{250/3}\lambda^{-14}/n \ge \lambda^{40}/n, \tag{3.25}$$

if λ is large enough. For $i = 1, 2$, applying Lemma 3.3 with $\omega = \omega_i$ and $t = t_0 + 10$, there is some ρ_i with $\omega_i/3 \le \rho_i \le 2\omega_i$ such that the event $F_0 \cap F_1 \cap \{|X_{t_0+10}| = \rho_i\}$ described in Lemma 3.3 has probability π_i satisfying

$$\pi_i \ge \lambda_\star^{g(t_0+10-\log\omega_i/\log\lambda)}/(3\lambda\omega_i) \ge \lambda^{39}/(3n\omega_i) \ge \lambda^{25}/n, \tag{3.26}$$

using (3.25). Let $\widetilde{E}_{\rho_i}(x)$ denote the event that $x \notin B_1$ and the neighbourhoods of x up to distance $t_0 + 10$ form a tree that, when viewed as a branching process, satisfies the conditions $F_0 \cap F_1 \cap \{|X_{t_0+10}| = \rho_i\}$. By (3.12) and Lemma 3.4, we have $\mathbb{P}(\widetilde{E}_{\rho_i}(x)) \sim \pi_i + o(n^{-2})$. Since π_i is much larger than n^{-2}, it follows that $\mathbb{P}(\widetilde{E}_{\rho_i}(x)) \sim \pi_i \ge \lambda^{25}/n$.

Let $E_{\rho_i}(x)$ be the event that $\widetilde{E}_{\rho_i}(x)$ holds and $x \in V_0$, so the only additional condition is that $x \notin B_2$. Let $P_i = \mathbb{P}(E_{\rho_i}(x))$. Since $\mathbb{P}(x \in B_2) = o(1/n)$, we have

$$P_i \sim \mathbb{P}(\widetilde{E}_{\rho_i}(x)) \sim \pi_i \ge \lambda^{25}/n. \tag{3.27}$$

Note also for later that, writing t_i and a_i for the integer and fractional parts of $t_0 + 10 - \log\omega_i/\log\lambda$, and writing $F_0(x)$ for the event that the neighbourhoods of x satisfy the diamond condition to distance t_i (corresponding to F_0 in Lemma 3.3), then starting from the first statement of Lemma 3.3 and arguing as above we have

$$\mathbb{P}\big(F_0(x) \cap \{|\Gamma_{t_i+1}(x)| \le \lambda^{1-a_i}\}\big) \sim \lambda_\star^{g(t_0+10-\log\omega_i/\log\lambda)}.$$

Using the first inequality in (3.26) it follows that

$$P_i \geq \lambda^{-15}\mathbb{P}\big(F_0(x) \cap \{|\Gamma_{t_i+1}(x)| \leq \lambda^{1-a_i}\}\big). \tag{3.28}$$

In other words, once we have explored the neighbourhoods to the 'branching vertex' x_0, and found few neighbours in the next step, it is not that unlikely that $E_{\rho_i}(x)$ holds.

Given distinct vertices x and y, as in (3.17) the probability that $E_{\rho_1}(x)$ and $E_{\rho_2}(y)$ hold and $d(x,y) > 2(t_0+10)+\ell(\rho_1)+\ell(\rho_2)$ is $(1+o(1))P_1P_2$. Furthermore, conditional on this holding, then by a variant of Lemma 3.5 that simply includes extra conditions on the neighbourhoods of a vertex up to distance t_0+10, the conditional probability P that $d(x,y) \geq 2(t_0+10)+k_{\max}$ satisfies

$$P = \exp\left(-\frac{\rho_1\rho_2}{n}(1+o(1))\sum_{i=1}^{k_{\max}-1}\lambda^i\right) + o(n^{-3}). \tag{3.29}$$

Since $\rho_i \leq 2\omega_i = 2000r_i$, using (3.23) shows that $P \geq \exp(-O(1))$, so $P = \Theta(1)$.

Let us call an ordered pair (x,y) a *regular far pair* if $E_{\rho_1}(x)$ and $E_{\rho_2}(y)$ hold, and $d(x,y) \geq 2(t_0+10)+k_{\max}$, and let N denote the number of regular far pairs; our aim is to show that $N \geq 1$ holds whp. From (3.27) we have $nP_1, nP_2 \geq (1+o(1))\lambda^{25} \to \infty$, so

$$\mathbb{E}\,N \sim n^2P_1P_2P \to \infty.$$

Unfortunately, we cannot use the trick from Subsection 2.2 to complete the proof: this trick, which allowed us to avoid considering the second moment of the number of *pairs* of vertices at large distance, needed $P \sim 1$. This will in fact hold for almost all values of the parameters in the present setting, but not all. Moreover, we now have less tolerance in the final estimate of the diameter, and consequently less flexibility. Instead we apply the second moment method directly to N. In the arguments that follow we shall avoid using the fact that $P = \Theta(1)$, using only

$$P \geq n^{-1/20}, \tag{3.30}$$

say; this will be useful later.

Let $M = \mathbb{E}(N^2)$ denote the expected number of pairs $((x,y),(z,w))$ of regular far pairs; our aim is to show that $\mathbb{E}\,M \sim (\mathbb{E}\,N)^2$. Note that the number of distinct vertices in $\{x,y,z,w\}$ may be 2, 3 or 4. The contribution to M from sets with 2 distinct vertices is trivially at most $2\,\mathbb{E}\,N = o((\mathbb{E}\,N)^2)$ (the factor 2 arises only if $\rho_1 = \rho_2$). Let us leave aside the case of 3 vertices, noting only that we expect the contribution from pairs with $x = z$, say, to be asymptotically

$$nP_1(nP_2)^2P^2 \sim (\mathbb{E}\,N)^2/(nP_1) = o((\mathbb{E}\,N)^2),$$

since $nP_1 \to \infty$. The argument for the case of 4 distinct vertices that we shall now give adapts easily to show this.

Let M_0 be the contribution to M arising from sets of 4 distinct vertices $\{x,y,z,w\}$ whose neighbourhoods up to distance $t_0+10+\ell(\rho_i)$ are all disjoint, where $i = 1$ or 2 as appropriate. To estimate M_0, explore from four distinct vertices, and test whether the relevant events $E_{\rho_i}(\cdot)$ hold with the neighbourhoods disjoint. As in (3.17), this has probability $(1+o(1))P_1^2P_2^2$. Our aim is to bound from above the conditional probability that $d(x,y), d(z,w) \geq 2(t_0+10)+k_{\max}$, showing that it is at most $(1+o(1))P^2$. Since none of x, y, z, w is in B_2, the neighbourhoods have already reached size at least $\log^6 n$. From this point onwards, as before, we may assume they grow at almost exactly the expected rate. Note that we may ignore events of conditional probability $o(n^{-1/10}) = o(P^2)$, since we have already conditioned on an event of probability $(1+o(1))P_1^2P_2^2$.

Since we stop the explorations when the neighbourhoods are no larger than $n^{3/5}$, say, we may assume that any intersections between neighbourhoods are small, involving at most a fraction $n^{-1/3}$ of the vertices in a neighbourhood. Such small intersections do not materially affect the calculations in Lemma 3.5(b), so the conditional probability that $d(x, y), d(z, w) \geq 2(t_0 + 10) + k_{\max}$ is indeed $(1 + o(1))P^2$.

It remains to deal with cases where some of the neighbourhoods meet within distance $t_0 + 10 + \ell(\rho_i)$ from the respective vertices. As above just after (3.25), let t_i be the relevant parameter t' in Lemma 3.3, where $i = 1$ or 2 depending on which vertex we consider. Note that to have the property $E_{\rho_i}(v)$, all our starting vertices v must have the property that $\Gamma_{t_i}(v)$ contains a unique vertex v_0. Also, within the tree up to this point, v must be the unique vertex at maximal distance from v_0, so our 'diamond' condition holds. As in Subsection 2.2, it follows that in a quadruple contributing to M, the neighbourhoods cannot meet before the corresponding vertices v_0, so the minimum possible distance between starting vertices is $t_i + t_j$.

Returning to the random graph without conditioning, let us explore the neighbourhoods of our 4 distinct vertices x, y, z, w out to distance $t_i - 1$ in each case, assuming these explorations are disjoint, and that there are no edges between the final sets (such an edge would give distance $t_i + t_j - 1$). Furthermore, let us test for each of these vertices v *how many* neighbours $\Gamma_{t_i-1}(v)$ has in the remaining set U of 'unused' vertices, but not *which* neighbours it has. If our quadruple is to contribute, in each case there must be exactly one neighbour, v_0. Now conditional on the information so far, the probability that $x_0 = z_0$, say, is exactly $1/|U| \sim 1/n$. If this happens, then going forwards, the remaining calculations are exactly as if we had $x = z$ in the beginning. Summing the corresponding contributions to M, the total from cases with $x \neq z$ but $x_0 = z_0$ has an extra factor of n from the choice of z (compared to the case $x = z$), but also an extra factor that is asymptotic to $1/n$ as noted above. (There is also the extra factor of at most 1 from the condition on the neighbourhoods of z up to distance $t_i - 1$; we can ignore this). In total, the contribution here is at most that with $x = z$, which is $o((\mathbb{E}\,N)^2)$ as noted above. (The argument here is not circular; when considering here the three-vertex case, a collision of this form reduces to the two-vertex case.)

So we may assume that x_0, y_0, z_0 and w_0 are distinct. Repeating the trick above, let us first test how many neighbours each has among the unused vertices (not testing edges such as x_0z_0 for now). For our quadruple to contribute, by definition of $\widetilde{E}_{\rho_i}(\cdot)$ the numbers must be at most λ^{1-a_i} with $i = 1, 2$ as appropriate. Since there are $n - O(n^{1/4})$ unused vertices, the probability of this happening is very close to $\mathbb{P}(\mathrm{Po}(\lambda) \leq \lambda^{1-a_i})$. Using (3.28), it follows that the probability that all our tests so far, for the relevant events $\widetilde{E}_{\rho_i}(\cdot)$, succeed is at most $\lambda^{61}P_1^2P_2^2$. Hence, going forward, we may neglect any event of probability smaller than $n^{-1/4} = o(\lambda^{-61})$, say. So far we revealed the numbers of neighbours, which were all at most λ, but not which vertices they were. But the probability of a collision is $O(\lambda^2/n) = o(n^{-1/4})$, which is negligible.

Also, the probability of an edge between x_0 and z_0, say, is $O(\lambda/n) = o(n^{-1/4})$. Recall that any vertex in a pair counted in N, or a quadruple in M or M_0, has the property E_{ρ_i} for some i and is hence in $V_0 = V(G) \setminus (B_1 \cup B_2)$. Exploring further up to distance $10 + \ell(\rho_i)$ steps from each vertex v_0, where $i = 1$ or 2 as appropriate, assuming typical growth as we may, the probability that two neighbourhoods meet, starting as they do with at most λ neighbours of v_0, is $O(\lambda^{20+\ell(\rho_i)+\ell(\rho_j)}/n) = o(n^{-1/4})$. So we may assume this does not happen, and hence $M - M_0$ is negligible compared with M.

In summary, it follows that $M = \mathbb{E}(N^2) \sim n^4 P_1^2 P_2^2 P^2 \sim (\mathbb{E}\, N)^2 \to \infty$, so the second moment method shows that $N \geq 1$ whp. But then the diameter is at least $2(t_0 + 10) + k_{\max}$, completing the proof of the first half of Theorem 1.2.

The second part of the theorem states that for 'most' values of n the diameter is almost determined, and gives a formula. The general exact formula is a bit complicated if we want to include all values of the parameters, even restricting to those for which the diameter is almost determined. In formulating Theorem 1.2 we omitted some additional problematic values of n, giving a much simpler formula. One way to explain the source of the problematic cases is to observe that, although the difference between the upper and lower bounds (3.9) and (3.10) is usually negligible, when the typical diameter is close to jumping to the next integer, the fact that these bounds do not exactly match becomes important.

Writing $\{x\}$ for $x - \lfloor x \rfloor$, in proving the second part of the theorem we may assume that

$$\begin{array}{ccccc} 5\varepsilon & < & \{\log n/\log\lambda\} & < & 1-5\varepsilon, \\ 5\varepsilon & < & \{\log n/\log(1/\lambda_\star)\} & < & 1-5\varepsilon, \end{array} \tag{3.31}$$

where ε is some positive constant, which we may take to be smaller than 1/10.

Let us first consider some values of r that, as it will turn out, in many cases (i.e., for many values of n) typically determine the diameter of the random graph.

Define q_n to be the infimum of q such that $n^{-1} > \lambda_\star^{t_0+g(q)}$. From the definition of g, with λ fixed and q varying, $\lambda_\star^{g(q)}$ jumps by a factor of at most λ at each discontinuity. (With $X \sim \mathrm{Po}(\lambda)$, the ratios $\mathbb{P}(X \leq k+1)/\mathbb{P}(X \leq k)$ are between 1 and λ, while the ratio $\lambda_\star^{-1}\mathbb{P}(X \leq 1)/\mathbb{P}(X < \lambda)$ is asymptotically $1/\mathbb{P}(X < \lambda) \sim 2$.) Thus for large n

$$n^{-1} = \lambda_\star^{t_0+g(q_n)}/\xi \tag{3.32}$$

for some $\xi = \xi(n)$ between 1 and λ. We call n 'normal' if $q_n < \varepsilon$ and $g(q_n) > 4\lambda^{-\varepsilon}$. Taking logs in (3.32), since $\xi = \lambda_\star^{o(1)}$, while $t_0 = \lfloor \log n/\log(1/\lambda_\star) \rfloor$, we have $g(q_n) = \{\log n/\log(1/\lambda_\star)\} + o(1)$, so $g(q_n) \geq \varepsilon \geq 4\lambda^{-\varepsilon}$ if n is large. Since for any constant $0 < a < 1$ we have $g(a) \to 1$, while $g(q_n) \leq 1 - \varepsilon$, it follows that $q_n = o(1)$, so any (large enough) n satisfying (3.31) is normal.

Putting $t = t_0 + 10$ and $\omega = \lambda^{t_1}$ such that $t_1 = 10 - q_n - \log 5/\log\lambda$ in Lemma 3.2, we find

$$\mathbb{P}(0 < |X_{t_0+10}| < \lambda^{10-q_n}/10) \leq 3\lambda_\star^{g(t_0+q_n+\log 5/\log\lambda)} = 3\lambda_\star^{t_0+g(q_n+\log 5/\log\lambda)},$$

which is at most $3\lambda_\star^{t_0+g(q_n)}/\lambda^{5/4} = o(n^{-1})$ by (3.4) and (3.32). Hence, arguing as for (3.14), we only need to consider vertices in S_r with $r \geq \lambda^{10-q_n}/10$.

Put $b = \lfloor \log n / \log \lambda + 2q_n \rfloor$ and $\phi = \{\log n / \log \lambda + 2q_n\}$. Call n 'standard' if $3\varepsilon < \phi < 1 - 3\varepsilon$. Since $q_n < \varepsilon$ for normal n, any n satisfying (3.31) is standard.

As noted above, for normal n we only need to consider r_1 and r_2 at least $\lambda^{10-q_n-o(1)}$, and for such cases (3.18) gives

$$\begin{aligned} \widehat{\mu}(r_1, r_2, b-18) &\leq (1+o(1))\widehat{\mu}_1\widehat{\mu}_{r_2}\exp(-\lambda^{20-2q_n-o(1)+b-18-\log n/\log\lambda} + o(n^{-3})) \\ &= (1+o(1))\widehat{\mu}_{r_1}\widehat{\mu}_{r_2}\exp(-\lambda^{2-\phi-o(1)} + o(n^{-3})). \end{aligned}$$

For standard n the exponential above is at most $\exp(-\lambda^{1+\varepsilon-o(1)}) + o(n^{-3})$. Hence for such n the quantity $\widehat{\mu}(r_1, r_2, b-18)$ goes to 0 quickly unless $\widehat{\mu}_{r_1}$ or $\widehat{\mu}_{r_2}$ is much bigger than $e^{100\lambda}$ say. From arguments as above, we know this forces r_1 and r_2 to be much larger than the typical values of around λ^{10}, at least λ^{100}, say, and then $\widehat{\mu}(r_1, r_2, b-18)$ is much smaller. Using the argument that earlier permitted us to restrict parameters to the set R, such cases can be neglected. Thus, whp there are no vertices in sets $S_{r_1} \cap V_0$, $S_{r_2} \cap V_0$ that have distance greater than $2(t_0+10)+b-18$, for any r_1 or r_2. Hence the diameter is at most $2t_0+b+2$ whp for any n satisfying (3.31) (or indeed, though we won't need it, for any normal standard n).

Continuing with standard normal n, let $\omega = \lambda^{10}$. Then using (3.32) and since $g(q_n) \geq 4\lambda^{-\varepsilon}$, $\lambda_\star^{-1} = e^{\lambda+O(\log\lambda)}$ and $\xi = e^{O(\log\lambda)}$,

$$\begin{aligned} \lambda_\star^{g(t_0+10-\log\omega/\log\lambda)} &= \lambda_\star^{t_0} \\ &= \lambda_\star^{-g(q_n)}\xi/n \\ &> \exp(4\lambda^{1-\varepsilon} + O(\log\lambda))/n \\ &> \exp(3\lambda^{1-\varepsilon})/n. \end{aligned}$$

Since the final bound is larger than λ^{40}/n if λ is large enough, the bound (3.25) holds with $\omega_i = \omega$ for $i = 1, 2$. The calculations down to (3.28) go through as before, now with $\rho_1 = \rho_2 = \rho$ and $\lambda^{10}/3 \leq \rho \leq 2\lambda^{10}$. This time we have $P_1 = P_2 \sim \pi_i \geq \exp(3\lambda^{1-\varepsilon})\lambda^{-O(1)}/n$, using (3.26) and the bound above.

Writing N for the number of pairs of vertices with property E_ρ at distance at least $2(t_0 + 10) + b - 18$, as before we have $\mathbb{E}\, N \sim (P_1 n)^2 P$, with

$$P = \exp\left(-\frac{\rho^2}{n}(1+o(1)) \sum_{i=1}^{b-19} \lambda^i \right) + o(n^{-3})$$

in place of (3.29). Since $\rho \leq 2\lambda^{10}$ we have

$$\log(1/P) \leq (4+o(1))\lambda^{20+b-19-\log n/\log\lambda} \sim 4\lambda^{1+2q_n-\phi} \leq 4\lambda^{1-\varepsilon}$$

for normal standard n. Since $P_1 n \geq \exp(3\lambda^{1-\varepsilon} - O(\log\lambda))$, we thus have $\mathbb{E}\, N \to \infty$. The second moment argument goes through as before to show that whp $N \geq 1$, so the diameter is whp at least $2(t_0+10)+b-18$. (Note that we still have (3.30) since (3.31) forces $\lambda_\star$ to be much larger than $1/n$, and hence $\lambda = O(\log n)$, so $\log(1/P) = o(\log n)$.) Hence, from the upper bound shown above, the diameter of the graph is, for normal standard n, whp equal to

$$2t_0 + b + 2 = 2\lfloor \log n / \log(1/\lambda_\star) \rfloor + \lfloor \log n/\log\lambda + 2q_n \rfloor + 2.$$

Using (3.31) again, and recalling that $q_n < \varepsilon$, this is

$$2\lfloor \log n/\log(1/\lambda_\star)\rfloor + \lfloor \log n/\log\lambda\rfloor + 2,$$

which is in turn exactly the diameter claimed in (1.5), completing the proof of Theorem 1.2. □

Remark. With hindsight, it is easy to explain intuitively why the diameter in the last case treated above is given by (1.5). Indeed, with $t_0 = \lfloor \log n/\log(1/\lambda_\star)\rfloor$, the probability that a given vertex has $|\Gamma_{t_0}(v)| = 1$ is roughly $\lambda_\star^{t_0}$, which is significantly larger than $1/n$. On the other hand, typically no vertices will have $|\Gamma_{t_0+1}(v)| = 1$, or indeed $|\Gamma_{t_0+1}(v)|$ much smaller than λ. So the diameter is likely to come from two of these 'candidate' vertices with $|\Gamma_{t_0}(v)| = 1$. Each of these has a unique vertex at distance t_0. Let us call such vertices *active*. Any given pair of active vertices will usually be at distance $d = \lceil \log n/\log\lambda\rceil$ from each other. However, there are usually many candidate vertices (at least $\lambda_\star^{-\varepsilon}$, which is roughly $e^{\varepsilon\lambda}$), and hence (not necessarily, but usually) about the same number of active vertices. The expected number of paths of length d joining two given active vertices is roughly $\lambda^d/n = \lambda^{1-f}$, so we might expect the probability that a given pair is at distance $d+1$ to be of order $\exp(-\lambda^{1-f})$, where f is the fractional part of $\log n/\log\lambda$. The probability of no path of length $d+1$ is roughly $\exp(-\lambda^{2-f})$, which is much smaller than the reciprocal of the number of pairs of candidate vertices. So we expect the diameter to be $2t_0 + d + 1$ whp, as we have shown.

4 Just above the critical point

In this section we shall prove Theorem 1.3, which is the analogue of Theorem 1.1 for $G(n, \lambda/n)$, where now $\lambda = 1 + \varepsilon$ with $\varepsilon = \varepsilon(n)$ tending to zero at a suitable rate. Roughly speaking, we shall simply repeat the arguments in Section 2 more carefully; however, there are many additional complications that we shall contend with as we go. As mentioned in the introduction, we shall also prove a stronger result, describing the (normalized) limiting distribution of the correction term; we postpone the somewhat unpleasant statement of this result until Section 5.

Throughout this section we write λ for $1+\varepsilon$, always assuming that $0 < \varepsilon < 1/10$, and often that $\varepsilon = \varepsilon(n) \to 0$. As before, we write $\lambda_\star$ for the unique solution $\lambda_\star < 1$ to $\lambda_\star e^{-\lambda_\star} = \lambda e^{-\lambda}$, so

$$\lambda_\star = 1 - \varepsilon + \frac{2}{3}\varepsilon^2 - \frac{4}{9}\varepsilon^3 + O(\varepsilon^4). \tag{4.1}$$

Sometimes it will be convenient to note that

$$\lambda_\star > 1 - \varepsilon \tag{4.2}$$

for all $\varepsilon > 0$; this is easily seen using the fact that $\lambda_\star e^{-\lambda_\star}$ has positive derivative, and $(1-\varepsilon)e^{-(1-\varepsilon)} < \lambda e^{-\lambda}$. As before we write s for the survival probability of the branching process $\mathfrak{X}_\lambda$, so (from (1.1)), we have

$$s = 2\varepsilon + O(\varepsilon^2). \tag{4.3}$$

Note that as $\varepsilon \to 0$ we have
$$\log(1/\lambda_\star) \sim \varepsilon \sim \log \lambda. \tag{4.4}$$

The overall plan of the proof is as for the cases λ constant and $\lambda \to \infty$. We shall treat the second phase (regular growth) in Subsection 4.1 and the first phase, approximation by the branching process, in Subsection 4.2. To be able to carry out the third phase, we still need to study the distribution of the time the branching process takes to reach a large size. We do this in Subsection 4.3, and prove various other branching process lemmas we shall need in Subsection 4.4. In Subsection 4.5 we consider the typical distances in the 2-core. Finally, armed with all these results, we prove the lower bound on the diameter in Subsection 4.6, and the upper bound in Subsection 4.7; this turns out to be not as easy as one might expect, and both proofs involve considerable re-examination of the first phase, the early growth of the neighbourhoods.

One complication concerns the wedge condition used in Section 2; here this turns out to have probability $\Theta(\varepsilon^3)$, or $\Theta(\varepsilon^2)$ if we condition on the vertex being in the giant component. In Section 2, we used a much stronger 'diamond' condition, that allowed us to simply avoid dependence between the neighbourhoods of the vertices we considered. Unfortunately, the diamond condition corresponds roughly to two wedge conditions, and has probability $\Theta(\varepsilon^4)$ after conditioning on being in the giant component. When $\varepsilon \to 0$, we cannot afford to give up a factor ε^2 in the number of vertices we consider to develop neighbourhoods from in the third phase.

Except that Subsections 4.3 and 4.4 belong together, Subsections 4.1 to 4.4 may be read in any order. We have chosen the present order as the first two subsections are relatively simple, and may be seen as motivating the extensive branching process analysis that follows.

Throughout we write Λ for $\varepsilon^3 n$, and assume that $\Lambda \to \infty$. In particular, we allow ourselves to assume that Λ is 'sufficiently large' (i.e., larger than some implicit constant) whenever this is convenient. As noted in the introduction, in proving Theorem 1.3 we may assume that $\varepsilon \to 0$; correspondingly, we shall assume without comment that ε is 'sufficiently small' whenever convenient.

In what follows we shall use standard results about the component structure of $G(n,p)$ just above the phase transition; let us recall these here. We write $C_i(G)$ for the number of vertices in the ith largest component of a graph G.

Theorem 4.1. *Let $\lambda = 1+\varepsilon$, where $\varepsilon = \varepsilon(n) > 0$ satisfies $\varepsilon \to 0$ and $\Lambda = \varepsilon^3 n \to \infty$, and let $s = s(\lambda)$ denote the survival probability of $\mathfrak{X}_\lambda$. Then*

$$C_1(G(n,\lambda/n)) = sn + O_{\mathrm{p}}(\varepsilon n/\sqrt{\Lambda}), \tag{4.5}$$

and

$$C_2(G(n,\lambda/n)) = \delta^{-1}\left(\log\Lambda - \frac{5\log\log\Lambda}{2} + O_{\mathrm{p}}(1)\right),$$

where

$$\delta = \lambda - 1 - \log\lambda = \varepsilon^2/2 - \varepsilon^3/3 + O(\varepsilon^4).$$

□

This result, which extends results of Bollobás [7, 8] by removing a logarithmic lower bound on Λ from the conditions, is essentially due to Łuczak [30]. Note, however, that the actual formula for C_2 given in [30] is incorrect; see the discussion in Bollobás and Riordan [13, Section 3.4], where a proof of the above result based on branching processes is given.

Formally, by the *giant component* of $G = G(n, \lambda/n)$ we mean the component C_1 with the most vertices (chosen according to any rule if there is a tie). Recalling from (4.3) that $s \sim 2\varepsilon$, under the conditions of Theorem 4.1 we have

$$|C_1| = (2 + o_{\mathrm{p}}(1))\varepsilon n. \tag{4.6}$$

4.1 Large neighbourhoods and meeting in the middle

In this subsection we show that whp once the neighbourhoods of a vertex become large, they grow at the expected rate until reaching size $\sqrt{\varepsilon n}\log\Lambda$, say. Showing this is not quite as simple as proving Lemma 2.4, since when ε is small, even when the neighbourhoods are fairly large, the expected increase in size from one step to the next may still be smaller than the standard deviation. Hence it may well happen that $\Gamma_t(x)$ is smaller than $\Gamma_{t-1}(x)$ for some t. However, this is unlikely to happen for many consecutive t.

We shall start by proving a corresponding growth result for a Galton–Watson branching process. It may well be that a similar result exists in the literature, but we have not found it; the key point is the dependence of the bounds on the parameters of the branching process. The general theme here and throughout this section is that the behaviour of the branching process is only 'regular' once it reaches sizes larger than $1/\varepsilon$, and that it is best seen on time scales on the order of $1/\varepsilon$, the typical time required for a constant factor change in the size of a generation.

Given parameters $\mu = 1 + \varepsilon$ and n, consider a Galton–Watson branching process $(Z_t)_{t\geq 0}$ starting with a fixed number N_0 of particles, in which each particle has a binomial number of children in the next generation, with parameters n and μ/n. Let $N_t = |Z_t|$ denote the number of particles in generation t.

Lemma 4.2. *Let $0 < \varepsilon, \delta < 1$ and n be given, and define (N_t) as above, with $\mu = 1 + \varepsilon$. Writing ω for εN_0, the probability that*

$$(1 - \delta)\, N_0\mu^t \leq N_t \leq (1 + \delta)\, N_0\mu^t \tag{4.7}$$

holds for all $t \geq 0$ is at least $1 - O(e^{-c_0\delta^2\omega})$, where $c_0 > 0$ is an absolute constant, and the implicit constant in the $O(\cdot)$ notation is absolute.

Proof. We may and shall assume that $\delta^2\omega \geq 100$, say; otherwise, there is nothing to prove.

We may construct (Z_t) in small steps in the following standard way: let $A_1, A_2, \ldots$ be independent binomial $\mathrm{Bi}(n, \mu/n)$ random variables. As we construct the process, we number the particles in order of the time they are born; we start by numbering the particles of Z_0 with $1, 2, \ldots, N_0$ in any order. To define (Z_t), simply take A_i to be the number of children of the ith particle. Writing S_t for $\sum_{t'<t} N_{t'}$, we then have

$$N_t = N_0 + \sum_{i\leq S_t}(A_i - 1) = N_0 + B_{S_t} - S_t, \tag{4.8}$$

where $B_i = \sum_{j \le i} A_j$.

For $t \ge -1/\varepsilon$ set

$$\delta_t = \frac{\varepsilon\delta}{8} \int_{s=-1/\varepsilon}^{t} \mu^{-s/4}\,\mathrm{d}s,$$

and set $\delta_t = 0$ if $t < -1/\varepsilon$. Note that δ_t is an increasing function of t, with

$$0 \le \delta_t \le \frac{\varepsilon\delta}{8}\frac{4}{\log\mu}\mu^{1/(4\varepsilon)} \le \delta,$$

using $(1+\varepsilon)^{1/(4\varepsilon)} < e^{\varepsilon/(4\varepsilon)} = e^{1/4}$ and $\varepsilon/\log\mu = \varepsilon/\log(1+\varepsilon) \le 1/\log 2$.

The key property of δ_t is that if $t \ge 0$ and $r = t - 1/\varepsilon$, then

$$\delta_t - \delta_r \ge (t-r)\frac{\varepsilon\delta}{8}\mu^{-t/4} = \delta\mu^{-t/4}/8. \tag{4.9}$$

For $t \ge 1$, let E_t be the event that $N_t > (1+\delta_t)N_0\mu^t$ holds but $N_s \le (1+\delta_s)N_0\mu^s$ for all $0 \le s < t$. Suppose that the upper bound in (4.7) fails for some t. Then $N_t > (1+\delta_t)N_0\mu^t$ for this t, and it follows that one of the events E_t holds.

Suppose that E_t holds for some $t \ge 0$. Set $r = t - 1/\varepsilon$, and, for convenience, set $N_s = \mu^s N_0$ for all negative integers s, so $\sum_{s<0} N_s = N_0/\varepsilon$. Then, with all sums starting at $-\infty$ unless otherwise indicated,

$$\begin{aligned} S_t + N_0/\varepsilon &= \sum_{s<t} N_s \le \sum_{s<t}(1+\delta_s)\mu^s N_0 \\ &\le \sum_{s<r}(1+\delta_r)\mu^s N_0 + \sum_{r\le s<t}(1+\delta_t)\mu^s N_0 \\ &= \sum_{s<t}(1+\delta_t)\mu^s N_0 - \sum_{s<r}(\delta_t-\delta_r)\mu^s N_0 \\ &= \frac{N_0}{\varepsilon}\big((1+\delta_t)\mu^t - (\delta_t-\delta_r)\mu^{\lceil r\rceil}\big) \\ &< \frac{N_0}{\varepsilon}\left((1+\delta_t)\mu^t - (\delta_t-\delta_r)\frac{\mu^t}{4}\right), \end{aligned}$$

since $\mu^{t-\lceil r\rceil} = (1+\varepsilon)^{\lfloor 1/\varepsilon\rfloor} \le (1+\varepsilon)^{1/\varepsilon} < e < 4$.

For each fixed i, let $f(i) = (1+\varepsilon)i$ denote the expectation of $B_i = \sum_{j=1}^{i} A_j$. From the above, we have

$$f(S_t) - S_t + N_0 = \varepsilon S_t + N_0 = \varepsilon(S_t + N_0/\varepsilon) \le N_0(1+\delta_t)\mu^t - N_0(\delta_t-\delta_r)\mu^t/4.$$

On the other hand, since E_t holds we have $N_t > (1+\delta_t)\mu^t N_0$, so from (4.8) it follows that

$$B_{S_t} - S_t + N_0 = N_t > (1+\delta_t)\mu^t N_0.$$

Combining the two equations above, using (4.9), and recalling that $N_0 = \omega/\varepsilon$, we see that

$$B_{S_t} - f(S_t) > N_0(\delta_t-\delta_r)\mu^t/4 \ge N_0(\delta\mu^{-t/4}/8)\mu^t/4 = \delta\omega\varepsilon^{-1}\mu^{3t/4}/32. \tag{4.10}$$

On the other hand, from the bound on $S_t + N_0/\varepsilon$ above we have, very crudely,

$$S_t \le \frac{N_0}{\varepsilon}(1+\delta_t)\mu^t \le 2\omega\varepsilon^{-2}\mu^t. \tag{4.11}$$

From (4.10) and (4.11) it follows that $B_{S_t} - f(S_t) \ge g(S_t)$, where

$$g(i) = \max\left\{\delta\omega\varepsilon^{-1}/32, i^{3/4}\delta\omega^{1/4}\varepsilon^{1/2}/60\right\}.$$

Let F_i be the event that $B_i - f(i) = B_i - \mathbb{E}\, B_i \ge g(i)$. We have shown that if one of the events E_t holds, then so does one of the events F_i. At this point we could simply bound the probability of the union of the F_i by the sum of their probabilities, but as they are highly dependent, this is rather inefficient.

Let $T = \lceil \omega/\varepsilon^2 \rceil$, noting that $T \ge 100$ (when ε is small enough) and $T < 2\omega/\varepsilon^2$. For $k = 0, 1, 2, \ldots$, let G_k be the event $\bigcup_{kT < i \le (k+1)T} F_i$, so

$$\mathbb{P}\Big(\bigcup_{t\ge1} E_t\Big) \le \mathbb{P}\Big(\bigcup_{i\ge1} F_i\Big) = \mathbb{P}\Big(\bigcup_{k\ge0} G_k\Big) \le \sum_{k=0}^{\infty} \mathbb{P}(G_k).$$

Finally, let G'_k be the event that $B_{(k+2)T} - \mathbb{E}\, B_{(k+2)T} \ge g(kT)$. Let us estimate $\mathbb{P}(G'_k \mid G_k)$. We test whether G_k holds by examining each B_i in turn, stopping at the first $i > kT$ for which F_i holds. Suppose G_k does hold, and that we stop at $i = i'$, so $kT < i' \le (k+1)T$. Recalling that $B_i = \sum_{j \le i} A_j$, where the A_j are independent with distribution $\mathrm{Bi}(n, \mu/n)$, we have not yet examined any A_j, $j > i'$. Hence the conditional distribution of $\Delta = B_{(k+2)T} - B_{i'} = \sum_{i' < j \le (k+2)T} A_j$ is just its unconditional distribution, which is binomial with mean $(1+\varepsilon)((k+2)T - i') \ge T \ge 100$. It is easy to check (for example from the Berry–Esséen Theorem) that this binomial distribution is well approximated by a normal distribution, and in particular, Δ exceeds its mean with probability at least 1/3. But when this happens,

$$B_{(k+2)T} - \mathbb{E}\, B_{(k+2)T} = B_{i'} - \mathbb{E}\, B_{i'} + \Delta - \mathbb{E}\,\Delta \ge B_{i'} - \mathbb{E}\, B_{i'} \ge g(i') \ge g(kT),$$

since we are assuming $F_{i'}$ holds, and $g(\cdot)$ is non-decreasing. Thus, given G_k, the event G'_k holds with probability at least 1/3. Hence $\mathbb{P}(G'_k) \ge \mathbb{P}(G_k)/3$, so $\mathbb{P}(G_k) \le 3\mathbb{P}(G'_k)$.

Now G'_0 is the event that B_{2T}, a variable with binomial distribution with mean $\mu_0 = (1+\varepsilon)2T \le 4T \le 8\omega\varepsilon^{-2}$, exceeds its mean by at least $x_0 = \delta\omega\varepsilon^{-1}/32$. Since $x_0 \le \mu_0$, Lemma 2.3 applies, and we see that $\mathbb{P}(G'_0) \le 2\exp(-x_0^2/(3\mu_0)) \le 2\exp(-\delta^2\omega/24576)$.

For $k \ge 1$, G'_k is the event that $B_{(k+2)T}$, which has a binomial distribution with mean $\mu_k = (1+\varepsilon)(k+2)T \le 12k\omega\varepsilon^{-2}$, exceeds its mean by $x_k = g(kT) \ge g(k\omega\varepsilon^{-2}) \ge k^{3/4}\omega\varepsilon^{-1}\delta/60$. Since $x_k \le \mu_k$, by Lemma 2.3 we have $\mathbb{P}(G'_k) \le 2\exp(-c_0 k^{1/2}\delta^2\omega)$ for some absolute constant $c_0 > 0$. Hence, reducing c_0 if necessary,

$$\mathbb{P}\Big(\bigcup_t E_t\Big) \le 3\sum_k \mathbb{P}(G'_k) \le 2e^{-c_0\delta^2\omega} + \sum_{k\ge1} 2e^{-c_0 k^{1/2}\delta^2\omega} = O\big(e^{-c_0\delta^2\omega}\big),$$

recalling that $\delta^2\omega \ge 100$. As noted above, if the upper bound in (4.7) fails, then some E_t holds, so we have proved that the upper bound holds with the required probability.

The argument for the lower bound is almost identical. Let E'_t be the event that $N_t < (1-\delta_t)N_0\mu^t$ holds but $N_s \ge (1-\delta_s)N_0\mu^s$ for all $s < t$. Changing signs in the argument above, we see that if E'_t holds then the equivalent of (4.10) holds, namely

$$B_{S_t} - f(S_t) \le -\delta\omega\varepsilon^{-1}\mu^{3t/4}/32. \tag{4.12}$$

The proof of (4.11) only assumed $N_s \le (1+\delta_s)N_0\mu^s$ for $s < t$, which we now know to be true with the required probability. If (4.11) does hold, then (4.12) implies that $B_i - f(i) \le -g(i)$ holds for some i. We may bound the probability of this event just as for F_i above, completing the proof. □

Turning to the graph, Lemma 4.2 enables us to prove the required growth result. Our choice of the parameters here is somewhat arbitrary, but will be useful later. Recall that $V(G)$ denotes the vertex set of a graph G, and $\Gamma_r(x)$ the set of vertices at graph distance r from a vertex x.

Lemma 4.3. *Let $\varepsilon = \varepsilon(n) \le 1$ satisfy $\Lambda = \varepsilon^3 n \to \infty$. Set $\lambda = 1+\varepsilon$, $\omega = \Lambda^{1/6}$, and $M = \sqrt{\omega\varepsilon n}$. For $x \in V(G(n,\lambda/n))$ and $r \ge 0$, let $E_{x,r}$ be the event that*

$$(1-2\omega^{-1/3})\lambda^t|\Gamma_r(x)| \le |\Gamma_{r+t}(x)| \le (1+\omega^{-1/3})\lambda^t|\Gamma_r(x)|$$

holds for $0 \le t \le T = \log(\varepsilon M/\omega)/\log\lambda$. Then, for some absolute constant c_0,

$$\mathbb{P}\big(E_{x,r} \,\big|\, |\Gamma_0(x)|, \dots, |\Gamma_r(x)|\big) \ge 1 - O\big(\exp(-c_0\omega^{1/3})\big) = 1 - o(\Lambda^{-100})$$

whenever $\omega/\varepsilon \le |\Gamma_r(x)| \le 2\omega/\varepsilon$ and $\sum_{r' \le r} |\Gamma_r(x)| \le n^{2/3}$.

In other words, once we reach size ω/ε in the neighbourhood exploration, provided we have not so far used up too many vertices, the neighbourhoods grow at the expected rate until they reach size approximately M. Note that if $\Lambda = \varepsilon^3 n \ge (\log n)^{20}$, then the error term in the form $O(\exp(-c_0\omega^{1/3}))$ is $o(n^{-100})$, i.e., utterly negligible.

Proof. Condition on the result of the exploration up to step r, assuming that we find between ω/ε and $2\omega/\varepsilon$ vertices in the last generation and have seen at most $n^{2/3}$ vertices so far. Let $N'_t = |\Gamma_{r+t}(x)|$. The (conditional) distribution of the process $(N'_t)_{t\ge 0}$ is very similar to that of (N_t): the only difference is that each vertex gives rise to a binomial $\mathrm{Bi}(m,\lambda/n)$ number of children in the next generation, where m is the number of vertices not seen so far.

For the upper bound on the neighbourhood sizes, we simply note that $m \le n$, so (N'_t) is stochastically dominated by (N_t). The result thus follows immediately from Lemma 4.2.

For the lower bound, set $n' = n - 2n^{2/3}$. Note that if the upper bound holds, which it does with probability $1 - O(\exp(-c_0\omega^{1/3}))$, then by time T we have used at most $n^{2/3} + 10M/\varepsilon \le 2n^{2/3}$ vertices, so we still have at least n' left. For times t by which we have used up at most $2n^{2/3}$ vertices, the process (N'_t) stochastically dominates a process (N''_t) in which each particle has $\mathrm{Bi}(n',\lambda/n)$ children. This binomial has mean $\mu = \lambda n'/n = (1+\varepsilon)(1-2n^{-1/3})$. Applying Lemma 4.2 again, it follows that with probability $1 - O(\exp(-c_0\omega^{1/3}))$ we have $|\Gamma_{r+t}(x)| \ge (1-\omega^{-1/3})\mu^t|\Gamma_r(x)|$ for $0 \le t \le T$. Since $T = \log(\varepsilon^{3/2}n^{1/2}\omega^{-1/2})/\log\lambda \le \log(\Lambda^{1/2})/(\varepsilon/2) = \varepsilon^{-1}\log\Lambda \le n^{1/3}/\omega$, we have $\mu^t/\lambda^t \ge 1 - 3/\omega$ for $t \le T$, so the lower bound follows. □

Remark. Let us note that, while the various constants can certainly be improved, Lemmas 4.2 and 4.3 are tight in several ways. Firstly, since the survival probability of the branching process $\mathfrak{X}_\lambda = (X_t)$ is of order ε, if we start from a neighbourhood $\Gamma_r(x)$ of size a/ε, the neighbourhood exploration process will die quickly with probability $e^{-\Theta(a)}$. Hence, in order to make it very likely that the neighbourhoods grow at the right rate, we certainly need $|\Gamma_r(x)|$ to be much larger than $1/\varepsilon$. In other words, neighbourhoods are only 'large' over size ω/ε, for some $\omega \to \infty$. Similarly, it can be seen that the form $\exp(-\Omega(\delta^2\omega))$ of the error bound in Lemma 4.2 is best possible.

Finally, when ε is close to the lower end of the range we consider, we cannot extend Lemma 4.3 to growth much beyond size $\sqrt{\varepsilon n}$; shortly beyond this point, the number of vertices 'used up' — which is larger than $\sqrt{n/\varepsilon} = \sqrt{\varepsilon n}/\varepsilon$ since about $1/\varepsilon$ generations are roughly the same size $\sqrt{\varepsilon n}$ — is sufficient to slow the growth appreciably. Fortunately, neighbourhoods of two different vertices are likely to join up when they have size around $\sqrt{\varepsilon n}$, as we shall now see. The basic explanation for this is that the probability of the $\sqrt{n/\varepsilon}$ vertices seen near one vertex being distinct from the $\sqrt{n\varepsilon}$ at a given distance from the other vertex becomes small. (It is misleading to consider separately each of the $\sqrt{n/\varepsilon}$ vertices 'near' one vertex being distinct from $\sqrt{n/\varepsilon}$ vertices 'near' the other, since these events do not have the required independence.) The fact that neighbourhoods typically join up when they each have size $\sqrt{\varepsilon n}$ explains one factor of ε in the first log in (1.6).

For $x \in V(G)$ and $a > 0$, let $t_a(x)$ denote the smallest r for which $|\Gamma_r(x)| \geq a$, if such an r exists; otherwise $t_a(x)$ is undefined. The following simple lemma captures the observation that, for large a, we are unlikely to 'overshoot' our cutoff a by too much.

Lemma 4.4. *Let $\lambda = 1+\varepsilon$ and fix an integer a, a vertex x and $\delta > 0$. Given that $t_a(x)$ is defined, the probability that $|\Gamma_{t_a(x)}(x)|$ exceeds $(1+\delta)(1+\varepsilon)a$ is $e^{-\Omega(\delta^2 a)}$, where the implicit constant is absolute.*

Proof. The event that $t_a(x)$ is defined may be written as a disjoint union of events of the form $E = \{t_a(x) = t,\ |\Gamma_{\leq t-1}(x)| = m,\ |\Gamma_{t-1}(x)| = s\}$, where $0 < s < a$. Let us condition on one such event. Given that $|\Gamma_{\leq t-1}(x)| = m$ and $|\Gamma_{t-1}(x)| = s$, the distribution of $|\Gamma_t(x)|$ is binomial with parameters $n-m$ and $1-(1-\lambda/n)^s \leq s\lambda/n$. Hence, given E, the conditional distribution of $|\Gamma_t(x)|$ is that of a binomial distribution with mean at most $s\lambda = (1+\varepsilon)s < (1+\varepsilon)a$ conditioned to be at least a. It is easy to check that the probability that such a distribution exceeds $(1+\delta)(1+\varepsilon)a$ is maximal when s is maximal, and is then (from the Chernoff bounds) of the form $e^{-\Omega(\delta^2 a)}$. □

We now turn to the time neighbourhoods take to meet having reached some 'reasonably large' size.

Lemma 4.5. *Let $\varepsilon = \varepsilon(n)$ and $\lambda = 1+\varepsilon$ be such that $\varepsilon \to 0$ and $\Lambda = \varepsilon^3 n \to \infty$. Set $\omega = \Lambda^{1/6}$, and $t_2 = \log(\varepsilon^3 n/\omega^2)/\log\lambda$. Let x and y be two vertices of $G(n, \lambda/n)$. Writing E for the event that $t_{\omega/\varepsilon}(x) = r_1$, $t_{\omega/\varepsilon}(y) = r_2$, and the graphs $G_{\leq r_1}(x)$ and $G_{\leq r_2}(y)$ each contain at most $n^{2/3}$ vertices and are disjoint, we have*

$$\mathbb{P}\big(d(x,y) \geq r_1 + r_2 + t_2 + a \mid E\big) = e^{-(1+o(1))\lambda^a} + O(e^{-c_0\omega^{1/3}}) = e^{-(1+o(1))\lambda^a} + o(\Lambda^{-10}) \quad (4.13)$$

for any function $a = a(n) \geq -t_2/2$, and

$$\mathbb{P}\big(d(x,y) \leq r_1 + r_2 + t_2 - K \bigm| E\big) = o(1)$$

whenever $K = K(n)$ is such that $\varepsilon K \to \infty$.

Proof. It suffices to prove the first statement: since $\log \lambda = \Theta(\varepsilon)$, if $\varepsilon K \to \infty$ then $K \log \lambda \to \infty$, so $\lambda^{-K} \to 0$, and the second statement follows immediately from the first. In proving (4.13), we may assume that $a \leq a_{\max} = \log \omega / (2 \log \lambda)$: otherwise, $\lambda^a \geq \omega^{1/2}$, and the additive error term in (4.13), which is independent of a, dominates the main term.

We explore the neighbourhoods of x and y in the usual way, initially stopping each exploration when we first reach a neighbourhood of size greater than ω/ε. At this point, the conditions of the theorem allow us to assume that we have used up at most $n^{2/3}$ vertices in each exploration, and that the explorations, having taken r_1 steps from x and r_2 steps from y, have not met. By Lemma 4.4, with conditional probability at least $1 - e^{-\Omega(\omega)}$ the last generation in each exploration has size at most $(1+\sqrt{\varepsilon})\omega/\varepsilon \sim \omega/\varepsilon$.

We now continue both explorations. At the start of step i, $i = 0, 1, 2, \ldots$, we have explored the neighbourhoods of x out to distance $r_1 + \lceil i/2 \rceil$ and those of y out to distance $r_2 + \lfloor i/2 \rfloor$. During step i, we first test whether any of the $|\Gamma_{r_1+\lceil i/2\rceil}(x)||\Gamma_{r_2+\lfloor i/2\rfloor}(y)|$ 'cross-edges' between these two neighbourhoods is present. If so, $d(x,y) = r_1 + r_2 + i + 1$, and we stop. Otherwise, we uncover the next neighbourhood of either x or y as appropriate and continue, stopping if we have found no cross-edge by step $t_2 + a$, in which case $d(x,y) > r_1 + r_2 + t_2 + a$.

After $t_2 + a_{\max}$ steps as above, each extending either x's or y's neighbourhood, the typical size of the neighbourhood of x or y reached is $(\omega/\varepsilon)\lambda^{(t_2+a_{\max})/2} = \omega^{1/4}\sqrt{\varepsilon n}$. In particular, this size is much less than the quantity M defined in Lemma 4.3. Hence, by Lemma 4.3, we may assume that

$$|\Gamma_{r_k+j}(x_k)| \sim \lambda^j |\Gamma_{r_k}(x_k)| \sim \lambda^j \omega/\varepsilon$$

for $k = 1, 2$ and all $j \leq (t_2 + a_{\max})/2$, where $x_1 = x$ and $x_2 = y$. Furthermore, the error terms, which are factors of the form $(1 + O(\sqrt{\varepsilon}) + O(\omega^{-1/3}))$, are uniform in j.

It follows that at step i we test $(1+o(1))\lambda^i(\omega/\varepsilon)^2$ potential cross-edges, and that by any step $i \geq t_2/2$ we have tested in total

$$(1+o(1)) \sum_{j=0}^{i} \lambda^j (\omega/\varepsilon)^2 \sim (\omega/\varepsilon)^2 \sum_{j=-\infty}^{i} \lambda^j \sim (\omega/\varepsilon)^2 \varepsilon^{-1} \lambda^i$$

potential cross-edges. (The bound $i \geq t_2/2$ is used for convenience only, to allow us to approximate the sum from $j = 0$ by the sum from $j = -\infty$.)

Since each cross-edge tested is present with its original unconditional probability of $\lambda/n \sim 1/n$, it follows that up to a $O(e^{-c_0\omega^{1/3}})$ error term (from the conclusion of Lemma 4.3 not holding, etc), the probability that the explorations do not meet by step $t_2 + a$ is

$$p_{\geq a} = (1 - \lambda/n)^{(1+o(1))\omega^2 \varepsilon^{-3} \lambda^{t_2+a}}.$$

Since

$$\log(1/p_{\geq a}) \sim (1/n)\omega^2 \varepsilon^{-3} \lambda^{t_2} \lambda^a = \lambda^a,$$

the proof is complete. □

Roughly speaking, Lemma 4.5 tells us that once the neighbourhoods of two vertices reach a decent size, ω/ε, then whp these neighbourhoods then meet within $O(1/\varepsilon)$ steps of 'when they should', which is after an extra t_2 steps. To study the diameter of $G(n, \lambda/n)$, we shall need the full strength of the bound actually proved. Note that there is variation of order $1/\varepsilon$ in the actual time taken to meet, as may be seen from (4.13), where any $a = O(1/\varepsilon)$ gives a probability bounded away from 0 and 1.

In the light of Lemma 4.5, as in the case of λ constant or $\lambda \to \infty$, the key to understanding the diameter of $G(n, \lambda/n)$ is understanding the distribution of the time taken until the neighbourhoods of a vertex reach a reasonable size (in this case ω/ε); this will be our aim in the next few subsections. We shall take $\omega = \omega(n) = \Lambda^{1/6}$, but there is in fact a wide flexibility in the choice of the function $\omega = \omega(n)$: the requirements in what follows are that ω is at least a certain power of $\log \Lambda$, and at most a certain power of Λ. If Λ is large enough to allow $\omega/\log n \to \infty$, then many arguments simplify; we shall not assume this, however.

In the remainder of this section we explain why Lemma 4.5 already gives us the *typical* distance between vertices in the giant component, if $\Lambda = \varepsilon^3 n$ is at least $(\log n)^{20}$, say. Indeed, the neighbourhoods of a random vertex of $G = G(n, \lambda/n)$ behave much like the branching process $\mathfrak{X}_\lambda = (X_t)_{t\geq 0}$, at least to start with. Roughly speaking, a vertex is in the giant component if and only if the corresponding branching process survives, which it does with probability $s \sim 2\varepsilon$. So we will be interested in the expected size of $|X_t|$ conditioned on the process surviving.

Lemma 4.6. *Let $\mathcal{S}$ be the event that $\mathfrak{X}_\lambda$ survives. Then*

$$\mathbb{E}(|X_t| \mid \mathcal{S}) = \frac{\lambda^t - (1-s)\lambda_\star^t}{s},$$

which is asymptotically $\lambda^t/s \sim \lambda^t/(2\varepsilon)$ if $\varepsilon \to 0$ and $\varepsilon t \to \infty$.

Proof. Writing $\mathbf{1}_A$ for the indicator function of an event A, we have

$$\mathbb{E}(|X_t|\mathbf{1}_{\mathcal{S}}) = \mathbb{E}(|X_t|) - \mathbb{E}(|X_t|\mathbf{1}_{\mathcal{S}^c}) = \lambda^t - (1-s)\,\mathbb{E}(|X_t| \mid \mathcal{S}^c) = \lambda^t - (1-s)\lambda_\star^t,$$

since the distribution of $\mathfrak{X}_\lambda$ conditioned on $\mathcal{S}^c$ is that of $\mathfrak{X}_{\lambda_\star}$. The result follows. □

It is not hard to see that the 'typical' size of $|X_t|$ given $\mathcal{S}$ is of the same order as the expected size; we shall give some precise results on this later. Hence, for most vertices in the giant component, their neighbourhoods take time $\log\omega/\log\lambda$ to reach size ω/ε, so the typical distance is $2\log\omega/\log\lambda + t_2 = \log(\varepsilon^3 n)/\log\lambda$. More precisely, one can check that the distance between two random vertices of the giant component is $\log(\varepsilon^3 n)/\log\lambda + O_p(1/\varepsilon)$; we shall not give the details. The rest of the proof of Theorem 1.3 essentially shows that the other term in the formula in that theorem accounts for vertices whose neighbourhoods take an abnormally long time to start growing large.

4.2 Branching process to graph

At some point, we need to compare the probabilities of events defined in terms of our random graph $G = G(n, \lambda/n)$ with events in the branching process. It turns out that we have to consider events involving trees of height $\Theta(\log\Lambda/\log\lambda) = \Theta(\varepsilon^{-1}\log\Lambda)$, recalling that $\Lambda = \varepsilon^3 n$, with (it will turn out) up to around $1/\varepsilon$ vertices at each distance from the root. For this reason, we need to consider trees with at least $\Theta(\varepsilon^{-2})$ vertices. If ε is smaller than $n^{-1/4}$, then we cannot simply extend Lemma 2.2 to cover such trees using the same proof, since the error terms $|T|^2/n$ would be too large.

Fortunately, it is easy to prove a result that applies for the trees we need. Although this is in some sense a coupling result, the obvious coupling between $G^0_{\le t}(x)$ and $\mathfrak{X}_\lambda$ fails here. This obvious coupling is based on the fact that a $\mathrm{Po}(\lambda)$ and a $\mathrm{Po}(\lambda(1-\delta))$ distribution can naturally be coupled to agree with probability at least $1-\lambda\delta$. In fact, much better couplings are possible. Recall that $X_{\le t}$ denotes the first t generations of $\mathfrak{X}_\lambda$, seen as a rooted tree.

Lemma 4.7. *Let $\lambda = 1+\varepsilon$, where $\varepsilon = \varepsilon(n) = O(1)$. Let $\delta(n)$ be any function with $\delta > 0$ and $\delta \to 0$ as $n \to \infty$. Let $t = t(n) \ge 0$ and let $T = T(n)$ be a rooted tree of height t with $\varepsilon|T|^2 \le \delta n$, each generation of size at most $n^{1/3}$, and $|T| \le \delta n^{2/3}$. Then*

$$\mathbb{P}\big(G^0_{\le t}(x) \cong T\big) \sim \mathbb{P}\big(X_{\le t} \cong T\big)$$

and

$$\mathbb{P}\big(G_{\le t}(x) \cong T\big) \sim \mathbb{P}\big(X_{\le t} \cong T\big),$$

where the asymptotics is uniform over all such sequences $T(n)$.

Proof. Rather than couple, we simply calculate directly; it is convenient to order the vertices first. When constructing $\mathfrak{X}_\lambda$ starting from X_0, let us number the particles $1, 2, 3, \ldots$ in the order they appear, so the initial particle is particle 1, and test particles in numerical order to see how many children they have. We number the vertices uncovered in the neighbourhood exploration process by which we find $G^0_{\le t}(x)$ analogously, this time using any (deterministic or random) rule to decide in which order to number the children of a vertex.

For each numbering $T^\star$ of T that can arise in such an exploration, let $E_1(T^\star)$ be the event that $X_{\le t}$ is isomorphic to $T^\star$ with the labels matching. Then $\{X_{\le T} \cong T\}$ is the disjoint union of the events $E_1(T^\star)$, where $T^\star$ runs over all numberings of T; note that these events are equiprobable. Similarly, let $E_2(T^\star)$ be the event that $G^0_{\le t}(x) \cong T^\star$ with labels matching, so $\{G^0_{\le t}(x) \cong T\}$ is the disjoint union of the $E_2(T^\star)$. Fix one particular numbering $T^\star$. Since the probabilities of $E_1(T^\star)$ and $E_2(T^\star)$ do not depend on the numbering, it suffices to show that $\mathbb{P}(E_1(T^\star)) \sim \mathbb{P}(E_2(T^\star))$.

Let r be the number of vertices of T at distance t from the root, and $m = |T|$ the total number of vertices. For $1 \le i \le m-r$, let d_i denote the number of children in T of the ith vertex. Now $E_1(T^\star)$ is simply the event that for $i = 1, \ldots, m-r$, the ith particle of the branching process has exactly d_i children. Thus,

$$\mathbb{P}(E_1(T^\star)) = \prod_{i=1}^{m-r} \frac{\lambda^{d_i}}{d_i!} e^{-\lambda}.$$

Similarly, $E_2(T^\star)$ is the event that for every i, when exploring the neighbours of the ith vertex reached, we find exactly d_i new neighbours. Let $u_i = 1 + \sum_{j<i} d_j$ denote the number

of vertices already 'used' (reached) at the point that we look for new neighbours of the ith vertex. Then

$$\mathbb{P}(E_2(T^\star)) = \prod_{i=1}^{m-r} \mathbb{P}\big(\mathrm{Bi}(n-u_i, \lambda/n) = d_i\big) = \prod_{i=1}^{m-r} \frac{(n-u_i)_{(d_i)}}{d_i!} (\lambda/n)^{d_i}(1-\lambda/n)^{n-u_i-d_i}.$$

Hence,

$$\rho = \frac{\mathbb{P}(E_2(T^\star))}{\mathbb{P}(E_1(T^\star))} = \prod_{i=1}^{m-r} \frac{(n-u_i)_{(d_i)}}{n^{d_i}} \frac{(1-\lambda/n)^{n-u_i-d_i}}{e^{-\lambda}}.$$

Since $n - u_i \geq n/2$ and d_i is bounded by $n^{1/3}$, we have $n - u_i - j = (n-u_i)e^{O(n^{-2/3})}$ for $0 \leq j \leq d_i$, so

$$\begin{aligned} \log\left(\frac{(n-u_i)_{(d_i)}}{n^{d_i}}\right) &= \log\left(\frac{(n-u_i)^{d_i}}{n^{d_i}}\right) + O(n^{-2/3}d_i) \\ &= -\frac{u_i d_i}{n} + O(d_i(u_i/n)^2) + O(n^{-2/3}d_i) = -\frac{u_i d_i}{n} + O(n^{-2/3}d_i), \end{aligned}$$

using $u_i \leq |T| \leq n^{2/3}$ in the last step. Also,

$$\begin{aligned} (n-u_i-d_i)\log(1-\lambda/n) &= (n-u_i-d_i)(-\lambda/n + O(1/n^2)) \\ &= -\lambda + u_i\lambda/n + O(d_i/n) + O(n^{-1}). \end{aligned}$$

Hence,

$$\log\rho = \sum_{i=1}^{m-r}\left(u_i\frac{\lambda-d_i}{n} + O(n^{-2/3}d_i) + O(n^{-1})\right) = o(1) + \sum_{i=1}^{m-r} u_i\frac{\lambda-d_i}{n},$$

using $\sum_i d_i = m-1 = o(n^{2/3})$.

Now $\lambda = 1+\varepsilon$, and $\sum u_i \leq m^2$. By assumption $\varepsilon m^2 = o(n)$, so

$$\log\rho = o(1) + \sum_{i=1}^{m-r} u_i\frac{1-d_i}{n} = o(1) + \sum_{i=1}^{m-r} u_i\frac{1}{n} - \sum_{i=1}^{m-r} u_i\frac{d_i}{n}.$$

We can rewrite the final sum as $\sum_{i=1}^{m-r}\sum_j u_i/n$, where j runs over the children of i. Each j in the range 2 up to m appears exactly once in the double sum, so the sum is equal to $\sum_{j=2}^m u_{j'}/n$, where j' is the parent of j. For any vertex j, the vertex j' is in the generation before j, so $u_j - u_{j'}$ is at most twice the maximum number of vertices in a generation. We have assumed this maximum is at most $n^{1/3}$, so $|u_j - u_{j'}| \leq 2n^{1/3}$. Hence,

$$\begin{aligned} \log\rho &= o(1) + \sum_{i=1}^{m-r}\frac{u_i}{n} - \sum_{i=2}^{m}\frac{u_{i'}}{n} \\ &= o(1) + \frac{u_1}{n} + \sum_{i=2}^{m-r}\frac{u_i - u_{i'}}{n} - \sum_{i=m-r+1}^{m}\frac{u_{i'}}{n} \\ &= o(1) + o(1) + \sum_i O(n^{-2/3}) - O(rm/n) = o(1) + O(rm/n) = o(1), \end{aligned}$$

and the first statement follows.

For the second, it suffices to prove that $\mathbb{P}\big(G_{\le t}(x) \cong T \bigm| G^0_{\le t}(x) \cong T\big) \sim 1$. But this is immediate since there are at most $2n^{1/3}|T| = o(n)$ extra edges that we must test. $\square$

For any fixed k, Lemma 4.7 extends to k starting vertices and k trees, with virtually the same proof.

Lemma 4.8. *Fix $k \ge 2$. Let $\lambda = 1 + \varepsilon$, where $\varepsilon = \varepsilon(n) = O(1)$. Let $\delta(n)$ be any function with $\delta > 0$ and $\delta \to 0$ as $n \to \infty$. Let $T_1, \dots, T_k$ be rooted trees, with $\varepsilon |T_i|^2 \le \delta n$, each generation of T_i of size at most $n^{1/3}$, and $|T_i| \le \delta n^{2/3}$. Given distinct vertices $x_1, \dots, x_k$ of $G = G(n, \lambda/n)$, let $E = E(x_1, \dots, x_k, T_1, \dots, T_k)$ denote the event that $G_{\le t_i}(x_i) \cong T_i$ for $1 \le i \le k$, and $d(x_i, x_j) > t_i + t_j$ for $1 \le i < j \le k$, where t_i is the height of T. Then*

$$\mathbb{P}(E) \sim \prod_{i=1}^{k} \mathbb{P}\big(X_{\le t_i} \cong T_i\big),$$

where the asymptotics is uniform over all choices of $T_1, \dots, T_k$. $\square$

In other words, the event that the t_i-neighbourhood of each x_i is isomorphic to T_i, and these neighbourhoods are disjoint, has asymptotically the probability suggested by independent branching processes. One can prove Lemma 4.8 by adapting the proof of Lemma 4.7 in the obvious ways. Alternatively, it follows from Lemma 4.7 by simple calculations.

4.3 Slow initial growth: the branching process

In this subsection we study the probability that the branching process $\mathfrak{X}_\lambda$ survives, but takes much longer than usual to reach generations of some large size. One might expect the results we need to be in the branching process literature, and perhaps they are. However, we have not found them. The key point is that here λ is variable, tending down to 1 from above, so results for fixed λ are not of much use. Furthermore, although there is a natural scaling limit as $\lambda \to 1$ from above (described below), results about this limit are not directly applicable either: we wish to consider events of probability around $1/n$, and this probability tends to 0 as $\lambda \to 1$. In other words, we need explicit bounds on the rate of convergence of some properties of the branching process as $\lambda \to 1$. Fortunately, as is often the case, the branching process results we need are not hard to prove directly.

With $\lambda = 1 + \varepsilon > 1$ fixed for the moment, let $\mathfrak{X}_\lambda = (X_t)_{t \ge 0}$ and $\mathfrak{X}^+_\lambda = (X^+_t)_{t \ge 0}$ be defined as before, so $X^+_t \subset X_t$ is the set of particles in X_t which have descendants in all future generations. Recall that $\mathfrak{X}^+_\lambda$ is again a Galton–Watson branching process, with $|X^+_0| = 1$ or 0 depending on whether $\mathfrak{X}_\lambda$ survives, and with offspring distribution Z_λ. Here, as before, Z_λ denotes the distribution of a Poisson $\mathrm{Po}(s\lambda)$ random variable conditioned to be at least 1, where $s = s(\lambda)$ is the probability that $\mathfrak{X}_\lambda$ survives forever.

From standard results (see Athreya and Ney [3], for example), we have $|X_t|/\lambda^t \to Y = Y_\lambda$ a.s., and $|X^+_t|/\lambda^t \to Y^+ = Y^+_\lambda$ a.s., for some random variables Y_λ and Y^+_λ. Our first (standard, trivial) observation is that Y_λ and Y^+_λ coincide up to a constant factor.

Lemma 4.9. *We have $Y_\lambda^+ = sY_\lambda$ a.s.*

Proof. Fix $\delta > 0$, and let $N = N(\delta)$ be a suitably chosen large integer. From standard results, with probability 1, either $\mathfrak{X}_\lambda$ dies out, or there is some minimal t with $|X_t| \ge N$. Choosing N large enough, at this time t the inequalities $||X_t|/\lambda^t - Y| \le \delta$ and $||X_t^+|/\lambda^t - Y^+| \le \delta$ hold with probability at least $1-\delta$. But t is a stopping time, so given t and $|X_t|$, each particle in X_t survives independently with probability s, and from the Chernoff bounds, provided N was chosen large enough, the ratio between $|X_t^+|$ and $|X_t|$ is within a factor $1 \pm \delta$ of s with probability at least $1-\delta$.

Hence, with probability at least $1-3\delta$ either $Y_\lambda^+ = Y_\lambda = 0$, or $Y_\lambda^+ = (1+O(\delta))sY_\lambda + O(\delta)$. The result follows by letting $\delta \to 0$. □

Our ultimate aim is to estimate $\mathbb{P}(0 < |X_t| < \omega)$ in the range of parameters where this probability is very small (around $1/(\varepsilon^2 n)$, it will turn out). Essentially, this reduces to estimating the lower tail of Y; in the light of Lemma 4.9, we may study Y^+ instead. This turns out to be easier, since (X_t^+) is in some sense 'better behaved' than (X_t) when $\varepsilon \to 0$.

When studying $\mathfrak{X}_\lambda^+ = (X_t^+)$ it makes sense to condition on the event that X_0^+ is non-empty, i.e., that $\mathfrak{X}_\lambda$ survives. Let us write $\tilde{\mathfrak{X}}_\lambda^+ = (\tilde{X}_t^+)_{t\ge 0}$ for the conditioned process, i.e., a Galton–Watson process with offspring distribution Z_λ started with a single particle. Let $\tilde{Y}^+$ denote $\lim_{t\to\infty} |\tilde{X}_t^+|/\lambda^t$, which exists a.s. Thus $\tilde{Y}^+$ is simply Y^+ conditioned on $Y^+ > 0$, up to a set of measure 0. By standard results, $\tilde{Y}^+$ is a continuous random variable with strictly positive density on $(0,\infty)$.

It turns out that we will need both upper and lower tail bounds on $\tilde{Y}^+ = \tilde{Y}_\lambda^+$. The dependence of these bounds on $\varepsilon = \lambda - 1$ is very important. We start with the upper tail.

Lemma 4.10. *There is an absolute constant $c > 0$ such that for any $1 < \lambda < 2$ and any $x > 0$ we have $\mathbb{P}(\tilde{Y}_\lambda^+ > x) \le 2e^{-cx}$.*

Proof. Recall that Z_λ denotes a Poisson distribution with mean $s\lambda$ conditioned to be at least 1, where $s = s(\lambda)$ is the positive solution to $1 - s = e^{-\lambda s}$. Set

$$f_\lambda(x) = \mathbb{E}(x^{Z_\lambda}) = \sum_{k\ge 1} x^k \frac{(s\lambda)^k e^{-s\lambda}}{k!(1-e^{-s\lambda})} = \frac{(e^{xs\lambda}-1)e^{-s\lambda}}{s} = \frac{e^{(x-1)s\lambda} - e^{-s\lambda}}{s}.$$

Note that $f_\lambda(1) = 1$, and, expanding about $x = 1$, we have

$$f_\lambda(x) = 1 + \lambda(x-1) + \frac{s\lambda^2}{2}(x-1)^2 + \cdots = 1 + \lambda(x-1) + O(s\lambda^2(x-1)^2),$$

provided $s\lambda(x-1)$ is bounded. More precisely, recalling that $s \sim 2\varepsilon$ as $\lambda \to 1$, and that $\lambda \le 2$, it is easy to check that if $0 \le x \le 2$, say, then we have

$$f_\lambda(x) \le 1 + \lambda(x-1) + C_1\varepsilon(x-1)^2 \tag{4.14}$$

for some absolute constant C_1, which we shall take to be at least 1.

Suppressing the dependence on λ in the notation, for $t \geq 0$ let $g_t(\theta) = \mathbb{E}(e^{\theta|\tilde{X}_t^+|/\lambda^t}) - 1$. Since $|\tilde{X}_0^+|$ is always 1, we have $g_0(\theta) = e^\theta - 1 = \theta + O(\theta^2)$ for θ bounded; in particular,

$$g_0(\theta) \leq \theta + \theta^2 \tag{4.15}$$

if $\theta \leq 1$.

Given $N = |\tilde{X}_1^+|$, the conditional distribution of $|\tilde{X}_{t+1}^+|$ is simply the sum of N independent copies of $|\tilde{X}_t^+|$, so

$$\begin{aligned} g_{t+1}(\theta) = \mathbb{E}\big(\mathbb{E}(e^{\theta|\tilde{X}_{t+1}^+|/\lambda^{t+1}} \mid N)\big) - 1 &= \mathbb{E}\big(\mathbb{E}(e^{\theta|\tilde{X}_t^+|/\lambda^{t+1}})^N\big) - 1 \\ &= \mathbb{E}\big((1 + g_t(\theta/\lambda))^N\big) - 1 = f_\lambda(1 + g_t(\theta/\lambda)) - 1, \end{aligned}$$

since $N = |\tilde{X}_1^+| \sim Z_\lambda$. With θ and t fixed, set $y_r = g_r(\theta/\lambda^{t-r})$, so $y_t = g_t(\theta)$, $y_0 = g_0(\theta/\lambda^t)$, and $y_{r+1} = f_\lambda(1 + y_r) - 1$ for $0 \leq r \leq t-1$. From (4.14), if $y_r \leq 1$, then

$$y_{r+1} \leq \lambda y_r + C_1\varepsilon y_r^2 \leq \lambda y_r(1 + C_1\varepsilon y_r). \tag{4.16}$$

Suppose $\theta \leq 1/(100C_1) \leq 1/100$. Then we claim that

$$y_r \leq 2\theta/\lambda^{t-r} \tag{4.17}$$

holds for $r = 0, 1, \ldots, t$. This is certainly true for $r = 0$, since $y_0 = g_0(\theta/\lambda^t) \leq (\theta/\lambda^t)(1 + \theta/\lambda^t) \leq 2\theta/\lambda^t$. If (4.17) holds for $r = 0, 1, \ldots, s-1$, then in particular $y_r \leq 1$ for $r < s$, so from (4.15) and (4.16) we have

$$y_s = y_0 \prod_{r<s} \frac{y_{r+1}}{y_r} \leq \frac{\theta}{\lambda^t}(1 + \theta/\lambda^t)\lambda^s \prod_{r<s}(1 + C_1\varepsilon y_r) \leq \frac{\theta}{\lambda^{t-s}} \exp\left(\frac{\theta}{\lambda^t} + C_1\varepsilon \sum_{r<s} y_r\right).$$

Using (4.17) for $r < s$, we have $\sum_{r<s} y_r \leq \sum_{0\leq r\leq t} 2\theta/\lambda^{t-r} \leq 2\theta \sum_{r\geq 0} \lambda^{-r} = 2\theta(1+\varepsilon)/\varepsilon$. Since $\theta \leq 1/(100C_1)$, (4.17) for $r = s$ follows, completing the proof of (4.17) by induction.

Setting $r = t$ in (4.17), we have in particular that $y_t = g_t(\theta) \leq 1/(50C_1) \leq 1$ for $\theta \leq \theta_0 = 1/(100C_1)$. Hence the moment generating functions $\mathbb{E}(e^{\theta|\tilde{X}_t^+|/\lambda^t})$ are uniformly bounded by 2 for all $1 < \lambda \leq 2$ and all $\theta \leq \theta_0$.

With λ fixed, we have $|\tilde{X}_t^+|/\lambda^t \to \tilde{Y}^+$ a.s. By Fatou's Lemma, it follows that

$$\mathbb{E}\big(e^{\theta\tilde{Y}^+}\big) \leq \liminf_{t\to\infty} \mathbb{E}\big(e^{\theta|\tilde{X}_t^+|/\lambda_t}\big) = \liminf_{t\to\infty} g_t(\theta) \leq 2.$$

Applying Markov's inequality, it follows that for any x we have $\mathbb{P}(\tilde{Y}^+ \geq x) \leq 2e^{-\theta x}$, completing the proof of the lemma. □

Our main application of the upper tail bound above is to show that the sum of many independent copies of $\tilde{Y}^+$ is tightly concentrated.

Lemma 4.11. *Let c, A and δ be positive constants. There is a constant $\alpha = \alpha(c, A, \delta) > 0$ such that, if Z is any random variable satisfying the tail bound $\mathbb{P}(|Z - \mathbb{E}\,Z| > x) \leq Ae^{-cx}$ for all $x > 0$, and S_n is the sum of n independent copies of Z, then*

$$\mathbb{P}\big(|S_n/n - \mathbb{E}\,Z| \geq \delta\big) \leq e^{-\alpha n}$$

for all $n \geq 1$.

Proof. For $|\theta| < c$, let $\phi(\theta) = \mathbb{E}(e^{\theta(Z-\mu-\delta)})$ where $\mu = \mathbb{E}\, Z$; the tail bound on Z implies that $\phi(\theta)$ is finite. Then $\phi(0) = 1$, $\phi'(0) = -\delta$, and

$$\phi''(\theta) = \mathbb{E}\big((Z-\mu-\delta)^2 e^{\theta(Z-\mu-\delta)}\big),$$

which is bounded by a constant due to the tail bound. Hence there are positive constants c'' (which we may take smaller than c/δ) and c' such that $\phi(\theta) < 1 - c'\delta^2$ when $\theta = c''\delta$. Now with $Z_1, \dots, Z_n$ independent copies of Z and $S_n = \sum Z_i$, we have

$$\begin{aligned}
\mathbb{P}(S_n \geq n(\mu+\delta)) &\leq \mathbb{E}\, e^{\theta S_n} e^{-\theta(\mu+\delta)n} \\
&= \mathbb{E}\, e^{\sum_i \theta(Z_i-\mu-\delta)} \\
&= \phi(\theta)^n \quad \text{by independence of the } Z_i \\
&< (1-c'\delta^2)^n < e^{-c'\delta^2 n}.
\end{aligned}$$

An exponential upper bound on $\mathbb{P}(S_n \leq n(\mu-\delta))$ is obtained by considering $\hat{\phi}(\theta) = \mathbb{E}(e^{\theta(\mu-\delta-Z)})$ similarly. □

Using Lemma 4.11 it is easy to show that up to an error probability that is exponentially small in ω, the martingale $|\tilde{X}_t^+|/\lambda^t$ has essentially converged to its (almost sure) limit $\tilde{Y}^+$ by the time that $|\tilde{X}_t^+|$ first reaches size ω. As before, it is crucial that the concentration we obtain is uniform in λ as $\lambda \searrow 1$.

Lemma 4.12. *Let $0 < \delta < 1$, $1 < \lambda \leq 2$ and $\omega \geq 1$ be given, let $t_\omega = \min\{t : |\tilde{X}_t^+| \geq \omega\}$, whenever this is defined, and let E be the event that $|\tilde{X}_t^+|/\lambda^t$ is within a factor $1 \pm \delta$ of $\tilde{Y}^+$ for all $t \geq t_\omega$.*

Then t_ω is defined with probability 1, and $\mathbb{P}(E) = 1 - e^{-\Omega(\omega)}$, where the implicit constant depends on δ but not on λ.

Proof. The sequence $|\tilde{X}_t^+|$ is non-decreasing, and increases with probability bounded away from zero (at least $\mathbb{P}(Z_\lambda > 1)$) at each step, so $\tilde{X}_t^+ \to \infty$ a.s., and t_ω is indeed defined with probability 1.

Let A be the event

$$A = \left\{(1-\delta/10)|\tilde{X}_{t_\omega}^+|/\lambda^{t_\omega} \leq \tilde{Y}^+ \leq (1+\delta/10)|\tilde{X}_{t_\omega}^+|/\lambda^{t_\omega}\right\}.$$

Our first aim is to show that A is very likely to hold. Let us condition on the event $t_\omega = t$, where $t \geq 0$, and also on $|\tilde{X}_t^+|$. Since t_ω is a stopping time, given that $t_\omega = t$ and $|\tilde{X}_t^+| = m$, the descendants of the $m \geq \omega$ particles in $\tilde{X}_t^+$ form independent copies of the original process. Let $n_{t',i}$ denote the number of descendants in generation t' of the ith particle in $\tilde{X}_t^+$. Then for each i we have $n_{t',i}/\lambda^{t'-t} \to \tilde{Y}_i^+$ a.s., where the $\tilde{Y}_i^+$ are independent and have the distribution of $\tilde{Y}^+$. It follows that $\tilde{Y}^+ = \sum_{i=1}^m \tilde{Y}_i^+/\lambda^{t_\omega}$ a.s.

Now $\tilde{Y}^+$ has mean 1, and $m \geq \omega$. Applying Lemmas 4.10 and 4.11, we see that

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^m \tilde{Y}_i^+ - 1\right| \geq \delta/10\right) = e^{-\Omega(m)} = e^{-\Omega(\omega)},$$

so $\mathbb{P}(A) = 1 - e^{-\Omega(\omega)}$.

Let B_- be the event that t_ω is defined, and there is some $t > t_\omega$ for which $|\tilde{X}_t^+|/|\tilde{X}_{t_\omega}^+| \le (1-\delta/2)\lambda^{t-t_\omega}$. If B_- holds, let t_1 be the first such time. Then t_1 (which is not always defined) is again a stopping time so, arguing as above, given that $t_1 = t$ and $|\tilde{X}_{t_1}^+| = m$, we have $\tilde{Y}^+ = \sum_{i=1}^m \tilde{Y}_i^+/\lambda^{t_1}$, where the $\tilde{Y}_i^+$ are iid with the distribution of $\tilde{Y}^+$. This also holds if we condition on the entire history up to time t_1, and in particular on t_ω and $r = |\tilde{X}_{t_\omega}^+|$.

By definition of B_- we have $m \le m_0 = (1-\delta/2)\lambda^{t-t_\omega} r$, so recalling that the $\tilde{Y}_i^+$ are independent copies of $\tilde{Y}^+$,

$$\mathbb{P}\left(\sum_{i=1}^m \tilde{Y}_i^+ \ge (1+\delta/10)m_0\right) \le \mathbb{P}\left(\sum_{i=1}^{\lfloor m_0 \rfloor} \tilde{Y}_i^+ \ge (1+\delta/10)m_0\right) = e^{-\Omega(m_0)} = e^{-\Omega(\omega)},$$

using Lemma 4.11 and the fact that $\lambda^{t-t_\omega} r \ge r \ge \omega$ for the last step. If B_- and A both hold, then the event appearing on the left above also holds, so we have shown that $\mathbb{P}(A \mid B_-) = e^{-\Omega(\omega)}$. Hence, $\mathbb{P}(A \cap B_-) \le \mathbb{P}(A \mid B_-) = e^{-\Omega(\omega)}$.

Define B_+ to be the event that there is some $t > t_\omega$ for which $|\tilde{X}_t^+|/|\tilde{X}_{t_\omega}^+| \ge (1+\delta/2)\lambda^{t-t_\omega}$. A similar but simpler argument shows that $\mathbb{P}(A \cap B_+) = e^{-\Omega(\omega)}$. Hence with probability $1 - e^{-\Omega(\omega)}$ the event A holds, while neither B_- nor B_+ does, and the lemma follows. □

Returning to the original branching process $\mathfrak{X}_\lambda = (X_t)$, recall that this survives with probability $s = s(\lambda) = \Theta(\varepsilon)$, where $\lambda = 1 + \varepsilon$. Recall also that $X_t/\lambda^t \to Y$ a.s., where by standard results $\mathbb{E}(Y) = 1$, and $Y = 0$ if and only if the process dies out, so $\mathbb{P}(Y \ne 0) = s$. Also, recalling that $\tilde{Y}^+$ has the distribution of Y^+ conditioned on $Y^+ > 0$, Lemma 4.9 implies that the distribution of sY given that $Y \ne 0$ is exactly the distribution of $\tilde{Y}^+$.

The next lemma will be similar to Lemma 4.12, but concerning $\mathfrak{X}_\lambda = (X_t)$. This will lead us to consider the sum S_N of N independent copies Y_i of Y, for ω large and $N \ge \omega/\varepsilon$. Given $0 < \delta < 1$, from concentration of the binomial distribution, with probability $1 - e^{-\Omega(\omega)}$ the number M of i with $Y_i \ne 0$ is within a factor $1 \pm \delta$ of its mean $sN = \Omega(\omega)$. Conditional on M, the variable sS_N is the sum of M independent copies of sY each conditioned to be positive, or equivalently of M independent copies of $\tilde{Y}^+$, so by Lemma 4.11, with probability $1 - e^{-\Omega(M)}$ this sum is within a factor $1 \pm \delta$ of its mean M. It follows that with probability $1 - e^{-\Omega(\omega)}$ we have $|S_N/N - 1| \le 3\delta$, say. Using this fact in place of concentration of the sum of ω copies of $\tilde{Y}^+$, the proof of Lemma 4.12 gives the following result, which is more or less a sharpening of Lemma 4.6. Recall that $Y = \lim_{t\to\infty} |X_t|/\lambda^t$.

Lemma 4.13. *Let $\delta > 0$, $1 < \lambda \le 2$ and $\omega \ge 1$ be given, and set $\varepsilon = \lambda - 1$. Let $t_{\omega/\varepsilon} = \min\{t : |X_t| \ge \omega/\varepsilon\}$, whenever this is defined, let $\mathcal{S}_\omega$ be the event that $t_{\omega/\varepsilon}$ is defined, let $E \subset \mathcal{S}_\omega$ be the event that $|X_t|/\lambda^t$ is within a factor $1 \pm \delta$ of Y for all $t \ge t_{\omega/\varepsilon}$, and let $\mathcal{S} = \{\forall t : |X_t| > 0\}$ be the event that the process survives.*

Then $\mathbb{P}(E \mid \mathcal{S}_\omega) = 1 - e^{-\Omega(\omega)}$, where the implicit constant depends on δ but not on λ. Furthermore, $\mathbb{P}(\mathcal{S} \mid \mathcal{S}_\omega) = 1 - e^{-\Omega(\omega)}$ and $\mathbb{P}(\mathcal{S}_\omega \setminus E) = O(\varepsilon e^{-\Omega(\omega)})$.

Proof. The first statement follows by modifying the proof of Lemma 4.12 as described above. The second is an immediate consequence (and also easy to verify directly). It implies in particular that $\mathbb{P}(\mathcal{S})/\mathbb{P}(\mathcal{S}_\omega)$ is bounded below, so $\mathbb{P}(\mathcal{S}_\omega) = O(\mathbb{P}(\mathcal{S})) = O(\varepsilon)$. The final statement then follows from the first. □

Lemma 4.13 tells us that for ω large enough, the probability that the branching process takes much longer than expected to reach size ω/ε is essentially determined by the tail of the distribution of $Y = Y_\lambda$ near 0. Lemma 4.12 will be useful in studying this tail indirectly.

Writing $\mathbb{R}^+$ for the set of non-negative reals, let $(\mathcal{Y}_t)_{t\in\mathbb{R}^+}$ be a standard Yule process. Thus $\mathcal{Y}_0$ consists of a single particle, and each particle in the process survives forever and gives rise to children according to a Poisson process with rate 1, independently of the other particles and of the history. Note that $|\mathcal{Y}_t|$ is a (random) non-decreasing function of t, and that $\mathbb{E}(|\mathcal{Y}_t|) = e^t$. It is well known that $\lim_{t\to\infty} |\mathcal{Y}_t|/e^t$ exists with probability 1 (see, for example, [3, Section III.7]); we denote this (random) limit by W.

It is not hard to see that as λ decreases to 1, the suitably rescaled process $\tilde{\mathfrak{X}}^+_\lambda$ converges in some sense to $(\mathcal{Y}_t)$. All we shall need is a very weak result of this form.

Lemma 4.14. *Let $T > 0$ be fixed. As $\lambda = 1 + \varepsilon$ tends to 1 from above, the distribution of $|\tilde{X}^+_{\lfloor T/\varepsilon\rfloor}|$ converges to that of $|\mathcal{Y}_T|$.*

Proof. We take snapshots of $(\mathcal{Y}_t)$ at times separated by ε, i.e., consider $Y_n = \mathcal{Y}_{n\varepsilon}$, $n = 0, 1, \ldots, T$. Each particle x in Y_n always survives to Y_{n+1}, has no children in Y_{n+1} with probability $\mathbb{P}(\mathrm{Po}(\varepsilon) = 0) = e^{-\varepsilon} = 1 - \varepsilon + O(\varepsilon^2)$, and has exactly one child in Y_{n+1} with probability $\mathbb{P}(\mathrm{Po}(\varepsilon) = 1) = \varepsilon + O(\varepsilon^2)$. Furthermore, the probability that this child (if it exists) has children of its own by time $(n+1)\varepsilon$ is $O(\varepsilon)$. Hence, the number Z' of descendants of x in Y_{n+1} is 1 with probability $1 - \varepsilon + O(\varepsilon^2)$, two with probability $\varepsilon + O(\varepsilon^2)$ and three or more with probability $O(\varepsilon^2)$. Hence Z' and Z_λ, the offspring distribution in $\tilde{\mathfrak{X}}^+_\lambda$, can be coupled to agree with probability $1 - O(\varepsilon^2)$.

Using the independence properties of $\tilde{\mathfrak{X}}^+_\lambda$ and of $(\mathcal{Y}_t)$, it follows that these processes can be coupled so that the event $E = \{|\tilde{X}^+_n| = |Y_n|, n = 0, 1, \ldots, \lfloor T/\varepsilon\rfloor\}$ *fails* to hold with probability at most

$$O(\varepsilon^2) \sum_{n\le T/\varepsilon} \mathbb{E}(|Y_n|) = O(\varepsilon^2) \sum_{n\le T/\varepsilon} e^{\varepsilon n} = O(\varepsilon)e^T = O(\varepsilon).$$

Since $\mathcal{Y}_{\varepsilon\lfloor T/\varepsilon\rfloor} = \mathcal{Y}_T$ with probability $1 - O(\varepsilon)$, the result follows. □

Corollary 4.15. *As $\lambda = 1 + \varepsilon$ tends to 1 from above, $\tilde{Y}^+_\lambda$ converges in distribution to W.*

Proof. Fix $\delta > 0$. It suffices to show that for ε sufficiently small we can couple $\tilde{Y}^+_\lambda$ and W so that they agree within a factor of $1 + O(\delta)$ with probability $1 - O(\delta)$.

Let ω be a constant to be chosen below, depending on δ but not on ε. Since $|\mathcal{Y}_t| \to \infty$ with probability 1, there is some T such that $\mathbb{P}(|\mathcal{Y}_T| < \omega) \le \delta$. From Lemma 4.14, if ε is sufficiently small, then we may couple $\tilde{\mathfrak{X}}^+_\lambda$ and $(\mathcal{Y}_t)$ so that with probability at least $1 - \delta$ we have $|\mathcal{Y}_T| = |\tilde{X}^+_{\lfloor T/\varepsilon\rfloor}|$. Then with probability at least $1 - 2\delta$ we have $|\mathcal{Y}_T| = |\tilde{X}^+_{\lfloor T/\varepsilon\rfloor}| \ge \omega$.

Let $n = \lfloor T/\varepsilon\rfloor$. Applying Lemma 4.12, it follows that if ω is chosen large enough (depending only on δ, not on ε), then with probability at least $1 - 3\delta$ the limit $\tilde{Y}^+_\lambda$ is within a factor $1 \pm \delta$ of $|\tilde{X}^+_n|/\lambda^n$. A similar result holds for $(\mathcal{Y}_t)$. (Indeed, since $|\mathcal{Y}_t|/e^t \to W$ a.s., there must be some constant T' such that with probability $1 - \delta$ we have $|\mathcal{Y}_t|/e^t$ within a factor $1 \pm \delta$ of W for all $t \ge T'$.) In particular, if ω is large enough, then with probability $1 - \delta$ the ratio $|\mathcal{Y}_T|/e^T$ is within a factor of $(1 \pm \delta)$ of W.

Putting the pieces together, and noting that $\lambda^n = (1+\varepsilon)^{\lfloor T/\varepsilon \rfloor} = e^T + O(\varepsilon)$, for ε small enough the quantities $\tilde{Y}_\lambda^+$, $|\tilde{X}_n^+|/\lambda^n$, $|\tilde{X}_n^+|/e^T$, $|\mathcal{Y}_T|/e^T$ and W agree up to factors of $1+O(\delta)$ with probability $1-O(\delta)$, completing the proof. □

It is well known, and not hard to check, that the (positive) random variable W associated to the Yule process has an exponential distribution with mean 1; this is an exercise in [3], for example. In particular,

$$\mathbb{P}(W \leq x) = 1 - e^{-x} \sim x \tag{4.18}$$

as $x \to 0$ from above. We are now ready to prove our bound on the lower tail of $\tilde{Y}_\lambda^+$.

Theorem 4.16. *Let $\lambda = 1+\varepsilon$. As ε and x tend to 0 from above we have*

$$\mathbb{P}(\tilde{Y}_\lambda^+ \leq x) \sim x^{\log(1/\lambda_\star)/\log\lambda}.$$

Note that we make no assumption on the relative rates at which ε and x tend to zero. With x fixed, the result would be immediate from (4.18) and Corollary 4.15.

Proof. Let $\delta > 0$ be given. We must show that there are constants $x_0 = x_0(\delta)$ and $\varepsilon_0 = \varepsilon_0(\delta)$ such that for all $0 < \varepsilon < \varepsilon_0$ and $0 < x < x_0$ we have $\mathbb{P}(\tilde{Y}_\lambda^+ \leq x) = e^{O(\delta)} x^{\log(1/\lambda_\star)/\log\lambda}$, where the implicit constant is absolute.

By (4.18), there is an $x_1 > 0$ such that for all $x \leq x_1$ we have

$$e^{-\delta} \leq \mathbb{P}(W \leq x)/x \leq e^{\delta}. \tag{4.19}$$

Fix such an x_1, and set $x_0 = \min\{x_1, \delta\}$.

Trivially, for $(1-\delta)x_0 \leq x \leq x_0$ and any λ, we have

$$\mathbb{P}\big(\tilde{Y}_\lambda^+ \leq (1-\delta)x_0\big) \leq \mathbb{P}(\tilde{Y}_\lambda^+ \leq x) \leq \mathbb{P}(\tilde{Y}_\lambda^+ \leq x_0).$$

As $\varepsilon \to 0$, from Corollary 4.15, for any constant a we have $\mathbb{P}(\tilde{Y}_\lambda^+ \leq a) \to \mathbb{P}(W \leq a)$. Applying this with $a = x_0$ and $a = (1-\delta)x_0$, it follows that there is an ε_0 such that

$$e^{-\delta}\mathbb{P}\big(W \leq (1-\delta)x_0\big) \leq \mathbb{P}(\tilde{Y}_\lambda^+ \leq x) \leq e^{\delta}\mathbb{P}(W \leq x_0) \tag{4.20}$$

for all $\varepsilon \leq \varepsilon_0$ and all x in the interval $I = [(1-\delta)x_0, x_0]$. We may and shall assume that $\varepsilon_0 < 1/10$, say. Since $\log(1/\lambda_\star)/\log\lambda \to 1$ as $\varepsilon \to 0$ (see (4.4)), reducing ε_0 if necessary, we have $x^{\log(1/\lambda_\star)/\log\lambda} = e^{O(\delta)}x$ uniformly in $x \in I$ and $\varepsilon \leq \varepsilon_0$. Using (4.19) and (4.20), it follows that

$$\mathbb{P}(\tilde{Y}_\lambda^+ \leq x) = e^{O(\delta)} x^{\log(1/\lambda_\star)/\log\lambda} \tag{4.21}$$

for all $x \in I$ and $\varepsilon \leq \varepsilon_0$, where the implicit constant is absolute.

At this point we return to the definition of $\tilde{Y}_\lambda^+$ in terms of $(\tilde{X}_t^+)$. Recall that $Z = Z_\lambda$ is a Poisson distribution with parameter $s\lambda$ conditioned on being non-zero, and that $\mathbb{E}(Z) = \lambda$ and $\mathbb{P}(Z = 1) = \lambda_\star$. Now $\tilde{Y}_\lambda^+$ has the distribution of the sum of Z independent copies $Y_1, \ldots, Y_Z$ of $\tilde{Y}_\lambda^+/\lambda$. Hence, for any x,

$$\mathbb{P}(\tilde{Y}_\lambda^+ \leq x) \geq \mathbb{P}(Z = 1,\, \tilde{Y}_\lambda^+ \leq x) = \mathbb{P}(Z = 1)\mathbb{P}(Y_1 \leq x) = \lambda_\star \mathbb{P}(\tilde{Y}_\lambda^+ \leq \lambda x).$$

Given any $x \le x_0$, there is some non-negative integer r such that $x' = x\lambda^r$ lies in I. From the inequality above it follows that $\mathbb{P}(\tilde{Y}_\lambda^+ \le x) \ge \lambda_\star^r \mathbb{P}(\tilde{Y}_\lambda^+ \le x')$. Applying (4.21) to bound the second probability, and noting that $\lambda_\star^r = \exp((-r)\log(1/\lambda_\star)) = \exp(\log(1/\lambda_\star)\log(x/x')/\log\lambda)$, it follows that

$$\mathbb{P}(\tilde{Y}_\lambda^+ \le x) \ge (x/x')^{\log(1/\lambda_\star)/\log\lambda} e^{O(\delta)} (x')^{\log(1/\lambda_\star)/\log\lambda} = e^{O(\delta)} x^{\log(1/\lambda_\star)/\log\lambda},$$

completing the proof of the lower bound.

For the upper bound we use the inequality

$$\begin{aligned}
\mathbb{P}(\tilde{Y}_\lambda^+ \le x) &= \mathbb{P}\big(Z=1,\, \tilde{Y}_\lambda^+ \le x\big) + \mathbb{P}\big(Z \ge 2,\, \tilde{Y}_\lambda^+ \le x\big) \\
&\le \mathbb{P}(Z=1)\mathbb{P}(Y_1 \le x) + \mathbb{P}(Z \ge 2)\mathbb{P}(Y_1 + Y_2 \le x) \\
&\le \mathbb{P}(Z=1)\mathbb{P}(\tilde{Y}_\lambda^+ \le \lambda x) + \mathbb{P}(Z \ge 2)\mathbb{P}(\tilde{Y}_\lambda^+ \le \lambda x)^2 \\
&= \lambda_\star \mathbb{P}(\tilde{Y}_\lambda^+ \le \lambda x)\left(1 + \frac{1-\lambda_\star}{\lambda_\star}\mathbb{P}(\tilde{Y}_\lambda^+ \le \lambda x)\right).
\end{aligned}$$

Given $x \le x_0$, as before there is a non-negative integer r such that $x' = x\lambda^r \in I$. For $0 \le i \le r$ let $x_i = x'/\lambda^i$, so $x_0 = x'$ and $x_r = x$. Let $p_i = \mathbb{P}(\tilde{Y}_\lambda^+ \le x_i)$, so

$$p_{i+1} \le \lambda_\star p_i (1 + \lambda_\star^{-1}(1-\lambda_\star)p_i)$$

and hence, by induction,

$$p_i \le \lambda_\star^i p_0 \prod_{j<i}(1 + (\lambda_\star^{-1} - 1)p_j) \le \lambda_\star^i p_0 \exp\left((\lambda_\star^{-1} - 1)\sum_{j<i} p_j\right). \tag{4.22}$$

Now $p_0 = \mathbb{P}(\tilde{Y}_\lambda^+ < x')$ and $x' \in I$, so, recalling that $x_0 \le \delta$, we have $p_0 = O(\delta)$. Thus $p_0 \le 1/10$, say, if we assume δ is small, which we may. It follows by induction on i that $p_i \le 2\lambda_\star^i p_0 \le \lambda_\star^i/5$. Indeed, if this holds for $j < i$, then the term inside the exponential in (4.22) is at most

$$(\lambda_\star^{-1} - 1)\sum_{j<i} \lambda_\star^j/5 \le \lambda_\star^{-1}(1-\lambda_\star)\sum_{j=0}^{\infty} \lambda_\star^j/5 = \lambda_\star^{-1}/5 \le 1/4,$$

and $e^{1/4} < 2$. Plugging $p_j \le 2\lambda_\star^j p_0$ back into (4.22), we see that $p_i \le \lambda_\star^i p_0 \exp(3p_0) = \lambda_\star^i p_0 e^{O(\delta)}$. Calculating as for the lower bound as above, this establishes the required upper bound. □

Remark. The method used above shows that for λ and x bounded above, $\mathbb{P}(\tilde{Y}_\lambda^+ \le x)$ is within a factor C of $x^{\log(1/\lambda_\star)/\log\lambda}$, where C depends only on the bounds we assume on λ and x. For λ constant and $x \to 0$, this is a standard result. Perhaps surprisingly, the conclusion of Theorem 4.16 does not hold in this case: as pointed out to us by Svante Janson, the limiting behaviour of $\mathbb{P}(\tilde{Y}_\lambda^+ \le x)/x^{\log(1/\lambda_\star)/\log\lambda}$ as $x \to 0$ is oscillatory. One period corresponds to changing x by a factor of λ, and the tail probability by a factor of $1/\lambda_\star$.

In the light of Lemma 4.9 and the fact that $\tilde{Y}_\lambda^+$ is just Y_λ^+ conditioned on being non-zero, an event of probability $s = s(\lambda) \sim 2\varepsilon$ by (4.3), Theorem 4.16 has the following corollary.

Corollary 4.17. *Let $\lambda = 1 + \varepsilon$. As ε and x tend to 0 from above we have*

$$\mathbb{P}\big(0 < Y_\lambda \le x/\varepsilon\big) \sim 4\varepsilon x^{\log(1/\lambda_\star)/\log\lambda}.$$

Proof. Let $s = s(\lambda)$ denote the survival probability of $\mathfrak{X}_\lambda$. Then

$$\begin{aligned}\mathbb{P}\big(0 < Y_\lambda \le x/\varepsilon\big) &= \mathbb{P}\big(0 < Y_\lambda^+ \le sx/\varepsilon\big) \\ &= \mathbb{P}(Y_\lambda^+ > 0)\mathbb{P}\big(Y_\lambda^+ \le sx/\varepsilon \mid Y_\lambda^+ > 0\big) = s\mathbb{P}\big(\tilde{Y}_\lambda^+ \le sx/\varepsilon\big),\end{aligned}$$

where the first step is from Lemma 4.9 and the rest are from the definitions. Applying Theorem 4.16, and using once again $s \sim 2\varepsilon$ and $\log(1/\lambda_\star)/\log\lambda \sim 1$, it follows that

$$\begin{aligned}\mathbb{P}(0 < Y_\lambda \le x/\varepsilon) &\sim s(sx/\varepsilon)^{\log(1/\lambda_\star)/\log\lambda} \\ &\sim 2\varepsilon\big((2+o(1))x\big)^{\log(1/\lambda_\star)/\log\lambda} \sim 4\varepsilon x^{\log(1/\lambda_\star)/\log\lambda},\end{aligned}$$

as claimed. □

In turn, Corollary 4.17 and Lemma 4.13 will give us the required estimate on the probability that the branching process $\mathfrak{X}_\lambda = (X_t)_{t\ge0}$ takes a long time to begin to have a large population.

Before turning to our tail bound, let us make a simple observation; the proof is analogous to, but simpler than, that of Lemma 4.4, so we omit it.

Lemma 4.18. *Let $\lambda = 1+\varepsilon$, $M \ge 1$ and $\delta > 0$, and let $t_M = \min\{t : |X_t| \ge M\}$, whenever this is defined. Given that t_M is defined, the probability that $|X_{t_M}|$ exceeds $(1+\delta)(1+\varepsilon)M$ is $e^{-\Omega(\delta^2 M)}$, where the implicit constant is absolute.* □

In the following result, $t_{\omega/\varepsilon}$ denotes $\min\{t : |X_t| \ge \omega/\varepsilon\}$, whenever this is defined.

Theorem 4.19. *Let $\lambda = 1+\varepsilon$, and suppose that $\varepsilon = \varepsilon(n) \to 0$, $\omega = \omega(n) \to \infty$, and $t = t(n)$ satisfy $t \le 100\log\omega/\log\lambda$ and $\varepsilon t \to \infty$. Then, with $t_1 = \log\omega/\log\lambda$, we have*

$$\mathbb{P}(t_{\omega/\varepsilon} > t_1 + t) \sim 4\varepsilon\lambda_\star^t,$$

$$\mathbb{P}\big(0 < |X_r| < \omega/\varepsilon,\ 0 \le r \le t_1 + t\big) \sim 4\varepsilon\lambda_\star^t,$$
$$\mathbb{P}\big(0 < |X_{t_1+t}| < \omega/\varepsilon\big) \sim 4\varepsilon\lambda_\star^t$$

and

$$\mathbb{P}\big((X_r) \text{ survives and } 0 < |X_{t_1+t}| < \omega/\varepsilon\big) \sim 4\varepsilon\lambda_\star^t.$$

Proof. We first show for any fixed $\delta > 0$ we have

$$\mathbb{P}\big(t_{\omega/\varepsilon} > t_1 + t\big) = (1+O(\delta))4\varepsilon\lambda_\star^t + O(\varepsilon e^{-\Omega(\omega)}), \tag{4.23}$$

$$\mathbb{P}\big(0 < |X_r| < \omega/\varepsilon,\ 0 \le r \le t_1 + t\big) = (1+O(\delta))4\varepsilon\lambda_\star^t + O(\varepsilon e^{-\Omega(\omega)}), \tag{4.24}$$

$$\mathbb{P}\big(0 < |X_{t_1+t}| < \omega/\varepsilon\big) = (1+O(\delta))4\varepsilon\lambda_\star^t + O(\varepsilon e^{-\Omega(\omega)}) \tag{4.25}$$

and

$$\mathbb{P}\big((X_r) \text{ survives and } 0 < |X_{t_1+t}| < \omega/\varepsilon\big) = (1+O(\delta))4\varepsilon\lambda_\star^t + O(\varepsilon e^{-\Omega(\omega)}). \tag{4.26}$$

Note that the events considered in (4.23) and (4.24) are not quite the same: by the event $A = \{t_{\omega/\varepsilon} > t_1 + t\}$ we mean the event that $t_{\omega/\varepsilon}$ is defined and greater than $t_1 + t$; this certainly implies the event considered in (4.24), but the latter may also hold with $T = t_{\omega/\varepsilon}$ undefined. Let $\mathcal{S}$ be the event that the process survives, noting that

$$\mathbb{P}(\mathcal{S} \mid \{t_{\omega/\varepsilon} \text{ is defined}\}) \geq 1 - (1-s)^{\omega/\varepsilon} = 1 - e^{-\Omega(\omega)}. \tag{4.27}$$

In particular, for large n this conditional probability is at least $1/2$, so the probability that T is defined is at most $2s = O(\varepsilon)$.

Let B_1 be the 'bad' event that $T = t_{\omega/\varepsilon}$ is defined and there is an $r \geq T$ with $|X_r|/\lambda^r$ outside the interval $(1 \pm \delta)Y_\lambda$. By Lemma 4.13, we have $\mathbb{P}(B_1 \mid T \text{ defined}) = e^{-\Omega(\omega)}$, so $\mathbb{P}(B_1) = O(\varepsilon e^{-\Omega(\omega)})$. Let B_2 be the event that T is defined and $|X_T| \geq (1+\delta)\omega/\varepsilon$. If ε is small enough, which we may assume, then $(1+\delta) \geq (1+\varepsilon)(1+\delta/2)$, and from Lemma 4.18 we have $\mathbb{P}(B_2) = e^{-\Omega(\omega/\varepsilon)} = O(\varepsilon e^{-\Omega(\omega)})$.

Suppose that B_1 does not hold. Then if $t_{\omega/\varepsilon}$ is defined, the process survives. Thus, off B_1, the event that T is defined coincides with $\mathcal{S}$ and hence with the event $Y_\lambda > 0$. Moreover, off $B_1 \cup B_2$, whenever $Y_\lambda > 0$ we have $Y_\lambda = (1+O(\delta))|X_T|/\lambda^T = (1+O(\delta))(\omega/\varepsilon)/\lambda^T$. Thus, off $B_1 \cup B_2$, for all sufficiently large constants a, $b > 0$,

(i) $Y_\lambda > (1+a\delta)(\omega/\varepsilon)/\lambda^{t_1+t} = (1+a\delta)/(\varepsilon\lambda^t)$ implies $T \leq t_1 + t$, and

(ii) $0 < Y_\lambda \leq (1-b\delta)/(\varepsilon\lambda^t)$ implies $T > t_1 + t$.

Since $\varepsilon t \to \infty$ we have $1/\lambda^t \to 0$. Thus using Corollary 4.17 to bound the probabilities of the events on the left in (i) and (ii) above, and recalling that $\mathbb{P}(B_1 \cup B_2) = O(\varepsilon e^{-\Omega(\omega)})$, we obtain the bound (4.23).

To deduce (4.24), it suffices to show that the probability that the indicated event holds but T is undefined is $o(\varepsilon\lambda_\star^t)$. Recall that up to probability 0 events, if $\mathcal{S}$ holds, then T is defined. So it suffices to bound the probability that $|X_{t_1+t}| > 0$ but $\mathcal{S}$ does not hold. Now

$$\mathbb{P}(|X_{t_1+t}| > 0,\, \mathcal{S}^{\mathrm{c}}) = \mathbb{P}(\mathcal{S}^{\mathrm{c}})\mathbb{P}(|X_{t_1+t}| > 0 \mid \mathcal{S}^{\mathrm{c}}) = (1-s)\mathbb{P}(|X^-_{t_1+t}| > 0) \sim \mathbb{P}(|X^-_{t_1+t}| > 0),$$

where (X^-_r) is the process conditioned on dying out, which has the distribution of $\mathfrak{X}_{\lambda_\star}$. As we shall see shortly (see Lemma 4.21), $\mathbb{P}(|X^-_a| > 0) = \Theta(\varepsilon\lambda_\star^a)$ as $\varepsilon \to 0$ and $a \to \infty$ with $a = \Omega(1/\varepsilon)$. Since $\log\lambda = \Theta(\varepsilon)$, we have $t_1 + t \geq t_1 = \Omega(1/\varepsilon)$, so

$$\mathbb{P}(|X_{t_1+t}| > 0,\, \mathcal{S}^{\mathrm{c}}) = \Theta(\varepsilon\lambda_\star^{t_1+t}) = \Theta(\varepsilon\lambda_\star^t(1/\omega)^{\log(1/\lambda_\star)/\log\lambda}) = o(\varepsilon\lambda_\star^t),$$

using $\log(1/\lambda_\star)/\log\lambda \sim 1$ (see (4.4)) and $\omega \to \infty$ for the last step. So (4.24) follows.

To see that (4.25) holds, note that we may extend the implication (i) above to imply $|X_{t_1+t}| > \omega/\varepsilon$. Also (ii) can trivially be extended to imply $|X_{t_1+t}| < \omega/\varepsilon$. Again applying Corollary 4.17 gives (4.25). Now (4.26) also follows since survival coincides with $Y_\lambda > 0$.

To deduce that the various statements in the theorem hold, observe that under the assumptions given on ε, ω and t but with $\delta > 0$ fixed, we have $\lambda_\star^t = \omega^{-O(1)}$, while any function that is $e^{-\Omega(\omega)}$ decreases faster than any power of ω, so the probabilities in (4.23)–(4.26) are all asymptotically $(1 + O(\delta))4\varepsilon\lambda_\star^t$. Since $\delta > 0$ is arbitrary, the same conclusion follows for $\delta \to 0$ slowly enough. □

Theorem 4.19 is the analogue of Lemma 2.1, giving (in the relevant range) the distribution of the time the branching process takes to grow to a certain size. It will turn out, however, that we need several further results about the branching process.

4.4 Further branching process lemmas

Theorem 4.19 gives good bounds on the probability that the branching process grows more slowly than expected. It will turn out that we also need a bound on the probability that it grows faster. Such a bound is immediate from Lemmas 4.9, 4.10 and 4.13. However (to handle the case where $\varepsilon^3 n$ grows slowly), when εt is small we shall need a bound that is stronger than the one obtained this way. This is easy to obtain directly using moment generating functions as in the proof of Lemma 4.10. Note that we study $(|X_t|)$ here rather than $(|\tilde{X}_t^+|)$.

Lemma 4.20. *Suppose that $0 < \varepsilon < 1/10$ and $\varepsilon t < 1/10$. Then for all $N \geq 20t$ we have*

$$\mathbb{P}(|X_t| \geq N) \leq t^{-1} e^{-N/(20t)}.$$

Proof. Let $m_r(\theta) = \mathbb{E}\, e^{\theta |X_r|}$ be the moment generating function of $|X_r|$, so $m_0(\theta) = e^\theta \leq 1+2\theta$ for $\theta < 1/2$. We have

$$m_{r+1}(\theta) = \mathbb{E}(m_r(\theta)^{|X_1|}) = e^{\lambda(m_r(\theta)-1)} \leq 1 + \lambda(m_r(\theta) - 1) + 2\lambda^2(m_r(\theta) - 1)^2$$

as long as $\lambda(m_r(\theta) - 1) \leq 3/2$. Let $g_r = m_r(1/(20t)) - 1$, noting that $g_0 \leq 2/(20t) = 1/(10t)$. Then, as long as $g_r \leq 2/5$, we have

$$g_{r+1} \leq \lambda g_r + 2\lambda^2 g_r^2 \leq g_r(1 + \varepsilon + 3g_r) \leq g_r \exp(\varepsilon + 3g_r).$$

We claim that for $r \leq t \leq \varepsilon^{-1}/10$ we have

$$g_r \leq g_0 \exp\Big(\varepsilon r + 3\sum_{i<r} g_i\Big) \leq g_0 \exp\big(1/10 + 3/10\big) < 2g_0 \leq 1/(10t).$$

The proof is by induction using the final bound $g_i < 1/(10t)$ for $i < r$ to establish the second inequality. Hence,

$$\mathbb{E}\big(e^{|X_t|/(20t)}\big) = 1 + g_t \leq 1 + 1/(10t).$$

Applying Markov's inequality to $e^{|X_t|/(20t)} - 1$, which is always non-negative, it follows that

$$\mathbb{P}(|X_t| \geq N) \leq \frac{1}{10t}\big(e^{N/(20t)} - 1\big)^{-1} \leq \frac{1}{5t} e^{-N/(20t)}$$

whenever $N \geq 20t$, as required. □

We next turn to various events associated to the subcritical branching process $\mathfrak{X}_{\lambda_\star} = (X_t^-)$. We start by estimating the probability that the process survives to time t, as well as a derived quantity associated to the wedge condition. If our aim is just to prove Theorem 1.3, then a considerably simpler form of the following lemma will do. However, we shall prove a more precise result useful also when it comes to studying the distribution.

Lemma 4.21. *Let $\varepsilon \to 0$ and set $s_t = s_t(\varepsilon) = \mathbb{P}(|X_t^-| > 0)$. Then for $t = o(1/\varepsilon)$ we have $s_t \sim 2/t$, while for $t \geq \varepsilon^{-2/3}$ we have $s_t \sim \frac{2\varepsilon}{\lambda_\star^{-t}-1}$. In particular, if $t = \Omega(1/\varepsilon)$, then $s_t = \Theta(\varepsilon\lambda_\star^t)$, and if $\varepsilon t \to \infty$, then $s_t \sim 2\varepsilon\lambda_\star^t$. Furthermore,*

$$\prod_{t=1}^{\infty}(1 - s_t) \sim \gamma_0\varepsilon^2$$

for some constant $\gamma_0 > 0$.

Proof. Let $\tilde{s}_t$ be the probability that a critical Poisson Galton–Watson branching process survives to time t, so $\tilde{s}_0 = 1$ and $\tilde{s}_{t+1} = 1 - \exp(-\tilde{s}_t)$. It is well known (see [29] or [3, Section I.9, Thm 1]) that $\tilde{s}_t \sim 2/t$ as $t \to \infty$, and indeed one can check that

$$\tilde{s}_t = 2t^{-1} + O(t^{-2}). \tag{4.28}$$

Moreover, $t\tilde{s}_t$ approaches 2 from below. Clearly, $s_t < \tilde{s}_t$, so we have

$$s_t < \tilde{s}_t < 2/t \tag{4.29}$$

for all $\varepsilon > 0$ and $t \geq 1$.

On the other hand, we may construct $\mathfrak{X}_{\lambda_\star}$ by first constructing a critical process, and then deleting each edge of the resulting tree with probability $1 - \lambda_\star \sim \varepsilon$, independently of all other edges. If the critical process survives to time t, then there is at least one path of length t witnessing this, and it follows that $s_t \geq \tilde{s}_t(1 - \lambda_\star)^t = \tilde{s}_t(1 - O(\varepsilon t))$. If $t = o(1/\varepsilon)$, then $(1 - \lambda_\star)^t \sim 1$, so $s_t \sim \tilde{s}_t$ and the first statement of the lemma follows. Note also for later that

$$s_t = \tilde{s}_t - O(\varepsilon t)\tilde{s}_t = \tilde{s}_t - O(\varepsilon). \tag{4.30}$$

For larger t we use the following iterative formula, obtained by considering the number of particles in X_1^- with descendants in generation $t + 1$:

$$s_{t+1} = \mathbb{P}\big(\mathrm{Po}(\lambda_\star s_t) > 0\big) = 1 - e^{-\lambda_\star s_t} = \lambda_\star s_t - \lambda_\star^2 s_t^2/2 + O(s_t^3),$$

where, since $\lambda_\star \leq 1$ and $s_t \leq 1$, the implicit constant is absolute. Note also that $s_{t+1} \leq \lambda_\star s_t$. Rewriting the formula above,

$$s_{t+1} = \lambda_\star s_t \exp\big(-\lambda_\star s_t/2 + O(s_t^2)\big). \tag{4.31}$$

We now simply 'guess' an approximate form for s_t (obtained by solving a differential equation, although things are not quite that simple): for $t \geq 1$, set

$$r_t = \frac{2(1 - \lambda_\star)}{\lambda_\star(\lambda_\star^{-t} - 1)} \sim \frac{2\varepsilon}{\lambda_\star^{-t} - 1}.$$

Since $a \geq 0$ implies $(1+a)^t - 1 \geq at$, we have $r_t \leq r_1/t = 2/t$ for all t. In particular, $r_t \leq 1/2$ for $t \geq 4$. Also,

$$\frac{\lambda_\star r_t}{r_{t+1}} = \frac{\lambda_\star(\lambda_\star^{-t-1} - 1)}{\lambda_\star^{-t} - 1} = \frac{\lambda_\star^{-t} - \lambda_\star}{\lambda_\star^{-t} - 1} = 1 + \frac{1 - \lambda_\star}{\lambda_\star^{-t} - 1} = 1 + \lambda_\star r_t/2.$$

In particular, $r_{t+1} \leq \lambda_\star r_t$. Furthermore, for $t \geq 4$, which implies $r_t \leq 1/2$, we have

$$r_{t+1} = \lambda_\star r_t(1 + \lambda_\star r_t/2)^{-1} = \lambda_\star r_t \exp\bigl(-\lambda_\star r_t/2 + O(r_t^2)\bigr). \tag{4.32}$$

Using (4.31) and (4.32), it is now not hard to show that s_t and r_t remain close for all large t. Set $T = \lfloor \varepsilon^{-2/3} \rfloor$, noting that $T \to \infty$ and $T = o(1/\varepsilon)$. Note that

$$\lambda_\star^{-T} = (1/\lambda_\star)^T = (1 + \varepsilon + O(\varepsilon^2))^T = 1 + T\varepsilon + O(T^2\varepsilon^2 + T\varepsilon^2) = 1 + T\varepsilon(1 + O(\varepsilon^{1/3})),$$

so

$$r_T, s_T = (1 + O(\varepsilon^{1/3}))2/T, \tag{4.33}$$

using (4.28) and (4.30) for s_T.

Let $\rho_t = s_t/r_t - 1$, noting that $\rho_T = O(\varepsilon^{1/3})$. Then, from (4.31) and (4.32),

$$\begin{aligned} 1 + \rho_{t+1} &= (1 + \rho_t)\exp\bigl(-\lambda_\star(s_t - r_t)/2 + O(r_t^2 + s_t^2)\bigr) \\ &= (1 + \rho_t)\exp\bigl(-\lambda_\star \rho_t r_t/2 + O(r_t^2 + s_t^2)\bigr). \end{aligned}$$

Since r_t and s_t are bounded, we have $\exp(O(r_t^2 + s_t^2)) \leq M(r_t^2 + s_t^2)$ for some absolute constant M. For ε small and $t \geq T$ we have $r_t \leq r_T \leq 1/10$, say. It follows that whatever the sign of ρ_t, the $\exp(-\lambda_\star \rho_t r_t/2)$ term 'pulls $(1 + \rho_t)$ towards 1' without overshooting, and hence that

$$|\rho_{t+1}| \leq |\rho_t| + (1 + |\rho_t|)M(r_t^2 + s_t^2).$$

Using $r_t \leq \lambda_\star^{t-T} r_T$ and $s_t \leq \lambda_\star^{t-T} s_T$, it follows that

$$|\rho_t| \leq |\rho_T| + 2M \sum_{0 \leq s \leq t-T} \lambda_\star^{2s}(r_T^2 + s_T^2),$$

provided $|\rho_s| < 1$ for $T \leq s < t$. Since $r_T^2 \sim s_T^2 \sim (4/T)^2 = \Theta(\varepsilon^{4/3})$, while $\sum_{s \geq 0} \lambda_\star^{2s} = O(1/\varepsilon)$, it follows easily that $|\rho_t|$ does remain bounded by 1, and in fact that $|\rho_t| = O(\varepsilon^{1/3})$ uniformly in $t \geq T$. In particular, $s_t \sim r_t$ for $t \geq T$, proving the second statement of the lemma. The next two statements follow.

Finally, we turn to the estimate on $\prod_{t \geq 1}(1 - s_t)$. From (4.28) we see that $\sum_t |\tilde{s}_t - 2/t| = \sum_t O(t^{-2})$ is bounded. It follows that $\sum_{t \geq 3} \log\left(\frac{1 - \tilde{s}_t}{1 - 2/t}\right)$ converges; let us write c for the value of this sum, which does not involve ε. Since $T \to \infty$, the sum truncated at T converges to c as $\varepsilon \to 0$. Hence, from (4.30),

$$\begin{aligned} \prod_{t<T}(1 - s_t) = \prod_{t<T}(1 - \tilde{s}_t + O(\varepsilon)) = e^{O(\varepsilon T)} \prod_{t<T}(1 - \tilde{s}_t) \\ \sim (1 - \tilde{s}_1)(1 - \tilde{s}_2)e^c \prod_{3 \leq t < T}(1 - 2/t) \sim \gamma_0 T^{-2}, \end{aligned}$$

for some constant $\gamma_0 > 0$. On the other hand, comparison with an integral shows that

$$\sum_{t\geq T} r_t = -2\log(\varepsilon T) + O(\varepsilon T) = -2\log(\varepsilon T) + o(1).$$

We have already seen that $\sum_{t\geq T} r_t^2 = o(1)$, and the same for s_t, so, using the bound on $|\rho_t|$ established above, it follows that

$$\begin{aligned}\log \prod_{t\geq T}(1-s_t) = o(1) - \sum_{t\geq T} s_t = o(1) - (1+O(\varepsilon^{1/3}))\sum_{t\geq T} r_t \\ = 2\log(\varepsilon T) + o(1) + O(\varepsilon^{1/3}\log(\varepsilon T)) = 2\log(\varepsilon T) + o(1).\end{aligned}$$

Thus $\prod_{t\geq 1}(1-s_t) \sim \gamma_0 T^{-2}(\varepsilon T)^2 = \gamma_0\varepsilon^2$, as claimed. □

The final statement of Lemma 4.21 shows that if we start one copy of $\mathfrak{X}_{\lambda_\star}$ at each time $t \geq 1$, the probability that for every t the tth copy dies within t generations is asymptotically $\gamma_0\varepsilon^2$.

Remark. Constructing $\mathfrak{X}_\lambda$ first by constructing $\mathfrak{X}_\lambda^+$, and then adding the subcritical trees, we see that $p_r = \prod_{t=1}^r (1-s_t)$ is exactly the probability that $|X_r| = 1$ given that $|X_r^+| = 1$. We have

$$\begin{aligned}\mathbb{P}\big(|X_r| = 1 \bigm| |X_r^+| = 1\big) = \mathbb{P}\big(|X_r^+| = 1 \bigm| |X_r| = 1\big)\mathbb{P}(|X_r| = 1)/\mathbb{P}(|X_r^+| = 1) \\ = s\mathbb{P}(|X_r| = 1)/\mathbb{P}(|X_r^+| = 1) = s\mathbb{P}(|X_r| = 1)/(s\lambda_\star^r) = \mathbb{P}(|X_r| = 1)/\lambda_\star^r.\end{aligned}$$

So the final statement of Lemma 4.21 is equivalent to the statement that for large r, $\mathbb{P}(|X_r| = 1) \sim \gamma_0\varepsilon^2\lambda_\star^r$ for some constant γ_0, which can presumably be seen more directly somehow.

Before turning to our next real lemma, let us get a simple observation out of the way. Trivially, $\mathbb{E}(|X_t^-|) = \lambda_\star^t$; a simple inductive calculation gives the standard formula $\mathbb{E}(|X_t^-|^2) = \lambda_\star^t(1+\lambda_\star+\cdots+\lambda_\star^t)$. Since $\lambda_\star > 1-\varepsilon$ (see (4.2)), this gives $\mathbb{E}(|X_t^-|^2) \leq \varepsilon^{-1}\lambda_\star^{2t}$, so $\mathrm{Var}(|X_t^-|) \leq \varepsilon^{-1}(\mathbb{E}\,|X_t^-|)^2$. If we start $\mathfrak{X}_{\lambda_\star}$ with $N \geq 10/\varepsilon$ particles in generation 0, and $r \leq 1/\varepsilon$, then the size of generation r has expectation $\mu \geq N(1-\varepsilon)^{1/\varepsilon} \geq N/3$, and, using independence of the offspring of different particles, variance at most $\varepsilon^{-1}\mu^2/N \leq \mu^2/10$. It follows by Chebyshev's inequality that

$$\mathbb{P}\big(|X_r^-| \geq N/6 \bigm| |X_0^-| = N\big) \geq 1/2 \tag{4.34}$$

whenever $N \geq 10/\varepsilon$ and $r \leq 1/\varepsilon$.

Let $(D_t)_{t\geq 0}$ denote the union of countably many independent copies of $\mathfrak{X}_{\lambda_\star}$, where the ith process starts with a single particle in generation i. Thus $|D_0| = 1$, while given $|D_t|$, the distribution of $|D_{t+1}|$ has the form $1 + \mathrm{Po}(\lambda_\star|D_t|)$.

Lemma 4.22. *Let $0 < \varepsilon < 1/10$ be given, and define $\lambda = 1+\varepsilon$ and $\lambda_\star = \lambda(1-s(\lambda))$ as usual. For $\omega \geq 20$ and $t \geq 0$ we have $\mathbb{P}(|D_t| \geq \omega/\varepsilon) = e^{-\Omega(\omega)}$, where the implied constant is absolute. Furthermore, for $T \geq 1/\varepsilon$,*

$$\mathbb{P}\big(\exists t : 0 \leq t \leq T,\ |D_t| \geq \omega/\varepsilon\big) = O(\varepsilon T e^{-\Omega(\omega)}).$$

Proof. Let $f_t(x) = \mathbb{E}\, x^{|D_t|}$ be the probability generating function of $|D_t|$. Then $f_0(x) = x$, while from the relationship between D_{t+1} and D_t above we have

$$f_{t+1}(x) = x f_t\big(e^{\lambda_\star(x-1)}\big)$$

for all $t \ge 0$ and all x. Fix $t \ge 0$ and let $x_0 = 1 + \varepsilon/10$, say. Inductively defining x_r by $x_{r+1} = e^{\lambda_\star(x_r-1)} > 1$, note that

$$f_t(x_0) = \prod_{r=0}^{t} x_r \le \prod_{r=0}^{\infty} x_r. \tag{4.35}$$

We claim that for every r we have

$$x_r \le 1 + (1-\varepsilon/3)^r \varepsilon/10, \tag{4.36}$$

say. This certainly holds for $r = 0$. Suppose then that (4.36) holds for some particular r. Since $\lambda_\star < (1-\varepsilon/2)$, it follows that $x_{r+1} \le \exp((1-\varepsilon/2)(1-\varepsilon/3)^r\varepsilon/10)$. Using $\exp(y) \le 1+y+y^2$ for $y \le 1$, we thus have

$$x_{r+1} \le 1 + (1-\varepsilon/2)(1-\varepsilon/3)^r\varepsilon/10 + (1-\varepsilon/3)^r\varepsilon^2/100 \le 1 + (1-\varepsilon/3)^{r+1}\varepsilon/10,$$

and (4.36) follows by induction.

Combining (4.35) and (4.36) we have, crudely, $\log(f_t(x_0)) \le 2\sum_r (1-\varepsilon/3)^r\varepsilon/10 = 6/10$, so $f_t(x_0) \le 2$. Recalling that $x_0 = 1+\varepsilon/10$, we thus have $\mathbb{P}(|D_t| \ge \omega/\varepsilon) \le f_t(x_0)/x_0^{\omega/\varepsilon} \le 2e^{-\Omega(\omega)}$, and the first statement of the lemma follows, for all $\omega \ge 2$, say.

For the second statement, suppose now that $\omega \ge 20$. Using (4.34), and simply ignoring the one new particle added in each generation, for $0 \le r \le 1/\varepsilon$, conditional on $|D_t| = N \ge 10/\varepsilon$, the probability that $D_{t+r} \ge N/6$ is at least $1/2$. Let $k = \lfloor 1/\varepsilon \rfloor$. Examining D_t, $D_{t+1}, \ldots$, one by one, stopping the first time any of these sets has size more than ω/ε, it follows that

$$\mathbb{P}\big(|D_{t+k}| \ge \omega/(6\varepsilon) \;\big|\; \exists t' : t \le t' \le t+k,\, |D_{t'}| \ge \omega/\varepsilon\big) \ge 1/2,$$

so

$$\mathbb{P}\big(\exists t' : t \le t' \le t+k,\, |D_{t'}| \ge \omega/\varepsilon\big) \le 2\mathbb{P}\big(|D_{t+k}| \ge \omega/(6\varepsilon)\big) = e^{-\Omega(\omega)},$$

using the first part for the final bound. Summing over $0 \le t \le T$ in steps of $k = \lfloor 1/\varepsilon \rfloor$, the second statement follows. □

Next we shall show that conditioning $\mathfrak{X}_{\lambda_\star}$ to survive to (at least) a certain time does not increase its expected total size too much.

Lemma 4.23. *Suppose that $\varepsilon > 0$ and $t \ge 1$. Let N denote the total number of particles in $\mathfrak{X}_{\lambda_\star}$. Then $\mathbb{E}(N \mid X_t^- \ne \emptyset) \le (t+1)/\varepsilon$.*

Proof. We shall use repeatedly the observation that for any μ, the distribution of a Poisson $\mathrm{Po}(\mu)$ random variable conditioned to be at least 1 is stochastically dominated by $1 + \mathrm{Po}(\mu)$. (This may be seen by considering the first point, if any, of a Poisson process in an interval.)

We may view the first generation of $\mathfrak{X}_{\lambda_\star}$ as the union of two sets: the set S_1 consisting of those children of the root that survive to time t, and the set S_2 of those that do not. The full process is then obtained by taking a copy of the process conditioned to survive for $t-1$ generations for each particle in S_1, and a copy conditioned to die within $t-1$ generations for each in S_2. The sets S_1 and S_2 have independent Poisson sizes. Conditioning on X_t^- being non-empty is equivalent to conditioning on $|S_1| \geq 1$. Let us instead simply add a new particle to S_1. By the observation at the start of the proof, this gives a process whose distribution dominates that of $\mathfrak{X}_{\lambda_\star}$. Our new process consists exactly of the standard process $\mathfrak{X}_{\lambda_\star}$, together with a copy of $\mathfrak{X}_{\lambda_\star}$ conditioned to survive at least $t-1$ generations started at time 1.

Applying the same procedure to the new copy (i.e., to the children of the extra particle in S_1, but *not* to those of the other particles in S_1), and continuing, it follows that the distribution of $\mathfrak{X}_{\lambda_\star}$ conditioned to survive to time t is dominated by the distribution of the union of $t+1$ copies of $\mathfrak{X}_{\lambda_\star}$, one started at each time r, $0 \leq r \leq t$. This has expected total size $(t+1)/(1-\lambda_\star) \leq (t+1)/\varepsilon$. □

Finally, we observe that if we condition on $\mathfrak{X}_\lambda$ surviving, this process quickly realizes its *conditional* expected size, which by Lemma 4.6 is a factor $(1+o(1))/s \sim 1/(2\varepsilon)$ larger than the unconditioned size.

Lemma 4.24. *Let $\lambda = 1+\varepsilon$, where $\varepsilon = \varepsilon(n) \to 0$. Let $\omega(n)$ and $\omega'(n)$ satisfy $\omega' \to \infty$ and $\omega/\omega' \to \infty$, and set $t_1 = \lfloor \log\omega/\log\lambda \rfloor$. Then $\mathbb{P}\big(|X_{t_1}| \geq \omega'/\varepsilon \mid (X_t) \text{ survives}\big) \to 1$ as $n \to \infty$.*

Proof. This is a simple consequence of Lemma 4.13 together with (a weak form of) our tail bound on the (a.s. defined) limit $Y = \lim_{t\to\infty} |X_t|/\lambda^t$. Starting with the tail bound, set $x = 2\omega'/\omega = o(1)$. From Corollary 4.17 we have

$$\mathbb{P}(0 < Y \leq x/\varepsilon) \sim 4\varepsilon x^{\log(1/\lambda_\star)/\log\lambda} = o(\varepsilon),$$

since $\log(1/\lambda_\star)/\log\lambda \sim 1$. Recalling that $Y > 0$ if and only if the process survives, it follows that $\mathbb{P}\big(Y \leq x/\varepsilon \bigm| (X_t) \text{ survives}\big) = o(1)$.

Conditional on survival, there is some generation with size at least ω'/ε with probability 1. By Lemma 4.13, with probability $1-o(1)$ the first such generation occurs at time $\log(\omega'/\varepsilon)/\log\lambda - \log Y/\log\lambda + O(1/\varepsilon)$. By the tail bound on Y above, this is less than t_1 with probability $1+o(1)$. Moreover, with probability $1-o(1)$, from this point on $|X_t|$ is within a factor 2, say, of $\lambda^t Y$. At time t_1, $\lambda^{t_1} Y \geq 2\omega'/\varepsilon$ unless $Y \leq 2\varepsilon^{-1}\omega'/\omega = x/\varepsilon$, an event of probability $o(1)$. □

4.5 Typical distances in the 2-core

We are now almost ready to prove our lower bound on the diameter of $G(n,p)$. It turns out that we need a result concerning typical distances in the 2-core. Unfortunately, this does not seem to follow easily from any published results, and our proof is a little painful. We first need a result that essentially bounds the kth moment of the size of the giant component.

Let $G = G(n, \lambda/n)$. We say that a k-tuple $(x_1, \ldots, x_k)$ of not necessarily distinct vertices of G is *useful* if for each i, either x_i is in a component of G containing a cycle, or it is joined by a path in G to some other x_j. It turns out that almost all useful k-tuples arise as they should, i.e., from vertices in the giant component. Recall that if N is the number of vertices in the giant component of $G(n, \lambda/n)$, then $N = (2 + o_{\mathrm{p}}(1))\varepsilon n$; see (4.6). An immediate consequence is that $\mathbb{E} N \geq (2 - o(1))\varepsilon n$. (In fact, it is well known that $\mathbb{E} N \sim 2\varepsilon n$.)

Lemma 4.25. *Let $\lambda = 1 + \varepsilon$, where $\varepsilon = \varepsilon(n) > 0$ satisfies $\varepsilon \to 0$ and $\Lambda = \varepsilon^3 n \to \infty$, and let $k \geq 1$ be fixed. Then the expected number of useful k-tuples in $G(n, \lambda/n)$ is $(1 + o(1))(2\varepsilon n)^k$.*

Proof. The lower bound is immediate, since the number of useful k-tuples is at least the kth power of the number N of vertices in the largest component of G, and $\mathbb{E} N^k \geq (\mathbb{E} N)^k \geq \big((2 - o(1))\varepsilon n\big)^k$.

Let $\psi = \psi(n)$ tend to infinity very slowly.

We first get a simple observation out of the way. Let us say that a k-tuple of vertices is *close* if each x_j, $j > 1$, is within distance ψ/ε of x_1 in G. Let C_k denote the number of close k-tuples. Set $t = \lfloor \psi/\varepsilon \rfloor$. Then $\mathbb{E} C_k = n \mathbb{E}(|G_{\leq t}(x)|^{k-1})$, where $x = x_1$ is any fixed vertex of G. Now $|G_{\leq t}(x)|$ is stochastically dominated by $|X_{\leq t}|$, the union of the first t generations of $\mathfrak{X}_\lambda$. (We are simplifying slightly here: a binomial $\mathrm{Bi}(n, p)$ is dominated by a Poisson with mean $-n \log(1 - p) = np + O(np^2)$, so we should consider the branching process with a parameter slightly larger than λ; the difference is negligible.) It is easy to check (for example by calculating inductively) that with r fixed, $\mathbb{E} |X_{\leq t}|^r = O(t^{2r-1}\lambda^{rt}) = O(\psi^{2r-1}e^{r\psi}\varepsilon^{-(2r-1)}) = \widetilde{O}(\varepsilon^{-(2r-1)})$, where we write $f = \widetilde{O}(g)$ if f is bounded by a function of ψ times g. It follows that

$$\mathbb{E} C_k = \widetilde{O}(n\varepsilon^{-(2k-3)}) = o(\varepsilon^k n^k), \tag{4.37}$$

provided ψ grows slowly enough, since $\varepsilon^{3k-3}n^{k-1} \to \infty$.

Turning to useful k-tuples, we shall proceed by induction on k. Let U_k denote the number of k-tuples of *distinct* vertices that are useful. Since $\varepsilon n \to \infty$, it suffices to prove that $\mathbb{E} U_k \sim (2\varepsilon n)^k$. We may then bound the total number of useful k-tuples in terms of $U_1, \ldots, U_k$.

From now on we insist that $x_1, \ldots, x_k$ are distinct. Let us say that a useful k-tuple is *reducible* if it contains a non-empty subset S which forms a close r-tuple within a component of G containing none of the remaining x_i. If this holds, then there is some set of edges present witnessing that S is a close r-tuple, and a disjoint set witnessing the event that the remaining set S^{c} is useful. (We may have $k - r = 0$; a 0-tuple is always useful.) By the van den Berg–Kesten inequality [5], the probability of this event is at most the probability that S is close times the probability that S^{c} is useful. Using (4.37) and the induction hypothesis, this probability is $o(\varepsilon^r)O(\varepsilon^{k-r}) = o(\varepsilon^k)$. Summing over r and over the $\binom{k}{r}$ sets S, we see that the expected number of reducible useful k-tuples is $o(\varepsilon^k n^k)$.

Finally, we estimate the number of irreducible useful k-tuples. To do so, let us pick $x_1, \ldots, x_k$ one-by-one; we do *not* fix them in advance. Each x_i is chosen uniformly from the remaining $n - i + 1$ vertices.

Having chosen x_i, let us explore its neighbourhoods as follows. First, if x_i itself is in the set R of vertices previously reached by such explorations, we do not explore at all, and declare x_i to be 'atypical for reason 1'. Otherwise, we explore the neighbourhoods of x_i as usual, except that we do not (for the moment) test for edges to R. Also, we stop as soon as either (i) we reach generation ψ/ε, or (ii) we find ψ/ε vertices in one generation t. (We then stop partway through this generation.) Let Γ_i denote the set of vertices reached. Our next step is to test all edges from Γ_i to R; if such an edge is present, x_i is 'atypical for reason 2'. We then test for non-tree edges within Γ_i, i.e., for edges between two vertices in Γ_i at distance t from x_i, or for 'redundant' edges between vertices at distances t and $t+1$. If we find such a non-tree edge, then x_i is 'atypical for reason 3'. Finally, if we have not yet labelled x_i as atypical, then we label x_i as 'good' if condition (i) or (ii) held, and 'bad' otherwise, i.e., if we ran out of vertices to explore.

Note that if any x_i is bad, then Γ_i is its entire component, this component is a tree, and every vertex of this tree is within distance ψ/ε of x_i. If $(x_1, \ldots, x_k)$ is useful and some x_i is bad, then it follows that at least one later x_j lies in Γ_i, so $(x_1, \ldots, x_k)$ is reducible. Thus we may bound the expected number of irreducible useful k-tuples by n^k times the probability that no x_i is bad. We do this by showing that the conditional probability that x_i is atypical or good given $x_1, \ldots, x_{i-1}$ and the associated explorations is at most $(1+o(1))2\varepsilon$.

The definition of the exploration ensures that each Γ_i contains at most ψ^2/ε^2 vertices, so $|R| \le k\psi^2\varepsilon^{-2}$ and the probability that x_i is atypical for reason 1 is $\widetilde{O}(\varepsilon^{-2}n^{-1}) = o(\varepsilon)$. Suppose this does not happen. Then $|\Gamma_i|$ is stochastically dominated by $|X_{\le \psi/\varepsilon}|$, which has expectation $\sum_{r \le \psi/\varepsilon} \lambda^r = O(\varepsilon^{-1}\lambda^{\psi/\varepsilon}) = \widetilde{O}(\varepsilon^{-1})$. At the end of the previous exploration, we have already uncovered all edges incident with all vertices of each Γ_j, $j < i$, except (possibly) for vertices in the last two generations. (Two because we may have stopped part way through a generation.) There are at most $2\psi k/\varepsilon = \widetilde{O}(\varepsilon^{-1})$ such vertices in total. Hence, given Γ_i, the conditional probability that x_i is atypical for reason 2 is at most $|\Gamma_i|\widetilde{O}(\varepsilon^{-1}/n)$, so the unconditional probability is at most $\widetilde{O}(\mathbb{E}\,|\Gamma_i|\varepsilon^{-1}/n) = \widetilde{O}(\varepsilon^{-2}n^{-1}) = o(\varepsilon)$.

Similarly, given Γ_i, the conditional probability that x_i is atypical for reason 3 is at most $|\Gamma_i|(2\psi/\varepsilon)\lambda/n$, since for each vertex we have to test edges to the at most $2\psi/\varepsilon$ other vertices in the same generation or the previous generation. Hence the probability that x_i is atypical for this reason is also $o(\varepsilon)$.

Finally, the exploration leading to Γ_i is dominated by $\mathfrak{X}_\lambda$, so the probability that x_i is good is bounded by the probability that the branching process $\mathfrak{X}_\lambda$ either reaches size ψ/ε, or lasts for at least ψ/ε generations. It is easy to check that the probability of this event is $(1+o(1))s \sim 2\varepsilon$; indeed, from Lemma 4.13 (say), the event that $\mathfrak{X}_\lambda$ reaches size ψ/ε coincides up to probability $o(\varepsilon)$ with the event that $\mathfrak{X}_\lambda$ survives, and Lemma 4.21 and the fact that $\mathfrak{X}_\lambda$ conditioned on dying is (X_t^-) show that the events that $\mathfrak{X}_\lambda$ survives for ψ/ε generations and that it survives forever agree up to probability $o(\varepsilon)$. □

Lemma 4.26. *Let $\lambda = 1+\varepsilon$, where $\varepsilon = \varepsilon(n) > 0$ satisfies $\varepsilon \to 0$ and $\Lambda = \varepsilon^3 n \to \infty$, and let C denote the 2-core of $G = G(n, \lambda/n)$. Then $N = |C|$ satisfies*

$$\mathbb{E}(N^k) \sim (2\varepsilon^2 n)^k \tag{4.38}$$

for each fixed k. Furthermore, if $d = \log\Lambda/\log\lambda - \omega/\varepsilon$ with $\omega = \omega(n) \to \infty$, then

$$\mathbb{E}\, M_d^{(k)} = o(\varepsilon^{2k}n^k), \tag{4.39}$$

where $M_d^{(k)}$ is the number of k-tuples of vertices of C some pair of which are within distance d.

One might expect the first statement to be known. Indeed, Pittel and Wormald [36] have shown that the distribution of the size of the 2-core is asymptotically normal, with mean $(2+o(1))\varepsilon^2 n$ and variance $(12+o(1))\varepsilon n = o(\varepsilon^4 n^2)$. Unfortunately, convergence in distribution does not imply convergence of the relevant moments, so we cannot simply deduce (4.38). We shall prove (4.38) using Lemma 4.25; it is then easy to deduce (4.39).

Proof. Fix k distinct vertices $x_1, \ldots, x_k$, and let A be the event that $x_1, \ldots, x_k$ are all in the 2-core. It suffices to show that $\mathbb{P}(A) \leq (1 + o(1))(2\varepsilon^2)^k$.

Let $G' = G - \{x_1, \ldots, x_k\}$, so G' has the distribution of $G(n', \lambda'/n')$ where $n' = n - k$ and $\lambda' - 1 \sim \lambda - 1$. Let U_r denote the number of useful r-tuples of not necessarily distinct vertices of G'. By Lemma 4.25, we have

$$\mathbb{E}\, U_r \leq (1 + o(1))(2\varepsilon n)^r \tag{4.40}$$

for any fixed r.

Suppose that A holds, and let E be a minimal set of edges witnessing A. Note that every vertex of $S = \{x_1, \ldots, x_k\}$ meets at least two edges of E. Also, since a vertex is in the 2-core if and only if it is on a cycle or on a path joining two cycles, E may be written as the union of k graphs with maximum degree at most 3, so at most $3k^2$ edges of E meet S.

Let E_0, E_1 and E_2 denote respectively the sets of edges of E with both ends in S, one end in S, and neither end in S. List the edges of E_1 as $a_i b_i$, $1 \leq i \leq r \leq 3k^2$, where each a_i is in S and each b_i in G'. From the minimality of E, each b_i is either joined to some other b_j by a path in E_2 (which may have length 0 if $b_i = b_j$), or is joined by a path in E_2 to a cycle in E_2. (Otherwise, removing pendant edges from E, we obtain a smaller witness to A.) It follows that the r-tuple $(b_1, \ldots, b_r)$ is useful *in the graph* G'. Let $t = |E_0|$.

Suppose first that $t = 0$ and $|E_1| = 2k$. (More precisely, suppose there is a (minimal) witness E with these properties.) Since each x_i meets at least two edges of E_1, it meets exactly two. Hence there is a $2k$-tuple $(b_1, \ldots, b_{2k})$ that is useful in G', with x_i joined to b_{2i-1} and b_{2i}. But from (4.40) and the independence of G' and the edges between S and G', the expected number of such $2k$-tuples is at most $(1 + o(1))(2\varepsilon n)^{2k}(\lambda/n)^{2k} \sim 2^{2k}\varepsilon^{2k}$. Since the $2k$-tuple is ordered, whenever there is one there are at least 2^k (swapping b_1 and b_2, etc), so the probability that a witness E exists with $t = 0$ and $|E_1| = 2k$ is at most $(1 + o(1))(2\varepsilon^2)^k$.

It remains to show that the probability that there is a witness E with $t > 0$ or $t = 0$ and $|E_1| > 2k$ is $o(\varepsilon^{2k})$, for which we simply bound the expected number of such witnesses. Since each vertex of S meets at least two edges of E_1, we have $r = |E_1| \geq 2k - 2t$, while, as noted above, $r \leq 3k^2$. Hence, setting $\Delta = 0$ if $t > 0$ and $\Delta = 1$ if $t = 0$, the expectation is bounded by

$$\sum_{t=0}^{\binom{k}{2}} \binom{k}{2}^t (\lambda/n)^t \sum_{r=2k-2t+\Delta}^{3k^2} k^r (\mathbb{E}\, U_r)(\lambda/n)^r,$$

since there are most $\binom{k}{2}$ choices for each of the t edges inside S, and, given r, at most k^r possibilities for which of the x_j each a_i is. (Some b_i may coincide, but we do not care.) By (4.40), each term in the sum may be bounded by a constant times

$$n^{-t}(2\varepsilon n)^r n^{-r} = O(n^{-t}\varepsilon^r) = O(n^{-t}\varepsilon^{2k-2t+\Delta}).$$

For $t = 0$ this is $O(\varepsilon^{2k+1}) = o(\varepsilon^{2k})$. For $t \geq 1$, since $\varepsilon^2 n \to \infty$, the final bound is $o(\varepsilon^{2k})$. It follows that $\mathbb{P}(A) \sim (2\varepsilon^2)^k$, completing the proof of (4.38).

Finally, as noted above, it is relatively easy to deduce (4.39) from (4.38). Let M be the number of k-tuples of vertices of C in which every pair is at distance larger than d. Then it suffices to show that $\mathbb{E} M \geq (1 + o(1))(2\varepsilon^2 n)^k$. In proving such a lower bound, we may consider k-tuples with additional properties that make the analysis easier.

Let $\psi = \psi(n) = o(\omega)$ tend to infinity very slowly, let E be the branching process event that at least two particles in generation 1 survive to generation $t = \psi/\varepsilon$, that these particles each have at least ψ/ε descendants in X_t, and that $|X_{t'}| \leq \psi^{10}\varepsilon^{-1}\lambda^{\psi/\varepsilon} = e^{O(\psi)}\varepsilon^{-1}$ for $0 \leq t' \leq t$. Recalling that, conditioned on survival, the branching process typically has size of order $\varepsilon^{-1}\lambda^{t'}$ in generations t' where t' is significantly larger than $1/\varepsilon$ (see Lemmas 4.24 and 4.13), it is easy to check that $\mathbb{P}(E) \sim s^2/2 \sim 2\varepsilon^2$, the asymptotic probability that two particles in generation 1 survive. Also, Lemma 4.7 applies to all trees consistent with E.

Given distinct vertices $x_1, \ldots, x_k$ of G, let E'_k denote the event that for every i the t-neighbourhood of x_i has the property corresponding to E, and these t-neighbourhoods are disjoint. Also, let E_k be the event that E'_k holds, every x_i is in the 2-core, and $d(x_i, x_j) > d$ for all i and j. By Lemma 4.8 we have $\mathbb{P}(E'_k) \sim \mathbb{P}(E)^k \sim (2\varepsilon^2)^k$. Since $\mathbb{E} M \geq (1+o(1))n^k \mathbb{P}(E_k)$, it thus suffices to show that $\mathbb{P}(E_k \mid E'_k) = 1 - o(1)$.

But after testing whether E'_k holds, we have not looked at any edges outside the relevant neighbourhoods. The expected number of paths of length at most d joining one pair of vertices in the last generation of these neighbourhoods is bounded by

$$\sum_{1 \leq i \leq d} n^{i-1}(\lambda/n)^i = n^{-1} \sum_{i \leq d} \lambda^i \sim \varepsilon^{-1} n^{-1} \lambda^d.$$

There are at most $\binom{k}{2} e^{O(\psi)}\varepsilon^{-2}$ pairs to consider, so the probability of finding any such path is at most

$$e^{O(\psi)}\varepsilon^{-3} n^{-1}\lambda^d = e^{O(\psi)}\Lambda^{-1}\Lambda\lambda^{-\omega/\varepsilon} = e^{O(\psi)}e^{-(1+o(1))\omega} = o(1).$$

Also, since for each of $x_1, \ldots, x_k$ we have two neighbours with many (at least ψ/ε) descendants in generation t, given E'_k it is very likely that these neighbourhoods continue to expand and eventually meet, so whp each x_i is in C. Thus $\mathbb{P}(E_k \mid E'_k) = 1 - o(1)$, as required. □

In fact, one can easily bound the expected number of pairs of vertices of C at distance significantly larger than $\log \Lambda / \log \lambda$, noting that all but at most $o(\varepsilon^4 n^2)$ such pairs also have the property E'_2. Using Lemma 4.5 it is then easy to extend the argument above to show that if x and y are chosen uniformly at random from C, then

$$d(x, y) = \log \Lambda / \log \lambda + O_{\mathrm{p}}(1/\varepsilon).$$

Furthermore, one can obtain the limiting distribution of the correction term without too much difficulty. We omit the details as this is not our focus, and Lemma 4.26 is all we shall need to know about the 2-core.

With the simple preliminaries of the last few subsections behind us, we are now ready to begin the proof of Theorem 1.3.

4.6 The lower bound on the diameter

In this section we shall prove the lower bound on the diameter in Theorem 1.3. As noted in Section 1, we may assume that $\varepsilon \to 0$. The argument we present will be rather complicated. It is difficult to explain why this is the case, other than to say that we have tried many promising simple approaches, and while several are extremely plausible, we could not make the details rigorous. Of course, a much simpler proof may nevertheless exist.

We must show that with high probability vertices x and y at large distance exist. In doing so we may focus on vertices x and y whose neighbourhoods satisfy certain restrictions, although if we are too restrictive, we will not get a good bound. Before turning to the graph, let us describe the corresponding restrictions on the branching process. Overall, our aim is to consider the event that a certain 'wedge' condition holds, and $t_{\omega/\varepsilon} > t$, for t near t_0+t_1, but to make our arguments work we need some additional technical conditions. We start by insisting that the process (X_t^+) consisting of those particles with infinitely many descendants has size 1 for a large number of generations, then bifurcates, and the non-surviving descendants of all the particles up to this point have died out before very long, in a way to be made precise. This condition will include an analogue of the weak wedge condition described in Subsection 2.2.

For the rest of this section let $\varepsilon = \varepsilon(n) > 0$ satisfy $\varepsilon \to 0$ and $\Lambda = \varepsilon^3 n \to \infty$. Set

$$\omega = \Lambda^{1/6},$$

and let

$$t_1 = \lfloor \log \omega / \log \lambda \rfloor$$

and

$$t_0 = \lfloor \log(\varepsilon^3 n)/\log(1/\lambda_\star) \rfloor,$$

as before. (The rounding to integers will always be irrelevant in calculations.) Later, we shall also consider

$$t_2 = \log(\varepsilon^3 n/\omega^2)/\log \lambda.$$

For $r, q = O(1/\varepsilon)$, set $T_0 = t_0 + r$ and $T_1 = t_0 + t_1 + q$. Recalling from (4.4) that $\log \lambda, \log(1/\lambda_\star) \sim \varepsilon$, note that $T_0, T_1 = O(\varepsilon^{-1} \log \Lambda)$. We shall assume that $|r| \le t_0/2$ and that $|r|, |q| \le t_1/10$; these conditions hold for n sufficiently large.

Let $A = A_r$ be the event that $|X_{T_0}^+| = 1$ and $|X_{T_0+1}^+| = 2$. Then $\mathbb{P}(A) = s\mathbb{P}(Z_\lambda = 1)^{T_0}\mathbb{P}(Z_\lambda = 2)$, where, as before, Z_λ is a Poisson with mean $s\lambda$ conditioned to be at least 1. From (1.8) we have $\mathbb{P}(Z_\lambda = 1) = \lambda_\star$, while from the definition of Z_λ we have $\mathbb{P}(Z_\lambda = 2)/\mathbb{P}(Z_\lambda = 1) = (s\lambda)/2 \sim \varepsilon$. Hence,

$$\mathbb{P}(A) \sim 2\varepsilon \lambda_\star^{T_0} \varepsilon \lambda_\star \sim 2\varepsilon^2 \lambda_\star^{T_0} \sim 2\varepsilon^{-1} n^{-1} \lambda_\star^r = \Theta(\varepsilon^{-1} n^{-1}). \tag{4.41}$$

When A holds, let x_i denote the unique particle in X_i^+ for $0 \le i \le T_0$, and let y, y' be the two particles in $X_{T_0+1}^+$.

Let $B = B_r$ be the event that $A = A_r$ holds, and the following conditions are satisfied:

(i) (the *strong wedge condition*) x_0 has no children other than x_1 and, for $1 \le i < T_0$, no children of x_i other than x_{i+1} or y, y' have descendants in generation $2i$.

(ii) no particles in X_{T_0+1} other than y and y' have descendants in $X_{T_1'}$, where $T_1' = t_0 + \lfloor t_1/2 \rfloor$.

Note that $T_0 < T_1' < T_1$. Also, since $T_1' - T_0 = \lfloor t_1 \rfloor/2 - r = t_1/2 + O(\varepsilon^{-1})$, we have $\varepsilon(T_1' - T_0) \to \infty$. For the moment we could simply write T_1 in place of T_1' in condition (ii), but for the distribution result in Section 5 it is convenient that T_1' does not depend on q.

Unfortunately, it takes some effort to examine the effect that condition (i) has upon the distribution of $t_{\omega/\varepsilon}$, the time the branching process takes to reach size ω/ε. (Condition (ii) presents no problems.) Constructing $\mathfrak{X}_\lambda$ from $\mathfrak{X}_\lambda^+$ by adding independent copies of the subcritical process $\mathfrak{X}_{\lambda_\star}$ starting at each particle, condition (i) says that for $i < T_0$ the subcritical process started at x_i dies by time $\max\{i, 1\}$ (measured from its starting time), and condition (ii) that for $i \le T_0$ the process started from x_i dies by time $T_1' - i$. Writing $d_t = 1 - s_t = \mathbb{P}(|X_t^-| = 0)$ for the probability that $\mathfrak{X}_{\lambda_\star}$ dies by time t, we thus have

$$\mathbb{P}(B \mid A) = d_1 \prod_{i=1}^{T_0} d_{\min\{i, T_1'-i\}},$$

so

$$d_1 \prod_{i=1}^{\min\{T_0, T_1'/2\}} d_i \ge \mathbb{P}(B \mid A) \ge d_1 \prod_{i=1}^{\infty} d_i \prod_{i=T_1'-T_0}^{\infty} d_i. \tag{4.42}$$

By Lemma 4.21, as $\varepsilon i \to \infty$ we have $s_i \sim 2\varepsilon\lambda_\star^i$, and so $\log(1-s_i) \sim -2\varepsilon\lambda_\star^i$. Since $\varepsilon(T_1' - T_0) \to \infty$, it follows that

$$\sum_{i \ge T_1'-T_0} \log(1 - s_i) \sim -2\varepsilon \sum_{i \ge T_1'-T_0} \lambda_\star^i = O\big(\lambda_\star^{T_1'-T_0}\big) = o(1).$$

Hence, $\prod_{i \ge T_1'-T_0} d_i \sim 1$. Similarly, since $\varepsilon \min\{T_0, T_1'/2\} = \varepsilon T_1'/2 \to \infty$, we have $\prod_{i \ge \min\{T_0, T_1'/2\}} d_i \sim 1$. From (4.42) it then follows that

$$\mathbb{P}(B \mid A) \sim d_1 \prod_{i=1}^{T_1'/2} d_i \sim d_1 \prod_{i=1}^{\infty} d_i \sim d_1 \gamma_0 \varepsilon^2 = e^{-\lambda_\star} \gamma_0 \varepsilon^2 \sim \gamma_0 e^{-1} \varepsilon^2, \tag{4.43}$$

using Lemma 4.21 to estimate the infinite product.

Let C be the event that A holds, and the particles y and y' each have at least ω'/ε descendants in X_{T_1}, where $\omega' = \sqrt{\omega} = \Lambda^{1/12}$. (Later we shall need to know that vertices corresponding to y and y' have many 'descendants' at distance T_1 from x_0; this will ensure that x_{T_0} is in the 2-core.) By Lemma 4.24, applied with $T_1 - (T_0 + 1) = t_1 + q - r - 1 = t_1 + O(1/\varepsilon)$ in place of t_1, i.e., with $\lambda^{t_1+q-r-1} = \Theta(\omega)$ in place of ω, we have

$$\mathbb{P}(C \mid A) = 1 - o(1).$$

We would like to impose the condition that $|X_t| < \omega/\varepsilon$ for $0 \le t \le T_1$; however, for technical reasons we must consider the descendants of x_{T_0} separately from the remaining particles.

Let D_1 be the event that A holds, and between them the particles y and y' have fewer than $(\omega - 2\omega')/\varepsilon$ descendants in each set X_t, $T_0 + 1 \le t \le T_1$, noting that $\omega - 2\omega' \sim \omega$. Conditioning on A, the trees of descendants of the two particles y, y' form independent copies of $\mathfrak{X}_\lambda$, each conditioned on the event that it survives. By Lemma 4.13, whp as soon as the number of descendants of y in X_{T_0+1+r} is large compared to ε^{-1}, it then remains close to $\tilde{Y}\lambda^r$, where $\tilde{Y}$ has the distribution of $Y = Y_\lambda$ conditioned to be positive. Let $\tilde{Y}_2$ have the distribution of the sum of two independent copies of Y each conditioned to be positive. Then it follows that

$$\mathbb{P}(D_1 \mid A) = o(1) + \mathbb{P}\big(\tilde{Y}_2 \lambda^{T_1 - T_0 - 1} < (\omega - 2\omega')/\varepsilon\big).$$

Now $\lambda^{T_1-T_0-1} = \lambda^{t_1+q-r-1} = \omega\lambda^{q-r+O(1)} \sim \omega\lambda^{q-r}$, and $\omega - 2\omega' \sim \omega$, so

$$\mathbb{P}(D_1 \mid A) = o(1) + \mathbb{P}\big(\tilde{Y}_2 < (1+o(1))\lambda^{r-q}/\varepsilon\big) = o(1) + \mathbb{P}\big(s\tilde{Y}_2 < (2+o(1))e^{\varepsilon(r-q)}\big),$$

recalling that $s \sim 2\varepsilon$ and noting that, since $\varepsilon(r-q)$ is bounded and $\lambda = 1+\varepsilon$, we have $\lambda^{r-q} \sim \exp(\varepsilon(r-q))$. In a moment we shall sum over r; we can evaluate the sum of the corresponding terms above by relating it to a certain disjoint union of events and using Theorem 4.19. While this is aesthetically pleasing, we in fact know the asymptotic distribution of $\tilde{Y}_2$, so we shall just use it.

Recall from Lemma 4.9 and Corollary 4.15 that sY conditioned on $Y > 0$ has the distribution of $\tilde{Y}^+ = \tilde{Y}^+_\lambda$, which converges in distribution to an exponential with parameter 1 as $\varepsilon \to 0$. It follows that $s\tilde{Y}_2$ converges in distribution to the sum of two independent such exponentials, which has distribution function $\Psi(x) = \int_{y=0}^{x} e^{-y}(1-e^{-(x-y)})\,\mathrm{d}y = 1-(x+1)e^{-x}$. Thus

$$\mathbb{P}(D_1 \mid A) = \Psi(2e^{r'-q'}) + o(1),$$

where $r' = \varepsilon r$ and $q' = \varepsilon q$ and we use uniform continuity to remove the $(1+o(1))$ factor in the argument of Ψ.

Since $r' - q' = \Theta(1)$, we thus have $\mathbb{P}(D_1 \mid A) = \Theta(1)$, and hence the above equation can be written as $\mathbb{P}(D_1 \mid A) \sim \Psi(2e^{r'-q'})$. Since $\mathbb{P}(C \mid A) = 1 - o(1)$, it follows that $\mathbb{P}(C \cap D_1 \mid A) \sim \Psi(2e^{r'-q'})$. Given A, the events B and $C \cap D_1$ are independent, so

$$\mathbb{P}(C \cap D_1 \mid A \cap B) \sim \Psi(2e^{r'-q'}). \tag{4.44}$$

Turning to particles other than the descendants of y, y', first let D_1' be the event that A holds and, for $T_0 \le t \le T_1$, the set X_t contains at most ω'/ε particles that are descendants of x_{T_0} but not of y or y'. Given $A \cap B$, these particles form a copy of $\mathfrak{X}_\lambda$ starting at x_{T_0} and conditioned to die within $T_1' - T_0$ generations. This process may be viewed as $\mathfrak{X}_{\lambda_\star}$ conditioned to die by a certain time, so its distribution is dominated by that of $\mathfrak{X}_{\lambda_\star}$. Since the total expected size of $\mathfrak{X}_{\lambda_\star}$ is $O(\varepsilon^{-1})$, it follows that $\mathbb{P}((D_1')^{\mathrm{c}} \mid A \cap B) = o(1)$.

Let D_2 be the event that A holds and, for $0 \le t \le T_1$, the set X_t contains at most ω'/ε particles that are not descendants of x_{T_0}. Given $A \cap B$, the tree of particles that are not descendants of x_{T_0} has the distribution of one copy of $\mathfrak{X}_{\lambda_\star}$ started at each time t, $0 \le t < T_0$, conditioned on the various copies of $\mathfrak{X}_{\lambda_\star}$ dying by various times. This distribution is dominated by that studied in Lemma 4.22, so by Lemma 4.22 we have $\mathbb{P}(D_2^{\mathrm{c}} \mid A \cap B) = O(\varepsilon T_1 e^{-\Omega(\omega')}) = o(1)$, recalling that $T_1 = O(\varepsilon^{-1}\log\Lambda)$.

Let $D = D_1 \cap D_1' \cap D_2$. Since $\mathbb{P}((D_1')^c \cup D_2^c \mid A \cap B) = o(1)$, from (4.44), we have

$$\mathbb{P}(C \cap D_1 \cap D_1' \mid A \cap B) \sim \Psi(2e^{r'-q'}) \tag{4.45}$$

and

$$\mathbb{P}(C \cap D \mid A \cap B) \sim \Psi(2e^{r'-q'}).$$

Note for later that if D holds, then $|X_t| < \omega/\varepsilon$ for $t \le T_1$.

Finally, setting $E_{r,q} = A \cap B \cap C \cap D$, and recalling (4.41) and (4.43), we have

$$\mathbb{P}(E_{r,q}) \sim 2\varepsilon^{-1}n^{-1}\lambda_\star^r \gamma_0 e^{-1}\varepsilon^2 \Psi(2e^{r'-q'}) \sim 2\gamma_0 e^{-1}\varepsilon e^{-r'}\Psi(2e^{r'-q'})/n.$$

Since this estimate holds uniformly in r, q with $r, q = O(1/\varepsilon)$, it also holds uniformly in r, q with $|q|, |r| \le 2M/\varepsilon$, say, for some function $M = M(n)$ tending to infinity. For $|q| \le M/\varepsilon$, let $E_q = \bigcup_{-2M/\varepsilon \le r \le 2M/\varepsilon} E_{r,q}$. For fixed q, the events $E_{r,q}$ are disjoint, so we have

$$\begin{aligned}
\mathbb{P}(E_q) &\sim 2\gamma_0 e^{-1}\varepsilon n^{-1} \sum_{-2M/\varepsilon \le r \le 2M/\varepsilon} e^{-r'}\Psi(2e^{r'-q'}) \\
&= 2\gamma_0 e^{-1}\varepsilon n^{-1} e^{-q'} \sum_{-2M/\varepsilon - q \le r-q \le 2M/\varepsilon - q} e^{-(r'-q')}\Psi(2e^{r'-q'}).
\end{aligned}$$

The sum above simplifies considerably, since it corresponds to splitting a single event according to the time that (X_t^+) first subdivides. Rather than using this observation, we simply calculate. Since $\Psi(x) = O(1)$ as $x \to \infty$ and $\Psi(x) = O(x^2)$ as $x \to 0$, the sum above has exponentially decaying tails. Recalling that r' and q' simply denote εr and εq, it follows easily that

$$\mathbb{P}(E_q) \sim 2\gamma_0 e^{-1} n^{-1} e^{-\varepsilon q} \int_{-\infty}^{\infty} e^{-x}\Psi(2e^x)\,\mathrm{d}x.$$

A simple computation shows that the integral evaluates to 2, so

$$\mathbb{P}(E_q) \sim 4\gamma_0 e^{-1} n^{-1} e^{-\varepsilon q} \sim 4\gamma_0 e^{-1} n^{-1} \lambda_\star^q,$$

uniformly in $|q| \le M/\varepsilon$, provided $M = M(n)$ tends to infinity sufficiently slowly.

Note that the event E_q requires that $y, y' \in X_{T_0+1}^+$, an event depending on an infinite number of generations of the process $\mathfrak{X}_\lambda$. To work with the graph, we seek an event depending on a finite number of generations of $\mathfrak{X}_\lambda$. Let F_q be the event corresponding to E_q but depending only on the first $T_1 = t_0 + t_1 + q$ generations. More precisely, F_q is the event that there are exactly two particles, y and y', say, in some generation $T_0+1 = t_0+r+1$, $-2M/\varepsilon \le r \le 2M/\varepsilon$, with descendants in generation $T_1' = t_0 + \lfloor t_1/2 \rfloor$, each of these particles has at least ω'/ε descendants in X_{T_1}, y and y' have a common parent x_{T_0}, the equivalent of the strong wedge condition (i) holds, and $D = D_1 \cap D_1' \cap D_2$ holds. From the strong wedge condition, if F_q holds then, in the tree obtained from $\mathfrak{X}_\lambda$ by deleting all descendants of x_{T_0}, the initial particle is the unique particle at maximum distance from x_{T_0}.

If E_q holds, then so does F_q. Furthermore, $\mathbb{P}(E_q \mid F_q) = 1 + o(1)$, since for each of y and y', the probability that none of its at least ω'/ε descendants in generation T_1 goes on to survive forever is $O((1-s)^{\omega'/\varepsilon}) = o(1)$. Hence,

$$\mathbb{P}(F_q) \sim \mathbb{P}(E_q) \sim 4\gamma_0 e^{-1} n^{-1} \lambda_\star^q.$$

Let T be a tree of height $t = T_1$ consistent with F_q. Then $t = O(\varepsilon^{-1} \log \Lambda)$, while, since D holds, each generation contains at most $\omega\varepsilon^{-1} = \omega\Lambda^{-1/3}n^{1/3} = o(n^{1/3})$ vertices. Also, the total size $|T|$ of T is

$$O(\omega\varepsilon^{-2} \log \Lambda) = O(\omega\Lambda^{-2/3}n^{2/3} \log \Lambda) = o(n^{2/3}), \tag{4.46}$$

and $\varepsilon|T|^2 = O(\omega^2\varepsilon^{-3} \log^2 \Lambda) = O(\omega^2\Lambda^{-1}n \log^2 \Lambda) = o(n)$. Lemma 4.7 applies to all such trees, telling us that

$$\mathbb{P}\big(G_{\le t}(x) \cong T\big) \sim \mathbb{P}\big(G^0_{\le t}(x) \cong T\big) \sim \mathbb{P}\big(X_{\le t} \cong T\big).$$

Let $F_q(x)$ denote the event that $G_{\le T_1}(x)$ is a tree satisfying the property F_q, where $T_1 = t_0 + t_1 + q$. Summing over all such trees, we see that

$$\mathbb{P}(F_q(x)) \sim \mathbb{P}(F_q) \sim 4\gamma_0 e^{-1} n^{-1} \lambda_\star^q \tag{4.47}$$

uniformly in q such that $|\varepsilon q| \le M$, for some $M \to \infty$.

Let q_0 be chosen so that εq_0 tends to minus infinity very slowly, and let $F(x) = F_{q_0}(x)$. Let N be the number of vertices x for which $F(x)$ holds; then

$$\mathbb{E}\, N = n\mathbb{P}(F_{q_0}(x)) \sim 4\gamma_0 e^{-1} \lambda_\star^{q_0} \to \infty.$$

We are now almost finished: it remains to use a second moment argument to show that N is whp large, and then to bound the probability that two vertices satisfying the relevant condition are close.

Given distinct vertices x and y of $G = G(n, \lambda/n)$, let $A(x, y)$ be the event that $F(x)$ and $F(y)$ both hold, with the trees 'witnessing' this being disjoint. For trees T_1 and T_2 consistent with F_{q_0}, by Lemma 4.8 the probability that the relevant neighbourhoods of x and y are disjoint and isomorphic to T_1 and T_2 respectively is asymptotically the product of the individual probabilities. It follows easily that

$$\mathbb{P}(A(x, y)) \sim \mathbb{P}(F(x))\mathbb{P}(F(y)) = \mathbb{P}(F(x))^2. \tag{4.48}$$

At this point, it seems that there should be a simple argument involving 'pulling the trees off the 2-core and reattaching them randomly'. However, once again, we did not manage to make such an argument precise in a simple way.

Our next aim is to show that it is very unlikely that $F(x)$ and $F(y)$ hold and the trees witnessing these events overlap. Recall that if $F(x)$ holds, then there is a unique 'first' vertex in the neighbourhoods of x with two children with descendants in generation $t_0+t_1+q_0$. Let x' denote this vertex. Since the two children of x' each have at least $\omega'/\varepsilon = \Lambda^{1/12}/\varepsilon$ descendants in generation $t_0 + t_1 + q_0$, with probability at least $1 - o(\Lambda^{-100})$, say, their neighbourhoods continue to grow, and eventually meet, in which case x' is in the 2-core. Let $\tilde{F}(x)$ be the event that $F(x)$ holds and x' is in the 2-core, so $\mathbb{P}(\tilde{F}(x)) \sim \mathbb{P}(F(x))$. Also, let $\mathcal{B}_1$ be the 'global bad event' that there is some vertex x such that $F(x)$ holds but x' is not in the 2-core. Then

$$\mathbb{P}(\mathcal{B}_1) \le n\mathbb{P}(F(x))o(\Lambda^{-100}) = o(\lambda_\star^{q_0}\Lambda^{-100}) = o(1), \tag{4.49}$$

assuming, as we may, that $\varepsilon q_0 \ge -\log\log\Lambda$, say.

Similarly, if $F(x)$ and $F(y)$ hold, then it is very likely that x and y are in the same component. Writing $\mathcal{B}_2$ for the event that there are x and y in different components such that $F(x)$ and $F(y)$ hold, we have

$$\mathbb{P}(\mathcal{B}_2) = o(1). \tag{4.50}$$

For our second moment bound, we will study $\tilde{N}$, the number of vertices x such that $\tilde{F}(x)$ holds. Note that whp $\tilde{N}$ is equal to N, since $\mathcal{B}_1$ has probability $o(1)$. Also,

$$\mathbb{E}\,\tilde{N} = n\mathbb{P}(\tilde{F}(x)) \sim n\mathbb{P}(F(x)) = \mathbb{E}\,N \sim 4\gamma_0 e^{-1}\lambda_\star^{q_0} \to \infty.$$

Let $\tilde{A}(x,y)$ denote the event that $\tilde{F}(x)$ and $\tilde{F}(y)$ hold, with the trees witnessing $F(x)$ and $F(y)$ disjoint. If $\tilde{A}(x,y)$ holds, then so does $A(x,y)$. On the other hand, continuing to explore as before, we see that given $A(x,y)$, the vertices x' and y' are very likely to be in the 2-core, so

$$\mathbb{P}(\tilde{A}(x,y)) \sim \mathbb{P}(A(x,y)) \sim \mathbb{P}(F(x))^2 \sim \mathbb{P}(\tilde{F}(x))^2. \tag{4.51}$$

It remains to consider the case of overlapping trees.

We defined $F(x)$ in such a way that if $F(x)$ holds, then x' together with the component of $G-x'$ containing x forms a tree, in which x is the unique vertex at maximal distance from x'. If $\tilde{F}(x)$ holds, so x' is in the 2-core, then x is the unique vertex of this tree at maximal distance from the 2-core. Let T_x denote this tree, or, in general, the tree component containing x if we delete from G all edges lying in the 2-core. If $\tilde{F}(x)$ and $\tilde{F}(y)$ both hold, then from this uniqueness property, the trees T_x and T_y are disjoint, except possibly at x' and y': they are two distinct trees attached to the 2-core.

Let $\tilde{B}(x,y)$ be the event that $\tilde{F}(x) \cap \tilde{F}(y)$ holds and the trees T_x and T_y are disjoint (except possibly at x' and y'), but the trees witnessing $F(x)$ and $F(y)$ overlap. From the remarks above, for $x \neq y$,

$$\tilde{F}(x) \cap \tilde{F}(y) = \tilde{A}(x,y) \cup \tilde{B}(x,y). \tag{4.52}$$

To bound $\mathbb{P}(\tilde{B}(x,y))$, we first test whether $F(x)$ (*not* $\tilde{F}(x)$) holds, in a way that first uncovers the tree T_x. Roughly speaking, we would like to show that the number of trees T_x hanging off the 2-core is well behaved (i.e., its second moment is not too large). Then we could say that the attachment points to the 2-core are uniformly distributed, so it's unlikely that there are two trees attached to close points. The problem is that we need independence to get the second moment bound, and we do not have this, as we can't tell in advance when we have reached the 2-core and should stop exploring the tree from x. To get around this, we choose a stopping vertex in advance.

Given distinct vertices x and $\bar{x}$, let $F(x;\bar{x})$ be the event that $F(x)$ holds, with the division vertex x' equal to $\bar{x}$. Note that $F(x)$ is the disjoint union of the events $F(x;\bar{x})$, $\bar{x} \in V(G)\setminus\{x\}$, all of which are equally likely. Thus

$$\mathbb{P}(F(x;\bar{x})) = (n-1)^{-1}\mathbb{P}(F(x)) \sim n^{-1}\mathbb{P}(F(x)). \tag{4.53}$$

Let $T(x;\bar{x})$ be the event that $\bar{x}$ together with the component of $G-\bar{x}$ containing x forms a tree consistent with $F(x;\bar{x})$. In other words, $T(x;\bar{x})$ is the event that the part of G that we can reach from x if we do not allow ourselves to pass through $\bar{x}$ is one of a certain set of trees. Note that we do not insist that $\bar{x}$ is in fact in the 2-core, and that if $F(x;\bar{x})$ holds then $T(x;\bar{x})$ must hold.

Crucially, we may test whether $T(x;\bar{x})$ holds by exploring the neighbourhoods of x in the usual way, except that if we reach $\bar{x}$ at some point, we do not test for edges from $\bar{x}$ to unseen vertices. (Since we require the relevant neighbourhood to be a tree, we do test for edges between all pairs of reached vertices.) Also, given $T(x;\bar{x})$, we may test whether $F(x;\bar{x})$ holds by continuing to explore from $\bar{x}$; roughly speaking, the property required of this further exploration is captured by $C\cap D_1\cap D_1'$ above (this was the reason for 'splitting off' D_1' from D_2), and has probability essentially $\Theta(\varepsilon^2)$.

More precisely, suppose that $T(x;\bar{x})$ holds and let us condition on the particular tree T_x revealed by the exploration so far. Let $V' = V(G)\setminus V(T_x)\cup\{\bar{x}\}$. Then we have not yet examined any edges inside V', and the only edges outside V' are those of T_x. Since T_x is required to be consistent with $F(x;\bar{x})$, we know that $d(x,x_0)=t_0+r$ for some r with $|r|\le 2M/\varepsilon$, and, from (4.46), that T_x contains $o(n^{2/3})$ vertices.

Now (recalling that $F(x)=F_{q_0}(x)$), the event $F(x;\bar{x})$ holds if and only if the following conditions are satisfied as we explore a further $t=T_1-(t_0+r)=t_1+q_0-r$ steps from $\bar{x}$ in $G[V']$: (i) the graph we uncover is a tree, (ii) there are exactly two vertices (y and y') in $\Gamma_1(\bar{x})$ with 'descendants' in $\Gamma_{T_1'-(t_0+r)}(\bar{x})$, where $T_1'=t_0+\lfloor t_1/2\rfloor$, (iii) these two vertices each have at least ω'/ε descendants in $\Gamma_t(\bar{x})$, (iv) between them, y and y' have at most $(\omega-2\omega')/\varepsilon$ descendants in each $\Gamma_{t'}(\bar{x})$, $t'\le t$, and (v) the neighbours of $\bar{x}$ other than y and y' have in total at most ω'/ε neighbours in each of these sets. Indeed, (i) and (ii) together with the fact that T_x is consistent with $F(x;\bar{x})$ ensure that the event corresponding to $A\cap B\cap D_2$ in the definition of $F=F_{q_0}$ holds, (iii) ensures that C holds, (iv) that D_1 holds, and (v) D_1'.

Arguing as for (4.47), we can approximate the probability of these conditions holding by that of the corresponding branching process event (the conditions ensure that only $o(n^{2/3})$ vertices are involved in total). Then we may consider the infinite version of the branching process event, differing only in that we assume that y and y' are in X_1^+. Now we require that $|X_1^+|=2$; since $|X_1^+|\sim \mathrm{Po}(s\lambda)$, this has probability $\Theta(\varepsilon^2)$. Given this, in the branching process the remaining conditions corresponding to (iii), (iv) and (v) are *exactly* the conditions C, D_1 and D_1' considered earlier, except that now $\bar{x}$ plays the role of the initial particle x_0, and all generation numbers are offset by t_0+r. In particular, the conditional probability of these events is exactly the probability $\mathbb{P}(C\cap D_1\cap D_1' \mid A\cap B)$ evaluated in (4.45), with $r'=\varepsilon r$ and $q'=\varepsilon q_0$.

Let $\psi=\psi(n)$ be a function tending infinity to arbitrarily slowly (more slowly than the reciprocal of the implicit function in the $o(\cdot)$ notation in (4.39)), and let us write $f=\widetilde{\Theta}(g)$ if $f/g=\psi^{O(1)}$. Taking $M(n)$ to tend to infinity sufficiently slowly, from the comments above and (4.45), we see that

$$\mathbb{P}\big(F(x;\bar{x}) \mid T(x;\bar{x})\big)=\widetilde{\Theta}(\varepsilon^2)$$

whenever $T(x;\bar{x})$ holds. From (4.53) it follows that for all $\bar{x}\neq x$ we have

$$\mathbb{P}(T(x;\bar{x}))=\widetilde{\Theta}\big(\varepsilon^{-2}n^{-1}\mathbb{P}(F(x))\big)=\widetilde{\Theta}(\varepsilon^{-2}n^{-2}),$$

recalling (from (4.47)) that $\mathbb{P}(F(x))=\Theta(n^{-1}\lambda_\star^{q_0})=\widetilde{\Theta}(n^{-1})$.

Given $x \neq y$ and $\bar{x}$, $\bar{y}$, let $B'(x, y, \bar{x}, \bar{y})$ be the event that $T(x, \bar{x}) \cap T(y, \bar{y})$ holds, with the trees T_x and T_y edge disjoint. Note that if this event holds, then $\bar{x}, \bar{y} \notin \{x, y\}$. We may test whether $B'(x, y, \bar{x}, \bar{y})$ holds by exploring from x and y respectively (with the explorations modified at $\bar{x}$ and $\bar{y}$), and the two explorations cannot 'help' each other. Arguing as for (4.48) above, using Lemma 4.8, it follows that

$$\mathbb{P}(B'(x, y, \bar{x}, \bar{y})) \sim \mathbb{P}(T(x, \bar{x}))\mathbb{P}(T(y, \bar{y})) = \widetilde{\Theta}(\varepsilon^{-4} n^{-4})$$

for all $\bar{x}$, $\bar{y} \notin \{x, y\}$; the probability is 0 if $\bar{x}$ or $\bar{y} \in \{x, y\}$.

Fix vertices $x \neq y$, and let $\mathbf{x}$ and $\mathbf{y}$ be chosen independently and uniformly at random from $V(G)$. Note that

$$\mathbb{P}(B'(x, y, \mathbf{x}, \mathbf{y})) = \widetilde{\Theta}(\varepsilon^{-4} n^{-4}). \tag{4.54}$$

Let us condition on $B'(x, y, \mathbf{x}, \mathbf{y})$. Moreover, we condition on $V_x = V(T_x) \setminus \{\mathbf{x}\}$, on $V_y = V(T_y) \setminus \{\mathbf{y}\}$ and on the structure of the trees T_x and T_y, but *not* on $\mathbf{x}$ and $\mathbf{y}$. Given this information, $\mathbf{x}$ and $\mathbf{y}$ are independent and uniform from $V' = V(G) \setminus (V_x \cup V_y)$. Indeed, the given information says that certain trees T_x and T_y are attached to $\mathbf{x}, \mathbf{y} \in V'$. Each tree is equally likely to be attached to any vertex of V', so, given this, the attachment vertices are uniform on V'.

The event we have conditioned on does not depend on the edges in V'. Hence, the conditional distribution of $G[V']$ is that of $G' = G(n', \lambda/n)$, where $n' = n - |T_x| + 1 - |T_y| + 1$. From the definition of F_{q_0}, we have $|T_x|, |T_y| = o(n^{2/3})$, so $n' = n - o(n^{2/3})$. The edge probabilities in G' are thus λ'/n' where

$$\lambda' = \lambda n'/n = (1 + \varepsilon)(n - o(n^{2/3}))/n = 1 + \varepsilon - o(n^{-1/3}) = 1 + \varepsilon',$$

with $\varepsilon' \sim \varepsilon$.

Let $\tilde{B}'(x, y)$ be the event that $\tilde{B}(x, y)$ holds, and $\mathbf{x} = x'$, $\mathbf{y} = y'$, so

$$\mathbb{P}(\tilde{B}'(x, y)) = n^{-2}\mathbb{P}(\tilde{B}(x, y)). \tag{4.55}$$

If $\tilde{B}'(x, y)$ holds, then so does $B'(x, y, \mathbf{x}, \mathbf{y})$. Furthermore, $\mathbf{x}$ and $\mathbf{y}$ must be in the 2-core of G, which is the same as the 2-core U of G'. Also, $\mathbf{x}$ and $\mathbf{y}$ must be *close*, i.e., within distance $d = 2t_1 + 4M(n)/\varepsilon \sim 2t_1$.

From the remarks above, we may bound $\mathbb{P}\big(\tilde{B}'(x, y) \mid B'(x, y, \mathbf{x}, \mathbf{y})\big)$ by the conditional probability (given the trees T_x, T_y etc but not $\mathbf{x}$, $\mathbf{y}$) that $\mathbf{x}$ and $\mathbf{y}$ are close in U, and hence by

$$|G'|^{-2}\, \mathbb{E}\, M_d(G') \sim n^{-2}\, \mathbb{E}\, M_d(G'),$$

where $M_d(G')$ is the number of close pairs in U, and the expectation is over the random graph G'.

Now G' has the distribution of $G(n', \lambda'/n')$, with $\lambda' = 1 + \varepsilon'$ and $\varepsilon' \sim \varepsilon$. Also, $d \sim 2t_1 \sim 2\log\omega/\varepsilon = \log(\varepsilon^3 n)/(3\varepsilon) \sim 3^{-1}\log((\varepsilon')^3 n')/\varepsilon'$. By Lemma 4.26, we thus have $\mathbb{E}\, M_d(G') = o(\varepsilon^4 n^2)$. Taking our slowly growing function $\psi(n)$ small enough, the expectation is smaller than $\varepsilon^4 n^2$ by at least a factor e^{ψ}, say. It follows that

$$\mathbb{P}\big(\tilde{B}'(x, y) \mid B'(x, y, \mathbf{x}, \mathbf{y})\big) \leq \Theta(\varepsilon^4 e^{-\psi}).$$

Using (4.54) it follows that $\mathbb{P}(\tilde{B}'(x,y)) = \widetilde{O}(n^{-4}e^{-\psi}) = o(n^{-4})$, and hence, from (4.55), we have

$$\mathbb{P}(\tilde{B}(x,y)) = o(n^{-2}). \tag{4.56}$$

It follows that whp there are no pairs (x,y) for which $\tilde{B}(x,y)$ holds. Recalling (4.49), (4.50) and (4.52), and noting that $\tilde{A}(x,y)$ trivially implies $A(x,y)$, we see that whp every pair of vertices $x \neq y$ for which $F(x) \cap F(y)$ holds has the properties

$$d(x,y) < \infty \quad \text{and} \quad A(x,y). \tag{4.57}$$

Using (4.51), (4.52) and (4.56), and recalling that $\tilde{N}$ denotes the number of vertices x such that $\tilde{F}(x)$ holds, we have

$$\mathbb{E}(\tilde{N}(\tilde{N}-1)) = \sum_x \sum_{y \neq x} \big(\mathbb{P}(\tilde{A}(x,y)) + \mathbb{P}(\tilde{B}(x,y))\big) = (1+o(1))(\mathbb{E}\,\tilde{N})^2 + o(1).$$

Since $\mathbb{E}\,\tilde{N} \to \infty$, it follows that $\mathbb{E}\,\tilde{N}^2 \sim (\mathbb{E}\,\tilde{N})^2$, and hence that $\tilde{N}$ is concentrated about its mean. Since $\mathbb{E}\,N \sim \mathbb{E}\,\tilde{N}$, and $\tilde{N}$ and N are whp equal, we thus have N concentrated about its mean also, where N is the number of x such that $F(x)$ holds.

Finally, the end of the proof is as in Section 2.2. Set $t_2 = \log(\varepsilon^3 n/\omega^2)/\log\lambda$, let $K\varepsilon \to \infty$ very slowly, let N be the number of vertices x for which $F(x)$ holds, and let M be the number of pairs x, y for which $A(x,y)$ holds (i.e., $F(x)$ and $F(y)$ hold disjointly) but $d(x,y) \leq d$, where

$$\begin{aligned} d = 2(t_0+t_1+q_0)+t_2-K &= 2\frac{\log(\varepsilon^3 n)}{\log(1/\lambda_\star)} + 2\frac{\log\omega}{\log\lambda} + \frac{\log(\varepsilon^3 n/\omega^2)}{\log\lambda} + O(1) + 2q_0 - K \\ &= \frac{\log(\varepsilon^3 n)}{\log\lambda} + 2\frac{\log(\varepsilon^3 n)}{\log(1/\lambda_\star)} + O(1) + 2q_0 - K. \end{aligned}$$

Given that $F(x)$ and $F(y)$ hold disjointly, the $(t_0+t_1+q_0)$-neighbourhoods of x and y each contain at most ω/ε vertices. Exploring from x and y in the obvious way, the rest of the graph is 'unseen', and the expected number of paths of length at most t_2-K joining one neighbourhood to the other is at most

$$\begin{aligned} (\omega/\varepsilon)^2 \sum_{k \leq t_2-K} n^{k-1}(\lambda/n)^k = \omega^2\varepsilon^{-2}n^{-1} \sum_{k\leq t_2-K} \lambda^k \\ = O(\omega^2\varepsilon^{-3}n^{-1}\lambda^{t_2-K}) = O(\lambda^{-K}) = o(1). \end{aligned}$$

(Here it is important that we work with $F(x)$ and not $\tilde{F}(x)$.) Hence, the conditional probability that $d(x,y) \leq d$ is $o(1)$, so $\mathbb{E}\,M = o\big(n(n-1)\mathbb{P}(A(x,y))\big) = o((\mathbb{E}\,N)^2)$, using (4.48). It follows that whp there are at least $\mathbb{E}\,N/2 \geq 2$ vertices x for which $F(x)$ holds, but at most $(\mathbb{E}\,N)^2/5$ pairs of vertices with $A(x,y)$ holding but $d(x,y) \leq d$. Using (4.57), it follows that $\mathrm{diam}(G) \leq d$ whp. Recalling that both $-q_0\varepsilon > 0$ and $K\varepsilon$ may be taken to tend to infinity arbitrarily slowly, this completes the proof of the lower bound in Theorem 1.3.

4.7 The upper bound

Throughout we fix a function $\varepsilon = \varepsilon(n) > 0$ satisfying $\varepsilon \to 0$ and $\varepsilon^3 n \to \infty$. As before we shall often write Λ for $\varepsilon^3 n$, and set

$$\omega = \Lambda^{1/6}.$$

As before, let $t_0 = \log(\varepsilon^3 n)/\log(1/\lambda_\star)$, $t_1 = \log\omega/\log\lambda$, and $t_2 = \log(\varepsilon^3 n/\omega^2)/\log\lambda$; we ignore rounding to integers, which makes no essential difference in our calculations.

Let $K = K(n)$ be such that $K\varepsilon \to \infty$, and let

$$d_0 = \log(\varepsilon^3 n)/\log\lambda + 2\log(\varepsilon^3 n)/\log(1/\lambda_\star) = 2t_0 + 2t_1 + t_2, \tag{4.58}$$

so our aim is to prove that $\mathrm{diam}(G) \le d_0 + K$ holds whp, and we may assume if we like that $K\varepsilon$ grows slower than any given function of n tending to infinity. The basic idea is to simply estimate the expected number of pairs x, y with $d(x,y) \ge d_0 + K$. However, the calculations in the previous sections imply that on its own, this will not work; the expectation turns out to be roughly ε^{-4} if $K\varepsilon$ grows slowly. The reason is that, given that a tree hanging off the 2-core has height at least h, the expected number of vertices it contains at distance at least h from the 2-core is of order ε^{-2}.

To get around this, we need to impose a version of the wedge condition; we should like to consider only vertices x that are at maximal distance from the 2-core in their tree. (Note that we cannot insist that x is the unique vertex at this distance in its tree, as we did before.) This suggests the *weak wedge condition*: roughly speaking, we should like any 'side branches' starting from $\Gamma_t(x)$ to have height at most t, one more than the height allowed in the strong wedge condition. This is all very well if the neighbourhoods of x out to the relevant distance form a tree, but in the upper bound we must consider *all* vertices x, so we must modify the condition. Unfortunately, most of the work in this section will be needed to show that we can rule out various unlikely cases (such as the diameter coming from a pair x, y where x is close to a short cycle).

Suppose that x and y are a pair of vertices at maximal distance, pick any $t \le d(x,y)$, and consider any shortest path P from x to y. Then, tracing P backwards from y to x, we first meet $G_{\le t}(x)$ at some vertex $v_t \in \Gamma_t(x)$. Since P is shortest, $d(v_t, y) = d(x,y) - t$, so continuing from v_t to x along the unique path in $G^0_{\le t}(x)$ joining these vertices, we find another shortest path P' from x to y that starts with $v_0 v_1 v_2 \cdots v_t$, a path in $G^0_{\le t}(x)$. We shall split the tree $G^0_{\le t}(x)$ into the *trunk* T, consisting of all vertices with descendants in $\Gamma_t(x)$, plus one *side branch* B_v for each $v \in T$. Here B_v consists of v together with all its descendants in $G^0_{\le t}(x)$ that are not descendants of another trunk vertex. (This corresponds roughly to the decomposition of $\mathfrak{X}_\lambda$ into $\mathfrak{X}^+_\lambda$ together with independent copies of $\mathfrak{X}_{\lambda_\star}$; the difference is that we only consider finitely many generations, as we must in the graph.)

Of course each v_i is a trunk vertex. The key observation is that for $0 \le i \le t$, the side branch B_{v_i} is either *short*, i.e., has height at most i, or is *reattached*, i.e., $B_{v_i} - v_i$ meets an edge of $G_{\le t}(x) \setminus G^0_{\le t}(x)$. Otherwise, let w be a vertex of B_{v_i} at maximum distance from v_i in B_{v_i}. Since B_{v_i} is not reattached, any path from w to y must pass via v_i. Since B_{v_i} is not short, the total length of such a path exceeds $d(x,y)$, contradicting the assumption that x and y are at maximum distance.

Given $1 \leq d \leq t$ and a vertex x, let S denote the set of vertices of $\Gamma_d(x)$ that have one or more descendants in $\Gamma_t(x)$, in the tree $G^0_{\leq t}(x)$. We say that x is (d,t)-*acceptable* if there is a vertex $v \in S$ such that every side branch in $G^0_{\leq t}(x)$ of the path $x = v_0 v_1 \cdots v_d = v$ is either short or reattached. From the observation above, if x and y are at maximal distance, then x and y must be (d,t)-acceptable for any $1 \leq d < t < d(x,y)$.

Set $h = \varepsilon^{-1} \log\log \Lambda$, say. (Here ε^{-1} times any slowly-enough growing function will do.) For $t > h$, let $A_t = A_t(x)$ be the event that x is (h,t)-acceptable, and let $B_t = B_t(x)$ be the event that $0 < |\Gamma_r(x)| < \omega/\varepsilon$ holds for $0 \leq r \leq t$. The following lemma will play a key role in our estimates.

Lemma 4.27. *Under the assumptions of Theorem 1.3 we have*

$$\mathbb{P}(A_t \cap B_t) \leq (1 + o(1)) 4\gamma_0 \varepsilon^3 \lambda_\star^{t-t_1} \tag{4.59}$$

uniformly in all t in the range $t_1 + 3h \leq t \leq 10\varepsilon^{-1} \log \Lambda$, where $\gamma_0 > 0$ is the constant appearing in Lemma 4.21.

Recall that, by the second part of Theorem 4.19, in the branching process $\mathfrak{X}_\lambda$ we have

$$\mathbb{P}\big(0 < |X_r| < \omega/\varepsilon,\, r = 0 \ldots t\big) \sim 4\varepsilon \lambda_\star^{t-t_1} \tag{4.60}$$

for any $t \leq 10\varepsilon^{-1} \log \Lambda$ such that $\varepsilon(t - t_1) \to \infty$; by Lemma 4.7, this carries over to the graph. Thus Lemma 4.27 says essentially that the conditional probability that our modified wedge condition holds is asymptotically $\gamma_0 \varepsilon^2$. We postpone the proof of the lemma for the moment.

Unfortunately, to handle the case when $\Lambda = \varepsilon^3 n$ grows slowly, it turns out that we need two further lemmas. The first is a very simple observation; once one thinks of the lemma, it is very easy to prove. We thought of it after seeing the preprint of Ding, Kim, Lubetzky and Peres [20].

Lemma 4.28. *Let $L = L(n)$ be any function satisfying $L = o(1/\varepsilon)$. Then, under the conditions of Theorem 1.3, whp the giant component of $G(n, \lambda/n)$ contains no cycle of length at most L.*

Proof. Fix $3 \leq \ell \leq L$ and a sequence $v_1, \ldots, v_\ell$ of distinct vertices of $G = G(n, \lambda/n)$. Let E be the event that this sequence forms a cycle, i.e., that the edges $v_1 v_2$, $v_2 v_3, \ldots, v_\ell v_1$ are all present, so $\mathbb{P}(E) = \lambda^\ell / n^\ell \sim n^{-\ell}$. Let F be the event that E holds and this cycle is in the giant component. First testing whether E holds, and then exploring outwards from this cycle, by comparison with the branching process as usual we see that $\mathbb{P}(F \mid E) = O(\ell s) = O(\varepsilon \ell)$, with the implicit constant universal. Hence $\mathbb{P}(F) = O(\varepsilon \ell n^{-\ell})$. Summing over all at most n^ℓ sequences, and dividing by 2ℓ to avoid overcounting, the expected number of ℓ-cycles in the giant component is thus $O(\varepsilon)$. Finally summing over $\ell \leq L$ and using Markov's inequality gives the result. □

Lemma 4.29. *Let $\psi = \psi(n)$ be some function of n tending to infinity slowly, with $\psi = O(\Lambda^{1/8})$ and $\psi = o(\varepsilon^{-1/10})$. Let $A^*(x)$ denote the event that $t_{\omega/\varepsilon}(x)$ is defined, x is (d,t)-acceptable for all $1 \leq d < t \leq t_{\omega/\varepsilon}(x)$, and $G_{\leq t}(x)$ is a tree for $t = \min\{t_{\omega/\varepsilon}(x), \varepsilon^{-1}/\psi\}$. Under the assumptions of Theorem 1.3 we have*

$$\mathbb{P}(A^*(x)) = O(\varepsilon^3 \psi^8) = O(\varepsilon^3 \Lambda).$$

(It is likely that the probability estimated above is $O(\varepsilon^3)$, at least if the quantity ε^{-1}/ψ in the definition of t is replaced by a small constant times ε^{-1}, but even a bound such as $O(\Lambda^{100}\varepsilon^3)$ would be more than enough for us here.)

Assuming Lemmas 4.27 and 4.29 for the moment, it is not hard to complete the proof of Theorem 1.3, calculating as in Section 2, by summing the expected number of pairs x, y with $t_{\omega/\varepsilon}$ in certain ranges and both having acceptable neighbourhoods.

Proof of Theorem 1.3. Let d_0 be defined by (4.58), and let $K = K(n)$ be such that $K\varepsilon \to \infty$ and $K \le \varepsilon^{-1} \log\log\Lambda$, say. Our aim is to show that whp there is no pair (x, y) of vertices in the same component with $d(x, y) \ge d_0 + K$. In the light of Łuczak's bound (1.7) from [32], and a standard duality argument, we need only consider the giant component. (In fact, Łuczak and Seierstad [33] have shown that in the random graph process, whp, for all densities in the range considered here, the diameter is realized by the giant component.)

Let us say that a vertex x is *tree-like* if $G_{\le\varepsilon^{-1}/\psi}(x)$ is a tree. By Lemma 4.28, whp *every* vertex in the giant component is tree-like, so it suffices to consider pairs (x, y) in which both x and y have this property.

As noted above, in any pair (x, y) at maximal distance greater than d_0, both x and y must be (d, t)-acceptable for any $d < t < d_0$. Set

$$t^+ = t_0 + t_1 + K/3,$$

noting that $t^+ < d_0/2$ and $t^+ > h = \varepsilon^{-1} \log\log\Lambda$. By Lemma 4.27, for any vertex x we have

$$\mathbb{P}(A_{t^+}(x) \cap B_{t^+}(x)) \le (4 + o(1))\gamma_0\varepsilon^3\lambda_\star^{t^+-t_1} = O(n^{-1}\lambda_\star^{K/3}) = o(n^{-1}),$$

so whp there is no vertex for which this event holds. Let $A'(x)$ be the event that $t_{\omega/\varepsilon}(x)$ is defined and at most t^+, and $A^*(x)$ holds, where $A^*(x)$ is defined in Lemma 4.29. Let us call (x, y) a *regular far pair* if $d(x, y) > d_0 + K$, and the events $A'(x)$ and $A'(y)$ hold. Then from the comments above it suffices to prove that whp there are no regular far pairs.

We may test whether $A'(x)$ holds by uncovering successive neighbourhoods of x, stopping at the first (if there is one) with at least ω/ε vertices, and then testing for acceptability and the tree condition, or stopping after t^+ steps if there is no such neighbourhood (in which case $A'(x)$ does not hold). By definition, each neighbourhood other than the last has at most ω/ε vertices. By Lemma 4.4, the probability that we find more than $2\omega/\varepsilon$ vertices in the last neighbourhood is at most $\exp(-\Omega(\omega/\varepsilon)) = \exp(-\Omega(\varepsilon^{-1/2}n^{1/6})) = o(n^{-100})$. Ignoring this event, testing $A'(x)$ involves uncovering

$$O(t^+\omega/\varepsilon) = O(\omega\log\Lambda/\varepsilon^2) = O(\Lambda^{1/3}\varepsilon^{-2}) = O(\Lambda^{1/3}\Lambda^{-2/3}n^{2/3}) = o(n^{2/3})$$

vertices. Also, we uncover $O(\omega/\varepsilon) = o(n^{1/3})$ vertices in each generation. Noting that $\varepsilon(t^+\omega/\varepsilon)^2 = O(\Lambda^{2/3}\varepsilon^{-3}) = O(n\Lambda^{-1/3}) = o(n)$, Lemmas 4.7 and 4.8 apply to the corresponding trees. By Lemma 4.8 it follows that for x and y distinct,

$$\begin{aligned}\mathbb{P}\big(A'(x) \cap A'(y) \cap \{d(x, y) > t_{\omega/\varepsilon}(x) + t_{\omega/\varepsilon}(y)\}\big)\\ = (1 + o(1))\mathbb{P}(A'(x))\mathbb{P}(A'(y)) + o(n^{-100}) = O(\Lambda^2\varepsilon^6), \quad (4.61)\end{aligned}$$

using $\mathbb{P}(A'(x)) \le \mathbb{P}(A^*(x))$ and Lemma 4.29 for the final bound. (In fact, we have glossed over something here: using Lemma 4.8 shows that the events that the explorations from x and y give certain trees consistent with $A'(x)$ and $A'(y)$ are asymptotically independent. However, the events $A'(z)$, $z = x, y$, depend not just on the trees, but also on any additional edges between the trees' vertices. Since these are present independently with probability λ/n, asymptotic independence of the trees gives asymptotic independence of the entire neighbourhoods.)

Suppose we have explored the neighbourhoods of x and y and found that the event described above holds, i.e., $A'(x)$ and $A'(y)$ hold disjointly. Then Lemma 4.5 applies, and the conditional probability that the explorations do not meet within $t_2 + 2\varepsilon^{-1}\log\log\Lambda$ further steps is $\exp\bigl(-(1+o(1))(\log\Lambda)^{2+o(1)}\bigr) + O(\Lambda^{-10}) = O(\Lambda^{-10})$. Summing over choices for x and y, we see that the expected number of regular far pairs with $d(x,y) \ge t_{\omega/\varepsilon}(x) + t_{\omega/\varepsilon}(y) + t_2 + 2\varepsilon^{-1}\log\log\Lambda$ is $O(n^2\Lambda^2\varepsilon^6\Lambda^{-10}) = O(\Lambda^{-6}) = o(1)$. Hence, whp there are no such pairs.

Set

$$t^- = t_0 + t_1 - 2\varepsilon^{-1}\log\log\Lambda,$$

noting that whp every vertex x in a regular far pair satisfies

$$t_{\omega/\varepsilon}(x) \ge d_0 + K - (t_2 + 2\varepsilon^{-1}\log\log\Lambda) - t^+ \ge t_0 + t_1 - 2\varepsilon^{-1}\log\log\Lambda = t^-. \tag{4.62}$$

This value is large enough that Lemma 4.27 applies.

(Let us remark that if $\Lambda \ge (\log n)^{20}$, say, then the argument above simplifies: we may replace $2\varepsilon^{-1}\log\log\Lambda$ by $2\varepsilon^{-1}\log\log n$, and the error probability given by Lemma 4.5 is then $o(n^{-100})$ (using the middle expression in (4.13)), so there is no need to check acceptability to conclude the equivalent of (4.62). In particular, there is no need for Lemma 4.29 in this case at all.)

For distinct vertices x and y and integers $t^- \le t, t' \le t^+$, let

$$E_{x,y,t,t'} = A_{[t]}(x) \cap \{t_{\omega/\varepsilon}(x) = t\} \cap A_{[t']}(y) \cap \{t_{\omega/\varepsilon}(y) = t'\} \cap \{d(x,y) \ge d_0 + K\},$$

where $[t]$ denotes the largest multiple of $\lfloor 1/\varepsilon \rfloor$ that is strictly smaller than t. From the comments above, to prove that $\operatorname{diam}(G) \le d_0 + K$ holds whp it suffices to prove that whp none of the events $E_{x,y,t,t'}$ holds. (Here we may impose whatever acceptability conditions we like: the reason for choosing exactly $A_{[t]}(x)$ will become clear in a moment.)

Using Lemma 4.8 as above, the probability that $E_1 = A_{[t]}(x) \cap \{t_{\omega/\varepsilon}(x) = t\}$ and $E_2 = A_{[t']}(y) \cap \{t_{\omega/\varepsilon}(y) = t'\}$ hold with disjoint witnesses is asymptotically $\mathbb{P}(E_1)\mathbb{P}(E_2)$. Noting that $d_0 + K - t - t' - t_2$ is within $5\varepsilon^{-1}\log\log\Lambda = o(t_2)$ of 0 and is hence at least $-t_2/2$, given that E_1 and E_2 hold disjointly, Lemma 4.5 tells us that the probability that $d(x,y) \ge d_0 + K$ is $\exp(-(1+o(1))\lambda^{d_0+K-t-t'-t_2}) + O(\Lambda^{-10})$.

Let $U = \bigcup_{x \ne y,\, t^- \le t,t' \le t^+} E_{x,y,t,t'}$. Then, writing $A \lesssim B$ for $A \le (1+o(1))B$,

$$\mathbb{P}(U) \lesssim n^2 \sum_{t=t^-}^{t^+} \sum_{t'=t^-}^{t^+} \mathbb{P}\bigl(A_{[t]}(x) \cap \{t_{\omega/\varepsilon}(x) = t\}\bigr)\mathbb{P}\bigl(A_{[t']}(x) \cap \{t_{\omega/\varepsilon}(x) = t'\}\bigr)$$
$$\Bigl(\exp\bigl(-(1+o(1))\lambda^{d_0+K-t-t'-t_2}\bigr) + O(\Lambda^{-10})\Bigr).$$

Grouping the sums into blocks of size $k = \lfloor 1/\varepsilon \rfloor$, and noting that if r is a multiple of k then

$$\sum_{t=r+1}^{r+k} \mathbb{P}\big(A_{[t]}(x) \cap \{t_{\omega/\varepsilon}(x) = t\}\big) = \mathbb{P}\big(A_r(x) \cap \{r < t_{\omega/\varepsilon}(x) \le r+k\}\big) \le \mathbb{P}(A_r(x) \cap B_r(x)),$$

we have

$$\mathbb{P}(U) \lesssim n^2 \sum_{t^- - k \le t \le t^+}{}' \sum_{t^- - k \le t' \le t^+}{}' \mathbb{P}(A_t(x) \cap B_t(x)) \mathbb{P}(A_{t'}(x) \cap B_{t'}(x)) \\ \left(\exp\big(-(1+o(1))\lambda^{d_0+K-t-t'-t_2-2k}\big) + O(\Lambda^{-10})\right),$$

where primes denote sums that run over multiples of k. From Lemma 4.27 we thus have

$$\begin{aligned} \mathbb{P}(U) &\lesssim n^2 \sum_{t,\,t'}{}' 16\gamma_0^2 \varepsilon^6 \lambda_\star^{t+t'-2t_1} \left(\exp\big(-(1+o(1))\lambda^{d_0+K-t-t'-t_2-2k}\big) + O(\Lambda^{-10})\right), \\ &= o(1) + n^2 \sum_{t,\,t'}{}' 16\gamma_0^2 \varepsilon^6 \lambda_\star^{t+t'-2t_1} \exp\big(-(1+o(1))\lambda^{d_0+K-t-t'-t_2-2k}\big), \end{aligned}$$

since there are at most $(\varepsilon t^+)^2 = O((\log \Lambda)^2)$ terms in the double sum (which has the same limits as before), so the contribution of the $O(\Lambda^{-10})$ term can be bounded by $16n^2\gamma_0^2\varepsilon^6(\log\Lambda)^2 O(\Lambda^{-10}) = O(\Lambda^{-8}(\log \Lambda)^2) = o(1)$.

Taking the final term in the sums above, we have t and t' at least $t^+ - k$, so the exponent of λ above is at least $d_0 + K - 2t^+ - t_2 - 4k = K/3 - 4k$, which is at least $K/4$ if n is large. Hence the exponential term above is *always* at most $\exp(-\lambda^{K/4}/2)$, say. Taking the final term in the sum, the corresponding $\lambda_\star^{\cdots}$ term is at most

$$\lambda_\star^{2t^+-2k-2t_1} = \lambda_\star^{2t_0+2K/3-2k} \le \lambda_\star^{2t_0} = \varepsilon^{-6} n^{-2}.$$

As $t+t'$ decreases from its maximum possible value in steps of k, the exponent of λ in the exponential increases by $k \sim 1/\varepsilon \sim 1/\log\lambda$, so the $\lambda^{\cdots}$ term increases by a factor that is asymptotically e and certainly at least 2. The $\lambda_\star^{\cdots}$ term increases by a factor of $\lambda_\star^{-k}$ which is asymptotically e and certainly at most 3. Also, after r steps, there are at most $r+1$ ways of realizing a given sum $t+t'$. It follows that

$$\mathbb{P}(U) \le o(1) + \sum_{r=0}^{\infty} 16\gamma_0^2 (r+1) 3^r \exp(-\lambda^{K/4} 2^{r-1}),$$

say. Since $\lambda^{K/4} \to \infty$, the exponential term in the final sum decreases extremely rapidly, and the whole sum is dominated by its first term, which is $o(1)$. This completes the proof of Theorem 1.3, assuming Lemmas 4.27 and 4.29. □

Let us note for later, when we come to consider the distribution of the diameter, that if we modify the definition of $E_{x,y,t,t'}$ by replacing $d(x,y) \geq d_0 + K$ by $d(x,y) \geq d_0 - K$, then we obtain

$$\mathbb{P}(U) \leq \sum_{r=0}^{\infty} 16\gamma_0^2(r+1)3^r \exp(-\lambda^{K/4-2K}2^{r-1}).$$

Indeed, everything is as before except that the exponent of λ has decreased by $2K$. Now this new sum is large, but the contribution from terms with $r \geq \log(\lambda^{3K})/\log 2 \sim 3K\varepsilon/\log 2$ is still small. Hence, the sum from terms in which one of t, t' is smaller than t^+ by more than $3\lfloor 1/\varepsilon \rfloor K\varepsilon/\log 2 \sim 3K/\log 2 \leq 5K$ is small. Since $\operatorname{diam}(G) \geq d_0 - K$ whp, it follows that whp the diameter is realized by vertices x and y which form a regular far pair in which each vertex z has $t_0 + t_1 - 5K \leq t_{\omega/\varepsilon}(z) \leq t^+ = t_0 + t_1 + K/3$. Since $K\varepsilon$ may be taken to tend to infinity arbitrarily slowly, this says that for a given error probability, it suffices to consider regular far pairs in which the vertices satisfy

$$t_{\omega/\varepsilon}(z) = t_0 + t_1 + O(1/\varepsilon). \tag{4.63}$$

It remains to prove Lemmas 4.27 and 4.29.

Proof of Lemma 4.27. Recall that $t_1 + 3h \leq t \leq 10\varepsilon^{-1}\log\Lambda$, and $h = \varepsilon^{-1}\log\log\Lambda < t/2$. Let $A = A(x)$ denote the event that x is (h,t)-acceptable, and $B_t = B_t(x)$ the event that $0 < |\Gamma_r(x)| < \omega/\varepsilon$ holds for $0 \leq r \leq t$. Our aim is to bound the probability of $A \cap B_t$; note that this event depends only on $G_{\leq t}(x)$.

To avoid dependence, we'd like to work with the branching process rather than the graph, but we cannot assume that the relevant neighbourhoods of x are trees. So let us model the pair $(G^0_{\leq t}(x), G_{\leq t}(x))$ by a pair $(T^\star, G^\star)$ as follows: first construct the branching process $(X_r)_{0\leq r\leq t}$, keeping track of the order in which the particles are born, as in the proof of Lemma 4.7. Let $T^\star$ be the corresponding labelled rooted tree of height at most t. Given $T^\star$, i.e., given (X_r), form $G^\star$ by starting with $T^\star$ and adding each of the following 'potential extra edges' independently with probability λ/n: all possible edges within X_r and, for each $v \in X_r$, all possible edges from v to children (in X_{r+1}) of earlier particles $v' \in X_r$. The potential extra edges correspond to edges that would not have been tested in the graph exploration, so the conditional distribution of $G^\star$ given $T^\star$ is the same as that of $G_{\leq t}(x)$ given $G^0_{\leq t}(x)$ (with an order on the vertices). If (T_0, G_0) is any possible value of $(G^0_{\leq t}(x), G_{\leq t}(x))$ consistent with $A \cap B_t$, then since B_t holds, T_0 is a tree to which Lemma 4.7 applies. So $\mathbb{P}(G^0_{\leq t}(x) \cong T_0) \sim \mathbb{P}(T^\star \cong T_0)$. It follows that $\mathbb{P}(G_{\leq t}(x) \cong G_0) \sim \mathbb{P}(G^\star \cong G_0)$. Hence, $\mathbb{P}((T^\star, G^\star) \in A \cap B_t)$ is asymptotically equal to the probability that $(G^0_{\leq t}(x), G_{\leq t}(x)) \in A \cap B_t$. From now on we consider the model $(T^\star, G^\star)$, forgetting about the graph $G(n, \lambda/n)$.

For technical reasons we modify $G^\star$ slightly as follows: recalling that each set X_r comes with an order, we only test for possible extra edges vw when both endvertices are among the first ω/ε vertices in the relevant set(s) X_r. This does not affect the probability of $A \cap B_t$, since when B_t holds (which is determined by $T^\star$), the distribution of $G^\star$ given $T^\star$ is unchanged.

Let S be the set of particles in X_h with descendants in X_t. To achieve independence between A and B_t, let us weaken $B = B_t$ to $B' = B'_t$, the condition that for every $v \in S$, the number of descendants of v in each X_r, $h \le r \le t$, is at most ω/ε. Our aim is to bound $\mathbb{P}(A\cap B)$ by $\mathbb{P}(A\cap B')$; to evaluate the latter we estimate $\mathbb{P}(B')$ and $\mathbb{P}(A \mid B')$.

Our first aim is to show that

$$\mathbb{P}(B') \sim \mathbb{P}(B) = p_0 \sim 4\varepsilon\lambda_\star^{t-t_1}, \tag{4.64}$$

where the final estimate is from (4.60). Note that $\mathbb{P}(B' \mid |S| = s) = p^s$, where p is the (unconditional) probability that $0 < |X_r| < \omega/\varepsilon$ holds for $0 \le r \le t-h$. From Theorem 4.19, we have $p \sim 4\varepsilon\lambda_\star^{t-h-t_1} \sim p_0\lambda_\star^{-h}$. Also, since $t \ge t_1 + 3h$, we have $p \le (1+o(1))\lambda_\star^{2h}$. Since $\varepsilon h \to \infty$ it follows that $p^2 \le (1+o(1))\lambda_\star^h p_0 = o(p_0)$.

Since $B \subset B'$, we have

$$\mathbb{P}\big(B\cap\{|S|\ge 2\}\big) \le \mathbb{P}\big(B'\cap\{|S|\ge 2\}\big) \le \mathbb{P}\big(B' \mid \{|S|\ge 2\}\big) \le p^2 = o(\mathbb{P}(B)).$$

Recalling that if B or B' holds then $|S|\ge 1$, to show that $\mathbb{P}(B')\sim\mathbb{P}(B)$ it suffices to show that $\mathbb{P}(B\cap\{|S|=1\}) \sim \mathbb{P}(B'\cap\{|S|=1\}) = p$. Let B'' be a strengthened version of B', where we replace the upper bound ω/ε by ω'/ε, with $\omega' = (1-1/\log\Lambda)\omega \sim \omega$. Applying Theorem 4.19 again with this new value of ω', we find that $\mathbb{P}(B'' \mid |S|=1) \sim p$. But given that $|S|=1$ and B'' holds, B certainly holds as long as the tree T formed by the descendants of the root that are not descendants of the unique particle in S contains at most $\omega/(\varepsilon\log\Lambda) > \Lambda^{1/10}/\varepsilon$ particles in each generation.

The distribution of T is dominated by that of the tree T' formed by starting one copy of $\mathfrak{X}_{\lambda_\star}$ in each generation $0 \le t < h$. (In T these copies are conditioned to die by a specific time.) The first $h-1$ generations of T' have exactly the distribution of the process (D_t) studied in Lemma 4.22. Hence, by the second part of that lemma, the probability that one of the first h generations of T exceeds size $\Lambda^{1/20}/\varepsilon$ is $O(\varepsilon h e^{-\Omega(\Lambda^{1/20})}) = o(1)$. From generation h onwards, the tree T evolves as a subcritical branching process, and from a standard martingale argument the probability that any later generation exceeds the size of generation h by a factor of $\Lambda^{1/20}$ is at most $1/\Lambda^{1/20} = o(1)$. Thus we do indeed have $\mathbb{P}(B \mid B''\cap\{|S|=1\}) \sim 1$, and it follows that $\mathbb{P}(B')\sim\mathbb{P}(B)$, as claimed.

Recalling that $p^2 = o(\mathbb{P}(B))$ and hence $p^2 = o(\mathbb{P}(B'))$, for $r\ge 2$ we have

$$\mathbb{P}(|S|=r \mid B') \le \mathbb{P}(B' \mid |S|=r)/\mathbb{P}(B') = o(p^{r-2}).$$

Summing, it follows that

$$\mathbb{E}(|S| \mid B') \sim 1. \tag{4.65}$$

We claim that (in the modified $G^\star$ model)

$$\mathbb{P}(A \mid B'\cap\{|S|=N\}) \le (1+o(1))\gamma_0 N\varepsilon^2. \tag{4.66}$$

Using $\mathbb{P}(A \mid B') = \sum_{N\ge 1}\mathbb{P}(|S|=N \mid B')\mathbb{P}(A \mid B'\cap\{|S|=N\})$, and (4.65) and (4.64), the required bound (4.59) on $\mathbb{P}(A\cap B) \le \mathbb{P}(A\cap B')$ then follows.

It remains only to prove (4.66). Recall that we are working with the model $(T^\star, G^\star)$. Let us construct $T^\star$ (which is simply the first t generations of $\mathfrak{X}_\lambda$) by decomposing it into the trunk and side branches exactly as in the graph. Thus the *trunk* consists of the subtree T' of

$T^\star$ consisting of all particles with descendants in X_t. Then $T^\star$ may be formed by adding for each v in generation r, $0 \le r < t$, of T' a copy W_v of the process $(X_{t'})_{0\le t'\le t-r}$ conditioned on X_{t-r} being empty. We may think of W_v as the subcritical process $\mathfrak{X}_{\lambda_\star}$ conditioned on dying out by time $t-r$.

Now whether B' holds is determined by T' together with the trees W_v for vertices v in sets X_r, $r \ge h$. Let us condition on T' and these trees W_v; the only remaining randomness is in the W_v for $v \in X_r$, $r < h$.

Let v be one of the $N = |S|$ vertices in S, and let $x = v_0 v_1 \cdots v_h = v$ be the path to v. Let $W_i = W_{v_i}$, for $0 \le i \le h-1$. Let A_v be the event that every W_i is either short or, when we come to $G^\star$, reattached. Note that A holds if and only if one of the events A_v holds, so it suffices to prove that the conditional probability of A_v is $(1+o(1))\gamma_0\varepsilon^2$. Since the different W_w are independent given T', the conditional distribution of each W_i (given T' and the W_w, $w \in X_r$, $r \ge h$) is just the unconditioned distribution. Writing, as before, $s_i = \mathbb{P}(|X_i^-| \ge 0)$ and $d_i = 1 - s_i$, the probability that W_i is *tall* (not short) is just

$$p_i = \mathbb{P}\big(|X_{i+1}| > 0 \bigm| |X_{t-i}| = 0\big) = \mathbb{P}\big(|X_{i+1}^-| > 0 \bigm| |X_{t-i}^-| = 0\big)$$
$$= \frac{d_{t-i} - d_{i+1}}{d_{t-i}} = \frac{s_{i+1} - s_{t-i}}{1 - s_{t-i}} = s_{i+1} - O(s_{t-i}) = s_{i+1} - O(s_h), \quad (4.67)$$

since $t \ge 2h$ and $i \le h$.

Now $h \ge 1/\varepsilon$, so (by Lemma 4.21) $s_h = O(\varepsilon\lambda_\star^h)$. Let w be the number of tall W_i. Then from the estimate above and Lemma 4.21,

$$\mathbb{P}(w=0) = \prod_{i=0}^{h-1}(1-p_i) = \exp(O(h\varepsilon\lambda_\star^h))\prod_{i=1}^{h}(1-s_i) \sim \prod_{i=1}^{\infty}(1-s_i) \sim \gamma_0\varepsilon^2,$$

since $\lambda_\star^h = \exp(-(1+o(1))\varepsilon h)$ and $\varepsilon h \to \infty$, so $h\varepsilon\lambda_\star^h = o(1)$. It thus suffices to show that $\mathbb{P}(A_v) \le (1+o(1))\mathbb{P}(w=0)$; then (4.66) follows by the union bound. In other words, we must show that $A_v \cap \{w > 0\}$ is much less likely than $w = 0$.

Let I be any subset of $\{0,1,2,\ldots,h-1\}$ with $|I| \ge 1$, and let us condition on precisely the corresponding trees $W_i : i \in I$ being tall. Let M_i, $i \in I$, be the number of vertices in each tall tree W_i, noting that these numbers are conditionally independent. Given that a particular W_i is tall, its average size is at most that of $\mathfrak{X}_{\lambda_\star}$ conditioned to survive to height $i+1$ (we also condition on dying out by height $t-i$). By Lemma 4.23 this is at most $(i+2)/\varepsilon$.

Let us now go through the tall trees in order, checking to see whether each is reattached. (We will be forced to skip some; see below.) Due to the way we modified $G^\star$, when checking if W_i is reattached, for each vertex u of W_i with $u \in X_r$ we need only check for edges of $G^\star \setminus T^\star$ between u and up to ω/ε vertices in each of X_{r-1}, X_r and X_{r+1}. For each u, the probability of finding such an edge is at most $p = 3(\omega/\varepsilon)\lambda/n \le 4\omega\varepsilon^{-1}n^{-1}$. The probability that the tall tree W_i reattaches is thus at most $\mathbb{E}(M_i p) = \mathbb{E}(M_i)p \le (i+2)\varepsilon^{-1}p \le 4(i+2)\omega\varepsilon^{-2}n^{-1}$.

When testing whether the first tall tree does reattach, we stop if we find one edge witnessing this. This edge may 'spoil' a later tall W_j by going to a vertex of that W_j. For $J \subset I$, let E_J be the event that the tall trees $W_j : j \in J$ are reattached by $|J|$ edges each with one end in the appropriate W_j and the other outside $\bigcup_{i\in J} W_i$. Given all the trees, the conditional probability of E_J is at most $\prod_{j\in J} M_j p$. Since, conditioning only on which trees are tall but not their sizes, the M_j are independent, it follows that

$$\mathbb{P}(E_J \mid I) \le \prod_{j\in J} \frac{4(j+2)\omega}{\varepsilon^2 n}.$$

If all $W_i : i \in I$ are reattached, then the testing algorithm above shows that E_J must hold for some J containing at least half of the first k elements of I for every $k \le |I|$, corresponding to the fact that we test trees in order, and each spoils at most one later one. Hence,

$$\mathbb{P}(A_v \mid I) \le \sum_J \prod_{j\in J} \frac{4(j+2)\omega}{\varepsilon^2 n},$$

with the sum restricted as above.

Suppose that $|I| = 2k-1$ or $|I| = 2k$, and list the elements of I as $i_1, i_2, \ldots$ in order. There are at most 4^k terms in the sum, and the largest has $J = \{i_1, i_3, i_5, \ldots, i_{2k-1}\}$, so given I, the probability of reattachment is at most

$$\frac{16(i_1+2)\omega}{\varepsilon^2 n} \frac{16(i_3+2)\omega}{\varepsilon^2 n} \cdots \frac{16(i_{2k-1}+2)\omega}{\varepsilon^2 n}.$$

Now the probability that the tall trees are exactly those indexed by I is

$$\mathbb{P}(w=0) \prod_{i\in I} \frac{p_i}{1-p_i} \le \mathbb{P}(w=0) \prod_{i\in I} 3p_i \le \mathbb{P}(w=0) \prod_{i\in I} \frac{10}{i+2},$$

say, noting that $p_i \le s_{i+1}$ and using the crude upper bound $3/(i+2)$ for s_{i+1}. Summing over I with $|I| \ge 1$ we find that

$$\begin{aligned}\mathbb{P}(A_v \cap \{w>0\}) & \\ &\le \mathbb{P}(w=0) \sum_{r\ge 1} \sum_{0\le i_1<i_2<\cdots<i_r<h} \frac{10}{i_1+2}\frac{10}{i_2+2} \cdots \frac{16(i_1+2)\omega}{\varepsilon^2 n}\frac{16(i_3+2)\omega}{\varepsilon^2 n} \cdots .\end{aligned}$$

The sum over even r, say $r = 2k$, may be crudely bounded by $\sum_{k=1}^{\infty} S^k$, where

$$S = \sum_{0\le a<b<h} \frac{10}{a+2}\frac{10}{b+2}\frac{16(a+2)\omega}{\varepsilon^2 n} \le \sum_{b<h} b \frac{1600}{b+2}\frac{\omega}{\varepsilon^2 n} \le 1600 h\omega\varepsilon^{-2} n^{-1}.$$

Since $h \le \varepsilon^{-1}\log\log\Lambda$, we have $S = o(1)$. Bounding the sum over odd r similarly, it follows that $\mathbb{P}(A_v \cap \{w>0\}) = o(\mathbb{P}(w=0))$, as required. □

Finally, we prove Lemma 4.29.

Proof of Lemma 4.29. Throughout this proof, let $K = \lceil \log(1/\varepsilon) \rceil$ and, for $1 \le k \le K$, let $t_k = \varepsilon^{-1}/(k\psi)$. (We ignore the irrelevant rounding to integers, noting that $t_K \to \infty$.)

For $2 \le k \le K$ let $E_k(x)$ denote the event that $t_k < t_{\omega/\varepsilon}(x) \le t_{k-1}$, the neighbourhoods of x to distance $t_{\omega/\varepsilon}(x)$ form a tree, and x is $(t_k/2, t_k)$-acceptable. Let $E_1(x)$ denote the event that B_{t_1} holds (i.e., $0 < |\Gamma_t(x)| < \omega/\varepsilon$ for $0 \le t \le t_1$), that $G_{\le t_1}(x)$ is a tree, and x is $(t_1/2, t_1)$-acceptable. Finally, let $E_\infty(x)$ denote the event that $t_{\omega/\varepsilon}(x) \le t_K$. Splitting into cases according to the value of $t_{\omega/\varepsilon}(x)$, we see that if $A^*(x)$ holds, then so does one of the events $E_1(x)$, $E_\infty(x)$ or $E_k(x)$, $2 \le k \le K$.

Let us start with a simple branching process observation related to that in Lemma 4.22, writing $t_{\omega/\varepsilon}$ for $\min\{t : |X_t| \ge \omega/\varepsilon\}$, as before, whenever this is defined. Suppose we have chosen some $t \ge 1$ in advance. If we explore the branching process step by step and find a generation X_r, $r \le t$, with size at least ω/ε, then it is easy to see that the conditional probability that $|X_t| \ge \omega/\varepsilon$ is at least 1/10, say. Thus $\mathbb{P}(|X_t| \ge \omega/\varepsilon) \ge \mathbb{P}(t_{\omega/\varepsilon} \le t)/10$, and hence

$$\mathbb{P}(t_{\omega/\varepsilon} \le t) \le 10\mathbb{P}(|X_t| \ge \omega/\varepsilon). \tag{4.68}$$

Using this observation and Lemma 4.20, we see that

$$\mathbb{P}(t_{\omega/\varepsilon} \le t_K) \le 10 t_K^{-1} e^{-\omega\varepsilon^{-1}t_K^{-1}/20} = 10 t_K^{-1} e^{-\omega\psi K/20} = o(\varepsilon^3),$$

since $\omega\psi \to \infty$ while $K \ge \log(1/\varepsilon)$. Comparing the graph and branching process as usual, it follows that $\mathbb{P}(E_\infty(x)) = o(\varepsilon^3)$.

Turning to $E_k(x)$ for $1 \le k \le K$, note that we may test whether this event holds by exploring at most $t_1 = O(1/\varepsilon)$ steps from x, stopping if we reach a neighbourhood of size ω/ε, and then checking that the neighbourhoods so far form a tree, and satisfy the relevant acceptability conditions. Arguing as above (4.61), Lemma 4.7 thus gives $\mathbb{P}(E_k(x)) = (1 + o(1))\mathbb{P}(E_k) + O(n^{-100})$, where E_k is the branching process event corresponding to $E_k(x)$. It thus suffices to show that

$$\sum_{k=1}^{K} \mathbb{P}(E_k) = O(\varepsilon^3\psi^8). \tag{4.69}$$

This statement involves only the branching process $\mathfrak{X}_\lambda$, so from now on we work with this rather than the graph.

Let A_k be the event that the branching process satisfies the condition corresponding to $(t_k/2, t_k)$-acceptability. To simplify the arguments, for $2 \le k \le K$ let E'_k be the event that A_k holds and $|X_{t_{k-1}}| \ge \omega/\varepsilon$. Only the second condition involves generations beyond t_k, so arguing as for (4.68) we have $\mathbb{P}(E'_k \mid E_k) \ge 1/10$, and hence $\mathbb{P}(E_k) \le 10\mathbb{P}(E'_k)$. Also, let E'_1 be the event that A_1 holds and $|X_{t_1}| > 0$. Then $E'_1 \supset E_1$. Hence

$$\mathbb{P}(E_k) \le 10\mathbb{P}(E'_k) \tag{4.70}$$

for all $1 \le k \le K$.

For $k \le 2$ let L_k be the event that $|X_{t_{k-1}}| \ge \omega/\varepsilon$; let L_1 be the event that $|X_{t_1}| > 0$, so $E'_k = A_k \cap L_k$. As before, let T be the *trunk* of $\mathfrak{X}_\lambda$ defined up to generation t_k, so T is the random tree consisting of all particles with descendants in X_{t_k}. If we condition on the first t_k generations of $\mathfrak{X}_\lambda$, then the conditional probability of L_k depends only on $|X_{t_k}|$. Since knowing the trunk T determines $|X_{t_k}|$, we thus have

$$\mathbb{P}(E'_k) = \mathbb{P}(A_k \cap L_k) = \sum_{T'} \mathbb{P}(T = T')\mathbb{P}(A_k \mid T = T')\mathbb{P}(L_k \mid T = T'),$$

where the sum runs over all possible trunks T'. Note that we may assume T' is non-empty, i.e., $|X_{t_k}| > 0$, as otherwise L_k cannot hold.

As before, given the trunk, we may reconstruct $X_{\le t_k}$ by adding independent random branches to each trunk vertex, with each branch a copy of $\mathfrak{X}_{\lambda_\star}$ conditioned to die by (absolute, not relative) time t_k. Let S be the set of trunk vertices in generation $t_{k/2}$, and $N = |S|$ the number of such vertices, so N is random but depends only on T. Since we are considering the branching process, which is by definition a tree, the acceptability condition A_k holds if and only if some $v \in S$ has the property that the side branch started at each v_i has height at most i for all $0 \le i \le t_k/2$, where $v_0 v_1 v_2 \cdots v_{t_k/2} = v$ is the chain of ancestors of v. For a given v, the probability of this event is exactly $\prod_{i=0}^{t_k/2}(1-p_i)$, where p_i is given by (4.67) with $t = t_k$ and h replaced by $t_k/2$. (The argument is as for (4.67).) It follows easily from the estimates in Lemma 4.21 that $s_{t_k/2} = O(t_k^{-1})$ and that

$$\prod_{i=0}^{t_k/2}(1-p_i) = \Theta(1)\prod_{i=1}^{t_k/2}(1-s_i) = O(t_k^{-2}).$$

So far we considered a single $v \in S$; by the union bound it follows that $\mathbb{P}(A_k \mid T = T') \le C t_k^{-2} N(T')$ for some absolute constant C. Hence,

$$\mathbb{P}(E'_k) \le C t_k^{-2} \sum_{T'} \mathbb{P}(T = T') N(T') \mathbb{P}(L_k \mid T = T'). \tag{4.71}$$

Let $n_0 = n_0(k) = (k\psi)^5$. Let μ_k^- and μ_k^+ denote respectively the contributions to the sum in (4.71) from trunks T' with $N(T') \le n_0$ and $N(T') > n_0$, so $\mathbb{P}(E'_k) \le \mu_k^- + \mu_k^+$. Trivially, we have

$$\mu_k^- \le C t_k^{-2} n_0 \sum_{T'} \mathbb{P}(T = T')\mathbb{P}(L_k \mid T = T') = C t_k^{-2} n_0 \mathbb{P}(L_k). \tag{4.72}$$

For $k = 1$ we have $\mathbb{P}(L_1) = \mathbb{P}(|X_{t_1}| > 0)$. Writing $\mathcal{S}$ for the event that the whole process survives, we have

$$\mathbb{P}(|X_t| > 0) = s + (1-s)\mathbb{P}(|X_t| > 0 \mid \mathcal{S}^{\mathrm{c}}) = s + (1-s)\mathbb{P}(|X_t^-| > 0).$$

By Lemma 4.21, it follows that for $t = o(1/\varepsilon)$ we have

$$\mathbb{P}(|X_t| > 0) \sim 2/t. \tag{4.73}$$

In particular, $\mathbb{P}(L_1) = O(1/t_1)$, so from (4.72)

$$\mu_1^- = O(t_1^{-3} n_0) = O(\varepsilon^3 \psi^3 \psi^5) = O(\varepsilon^3 \psi^8).$$

For $k \ge 2$, from Lemma 4.20 we have

$$\mathbb{P}(L_k) = \mathbb{P}\big(|X_{t_{k-1}}| \ge \omega/\varepsilon\big) \le t_{k-1}^{-1} e^{-\omega\psi(k-1)/20},$$

so, from (4.72), $\mu_k^- \le 10Ct_k^{-2}n_0t_{k-1}^{-1}e^{-\psi\omega(k-1)/20}$. Recalling that $t_k = \varepsilon^{-1}/(\psi k)$ and $n_0 = n_0(k) = k^5\psi^5$, it follows that

$$\sum_{k=2}^{K} \mu_k^- \le \sum_{k\ge 2} 10C\varepsilon^3k^8\psi^8e^{-\psi\omega(k-1)/20}.$$

Since ω and ψ are large for n large, the first term dominates, and this sum is $o(\varepsilon^3)$. Together with the bound for μ_1^- above this gives

$$\sum_{k=1}^{K} \mu_k^- = O(\varepsilon^3\psi^8). \tag{4.74}$$

It remains to bound μ_k^+. Noting that $N(N-1) \ge n_0N$ whenever $N > n_0$, and that $\mathbb{P}(L_k \mid T = T') \le 1$, from (4.71) we have

$$\mu_k^+ \le Ct_k^{-2}\sum_{T'} \mathbb{P}(T = T')n_0^{-1}N(T')(N(T')-1) = Ct_k^{-2}n_0^{-1}\,\mathbb{E}\big(N(N-1)\big),$$

where the final expectation is unconditional. Given $X_{t_k/2}$, each particle in this generation survives to generation t_k independently with probability $p = \mathbb{P}(|X_{t_k/2}| > 0) = O(t_k^{-1})$, from (4.73). Hence

$$\mathbb{E}\big(N(N-1)\big) = p^2\,\mathbb{E}\big(|X_{t_k/2}|(|X_{t_k/2}|-1)\big) = O(t_k^{-2})\,\mathbb{E}\big(|X_{t_k/2}|(|X_{t_k/2}|-1)\big).$$

A simple inductive formula, or a tree counting argument, gives $\mathbb{E}\big(|X_t|(|X_t|-1)\big) = \lambda^t(\lambda + \lambda^2 + \cdots + \lambda^t) \le t\lambda^{2t}$. With $t = t_k/2 \le 1/\varepsilon$, this is $O(t_k)$, so $\mathbb{E}\big(N(N-1)\big) = O(t_k^{-1})$. Hence,

$$\mu_k^+ = O(t_k^{-3}n_0^{-1}) = O\big(\varepsilon^3\psi^3k^3(k\psi)^{-5}\big) = O(\varepsilon^3k^{-2}).$$

Thus $\sum_{k=1}^{K}\mu_k^+ = O(\varepsilon^3)$. Recalling that $\mathbb{P}(E_k') \le \mu_k^- + \mu_k^+$, and using (4.74) and (4.70), this establishes (4.69). As noted earlier, the lemma follows. □

Remark. As noted earlier, in the first draft of this paper we needed the condition $\Lambda \ge e^{(\log^* n)^4}$. The changes that allowed us to eliminate this are the introduction of Lemma 4.28 (making checking for acceptability in the case when $t_{\omega/\varepsilon}(x) = o(1/\varepsilon)$ much simpler), the modification of Lemma 4.29 to include the tree condition, and the new proof of Lemma 4.29 above.

5 The distribution of the correction term

In this section we shall describe the limiting distribution of the correction term in Theorem 1.3 and, very briefly, that in Theorem 1.1. Surprisingly, although Theorem 1.3 is much harder to prove than Theorem 1.1, the study of the correction term is much easier in the former case. Indeed, with $p = \lambda/n$ and λ constant, even the description of the correction term is rather complicated. Let us start with the simpler case, assuming that $\lambda = 1+\varepsilon$ with $\varepsilon = \varepsilon(n) \to 0$.

It turns out that given the results of the previous section, not much extra work is needed to obtain the distribution. Essentially, only one natural extra idea is needed. Since the formal details would take some time to write out, we shall only sketch the arguments.

In Subsection 4.6, we obtained a lower bound on the diameter by considering vertices x with a certain property $F = F_q$, $q = q_0$, depending on the $t = t_0 + t_1 + q$ neighbourhoods, where $|q\varepsilon| \le M$ was essentially bounded. (We shall repeatedly use the observation that if some probability is $o(1)$ uniformly in $|q\varepsilon| \le M$ for any constant M, then it is $o(1)$ uniformly in $|q\varepsilon| \le M$ if $M = M(n)$ tends to infinity slowly enough. It is often easier to think of M as constant, although in the end we need $M \to \infty$.)

One aspect of this property F_q, or rather of the related property $\tilde{F}_q$, was that in the tree T_x containing x and attached to the 2-core, x is the unique vertex at maximal distance from the 2-core. It turns out that a positive fraction of the trees attached to the 2-core have more than one vertex at maximal distance, and to obtain a precise result we must also consider such trees. But we must only count each tree once. The solution is very natural: we consider an auxiliary random order $\prec$ on $V(G)$, and consider only vertices x such that, writing S_x for the set of vertices of T_x at maximal distance from the 2-core, x is the first vertex of S_x in the order $\prec$.

More precisely, we modify the definition of the branching process events E_q and F_q, by weakening the 'strong wedge condition' B(i) on page 65: instead of insisting that the 'side branch' starting at generation i dies within i generations, we insist that it dies within $i+1$ generations (this is the *weak wedge condition*), and also, writing S_i for the set of particles in the ith generation of the ith side branch, letting S be the union of the sets S_i together with the initial particle, and taking a random order on S, we insist that the initial particle comes first in this order; we call this the *medium wedge condition*.

We showed that the probability of the strong wedge condition was asymptotically $d_1 \prod_{i=1}^{\infty} d_i \sim d_1\gamma_0\varepsilon^2 \sim e^{-1}\gamma_0\varepsilon^2$, where $d_i = \mathbb{P}(|X_i^-| = 0) = 1 - s_i$ is the probability that the subcritical process dies by time i. Similarly, the probability of the weak wedge condition is asymptotically $\prod_{i=1}^{\infty} d_i \sim \gamma_0\varepsilon^2$.

If we condition on the weak wedge condition, then the distribution of S depends on ε. However, the conditional probability that S_i is non-empty is bounded by

$$\begin{aligned}\mathbb{P}\big(|X_i^-| > 0 \bigm| |X_{i+1}^-| = 0\big) &= 1 - \mathbb{P}\big(|X_i^-| = 0 \bigm| |X_{i+1}^-| = 0\big) \\ &= 1 - \frac{d_i}{d_{i+1}} = \frac{s_i - s_{i+1}}{1 - s_{i+1}} \sim s_i - s_{i+1}.\end{aligned}$$

From (4.29) we have $s_i < 2/i$ for all $i \ge 1$, so $\sum_i \mathbb{P}(S_i \neq \emptyset)$ converges uniformly as $\varepsilon \to 0$. Hence, for any $M(n) \to \infty$, the probability that any S_i, $i > M$, is non-empty tends to 0. For fixed i, the distribution of S_i converges as $\varepsilon \to 0$, in fact, to the distribution of the size of the ith generation of the exactly critical process $\mathfrak{X}_1$ given that the $(i+1)$st generation is empty. It follows that, in the branching process, $|S|$ converges in distribution to some random variable R not depending on ε. Modifying the arguments in Subsection 4.6, we find that when we replace the strong wedge condition by the medium wedge condition, in place of (4.47) we obtain the estimate

$$\mathbb{P}(F_q(x)) \sim \mathbb{P}(F_q) \sim 4\gamma_1 n^{-1}\lambda_\star^q \tag{5.1}$$

uniformly in $|q| \le M/\varepsilon$, where $\gamma_1 = \mathbb{E}(1/R)\gamma_0$, and γ_0 is the constant in Lemma 4.21.

Turning to the upper bound, after much work mostly involving ruling out pathological cases, we showed in Subsection 4.7 that for any function $M(n)$ tending to infinity, whp any vertex x that is part of a pair (x, y) at maximal distance satisfies the property $B^*(x)$, that $|t_{\omega/\varepsilon}(x) - t_0 - t_1| \le M/\varepsilon$, (see (4.63)) together with a certain unpleasant 'acceptability' condition $A^*(x)$. Moreover, Lemma 4.27 shows that the expected number of such vertices is bounded by some function of M. Thinking of M as constant for the moment, this expectation is bounded. Now given that a vertex has property B^*, it is likely that its relevant neighbourhood (up to $t_{\omega/\varepsilon}$) is a tree. (The expected number of edges within sets $\Gamma_t(x)$ is bounded by $\delta = \lambda n^{-1} t_{\omega/\varepsilon}\binom{\omega/\varepsilon}{2} = O(\omega^2(\log\Lambda)\varepsilon^{-3}n^{-1}) = O(\Lambda^{-2/3}\log\Lambda) = o(1)$; a similar bound holds for the expected number of 'redundant' edges between consecutive $\Gamma_t(x)$.) We had to consider the non-tree case, because δ may go to zero only slowly, but after reducing to vertices satisfying B^*, it is easy to check from the proof of Lemma 4.27 that the probability that $A^* \cap B^*$ holds and the neighbourhood is not a tree is $o(\mathbb{P}(A^* \cap B^*))$. It follows that (if M increases slowly enough), the expected number of vertices with $A^* \cap B^*$ holding and the neighbourhood not a tree is $o(1)$.

When considering tree neighbourhoods, acceptability becomes a much simpler condition, closely related to the weak wedge condition. So far we considered any vertex x in a pair (x, y) at maximal distance. Since we are only interested in the existence of a pair at a certain distance, we may restrict our attention to those x that are first in their tree T_x in our auxiliary random order. For vertices satisfying $A^* \cap B^*$, the conditional probability of this extra condition is asymptotically $\mathbb{E}(1/R)$, as above. Putting the pieces together, we find that whp the diameter is realized by some pair of vertices each of which satisfies a certain condition F'_q depending on its $t = t_0 + t_1 + q$ neighbourhood, where again $|q\varepsilon| \le M$. This condition is that $G_{\le t}(x)$ is a tree, and the event $A_t \cap B_t$ considered in Lemma 4.27 modified to the medium wedge condition holds. Also, modifying the proof of this lemma as indicated above, the probability that a vertex satisfies this condition is

$$\mathbb{P}(F'_q(x)) \sim \mathbb{E}(1/R)4\gamma_0\varepsilon^3\lambda_\star^{t-t_1} \sim 4\gamma_1\lambda_\star^q n^{-1}.$$

Now the precise details of $F_q(x)$ and $F'_q(x)$ are rather different. However, the definitions are such that $F_q(x)$ implies $F'_q(x)$. (Firstly, in defining $F_q(x)$ we insisted that $G_{\le t}(x)$ is a tree. Secondly, via the condition $D = D_1 \cap D'_1 \cap D_2$, we ensured that $|\Gamma_{t'}(x)| < \omega/\varepsilon$ for $0 \le t' \le t$. Thirdly, via A we ensured that for all t' up to $t_0 - r \ge t_0 - 2M/\varepsilon$, which is much larger than h, there is a unique particle in each generation t' with descendants in $\Gamma_t(x)$. Finally, we imposed the (there strong, but now medium) wedge condition on all the side branches starting up to time (at least) $t_0 - 2M/\varepsilon$. This implies the (modified) form of (h, t)-acceptability in F'_q.)

Since $\mathbb{P}(F'_q(x)) \sim \mathbb{P}(F_q(x))$, and the expected number of vertices with $F_q(x)$ is (for M fixed) $\Theta(1)$, it follows that for each q, whp *every* vertex with property $F'_q(x)$ also has $F_q(x)$. We shall essentially consider only a bounded number of values of q (again, a number that tends to infinity arbitrarily slowly), so this holds whp for all such values. Thus, whp, the diameter is equal to the maximum distance between vertices with property $F_q(x)$ for suitable q. This also applies if $M \to \infty$ slowly enough. We may thus forget about $F'_q(x)$.

Now the condition $F_q(x)$ says that the (medium) wedge condition holds, that $t(x) = t_{\omega/\varepsilon}(x) > t_0 + t_1 + q$, and that certain other technical conditions hold. We shall need to know a little more, namely roughly how large $t(x)$ is. From the remarks above, we may ignore x with $t(x) \geq t_0 + t_1 + M/\varepsilon$. For $-M^2 \leq i \leq M^2$, let $q_i = i/(M\varepsilon)$. Let us say that x is *of type* i if $F_{q_i}(x) \setminus F_{q_{i+1}}(x)$ holds; this corresponds roughly to the wedge condition plus $q_i < t(x) - t_0 - t_1 \leq q_{i+1}$. Let N_i be the number of type i vertices. With M constant, applying (5.1) twice shows that $\mathbb{E}\, N_i$ is asymptotically what it should be, and as usual this extends to $M \to \infty$ slowly enough, in which case

$$\mathbb{E}\, N_i \sim 4\gamma_1 e^{-q_i \varepsilon}/M,$$

since $\lambda_\star^q = (1 - \varepsilon + O(\varepsilon^2))^q \sim e^{-q\varepsilon}$ if $q\varepsilon$ does not grow too fast.

Let us say that x is *plausible* if it is of type i for some $-M^2 \leq i \leq M^2$. From the comments above, whp the diameter is realized by a pair of plausible vertices.

Now, the precise technical conditions in the definition of type i vertices are as in Subsection 4.6; as there, these allow us to calculate 2nd moments, and indeed rth moments for any fixed r. More precisely, given a sequence $\mathbf{i} = (i_1, \ldots, i_r)$, let us say that a sequence $(x_1, \ldots, x_r)$ of distinct vertices is an *r-tuple of type* $\mathbf{i}$ if each x_j is of type i_j. Such an r-tuple is *good* if the relevant trees witnessing this are disjoint, and *bad* otherwise. Arguing as in Subsection 4.6, the expected number of good r-tuples of type $\mathbf{i}$ is what it should be, namely $(1 + o(1)) \prod_{j=1}^r \mathbb{E}\, N_{i_j}$ (which is $\Theta(1)$ if M is fixed), and the expected number of bad r-tuples is $o(1)$. This shows that all fixed mixed moments of the sequence $(N_{-M^2}, \ldots, N_{M^2})$ converge to what we expect, and thus that (for M fixed) the sequence (N_i) converges in distribution to a sequence of independent Poisson random variables.

Turning to the diameter, let P be the number of unordered *pairs* (x, y) of plausible vertices with $d(x, y) \geq d = d_0 + c\varepsilon^{-1}$, where c is constant. We aim to understand $\mathbb{P}(P > 0)$ by evaluating the factorial moments $\mathbb{E}_k(P) = \mathbb{E}(P(P-1)\ldots(P-k+1))$. Now $\mathbb{E}_k(P)$ is the expected number of k-tuples of distinct pairs with the relevant property. It may be that several pairs involve the same vertex; in general we can write $\mathbb{E}_k(P)$ as a sum over integers $r \leq 2k$ and graphs H on $\{1, 2, \ldots, r\}$ with k edges of the expectation of the number of r-tuples of plausible vertices in which certain specified pairs are at distance at least d and the others are not. We evaluate this by summing over the types of the relevant vertices. Thus we must evaluate the expected number of r-tuples $(x_1, \ldots, x_r)$ of type $\mathbf{i}$ in which k specified pairs are at distance at least d and the others are not.

Since there are $o(1)$ bad r-tuples, we consider only good r-tuples. Finally, we test whether a particular sequence $(x_1, \ldots, x_r)$ has the required property by exploring the neighbourhoods of each x_j out to the relevant distance $(t_0 + t_1 + q_{i_j})$. By Lemma 4.8, the probability that the explorations are disjoint and each x_j is of the right type is 'what it should be', namely n^{-r} times the expected number of good r-tuples of type $\mathbf{i}$. Suppose this happens. Then we have not so far tested any edges outside these neighbourhoods.

Continuing to explore, the neighbourhoods grow at the expected rate whp. We explore $t_2/2 - O(1/\varepsilon)$ further steps, by which time the neighbourhoods have size $\Theta(\sqrt{\varepsilon n})$. (Recall that this is the size at which they typically meet.) By this time, there are very few (in expectation $O(1)$) vertices in two or more neighbourhoods, and whp none in three or more. It follows that the times at which different pairs of neighbourhoods meet are essentially independent, with distribution given by Lemma 4.5. This allows us to calculate $\mathbb{E}_k(P)$, and hence $\mathbb{P}(\mathrm{diam}(G(n, \lambda/n)) \geq d) \sim \mathbb{P}(P > 0)$.

Rather than give any further details, let us describe the limiting distribution we obtain. It should then be clear that all expectations being 'what they should be' corresponds to convergence to the corresponding values for this limiting distribution.

Let $\mathcal{P}$ be a Poisson process on $\mathbb{R}$ with density function $f(x) = 4\gamma_1 e^{-x}$. Note that $\int_{x' \geq x} f(x')\,\mathrm{d}x' = f(x) < \infty$ for any x, so with probability 1 we may list the points of $\mathcal{P}$ as $z_1, z_2, \ldots$ in decreasing order. For each $1 \leq i < j$, let T_{ij} be a random variable with $\mathbb{P}(T_{ij} > x) = \exp(-e^x)$, with these variables independent of each other and of $\mathcal{P}$. Finally, let $D = \sup\{z_i + z_j + T_{ij}\}$. It is not hard to check that with probability 1 D is finite, and the supremum is attained. Indeed, as $M \to \infty$, the probability that it is attained by some i, j with $z_i, z_j \geq -M$ tends to 1.

Theorem 5.1. *Let $\varepsilon = \varepsilon(n) > 0$ satisfy $\varepsilon \to 0$ and $\varepsilon^3 n \to \infty$, and let $\lambda = 1 + \varepsilon$. For any constant c we have*

$$\mathbb{P}\left(\mathrm{diam}(G(n, \lambda/n)) \geq \frac{\log(\varepsilon^3 n)}{\log \lambda} + 2\frac{\log(\varepsilon^3 n)}{\log(1/\lambda_\star)} + c/\varepsilon\right) \to \mathbb{P}(D \geq c)$$

as $n \to \infty$. □

In other words, the $O_{\mathrm{p}}(1/\varepsilon)$ correction term in (1.6) converges in distribution to D (after multiplication by ε).

We have proved Theorem 5.1 in outline above. There are a few further technical details (such as checking that the relevant sequences of moments do not grow too fast, so convergence of all fixed moments gives convergence in distribution), but we shall not describe these any further.

The description of the random variable D is somewhat complicated; however, it seems rather unlikely that this random variable will have a simpler description. Given this description, the branching process approach taken here seems with hindsight very natural: the description of D more or less forces us to consider the (exponentially distributed) times that the vertices take for their neighbourhoods to reach certain very large sizes, and then the time they take to meet after this.

Finally, let us comment very briefly on the case $p = \lambda/n$, λ constant. It is not that the proof is any harder in this case (it is much easier), but the result is much harder to describe. Again we consider vertices satisfying the medium wedge condition (which now has probability bounded away from 0), and, taking $\omega = (\log n)^6$, say, we study the distribution of $t_\omega(x)$ for such x, in the range where $\mathbb{P}(t_\omega(x) \geq t_0)$ is of order $1/n$. From Lemma 2.1 it is very easy to check that when $t_\omega(x)$ is very large, this is almost always because for many generations there is only one neighbour whose descendants do not die quickly, and we easily find asymptotic independence of the event $\{t_\omega(x) > t_0\}$ and the wedge condition.

Approximating by a branching process, it is easy to prove an equivalent of Theorem 4.19, showing that the distribution of $t_\omega(x)$ may be described (as in the $\lambda \to 1$ case) by the tail of $Y = Y_\lambda$ near 0. But now the first complication appears: this random variable no longer has a nice power-law tail, but asymptotically follows a power law multiplied by a function that oscillates periodically within a constant factor. Also, when we explore neighbourhoods and reach size ω, the current neighbourhood may have any size between ω and $\lambda\omega$; this constant factor affects the probability of joining up with another neighbourhood within a certain time. In the end it turns out that the distribution depends on the fractional parts of both $\log n/\log\lambda$ and $\log n/\log\lambda_\star$, as indeed it must from the form of (1.4). We omit the details, as a precise statement of the result would be rather lengthy.

Acknowledgements. This research started during the program ‘Random Graphs and Large-Scale Real-World Networks’ at the Institute for Mathematical Sciences, National University of Singapore in summer 2006; the authors are grateful to the Institute for its support. The authors would like to thank the anonymous referee for a careful reading of the paper and many helpful suggestions concerning the presentation.

References

[1] L. Addario-Berry, N. Broutin and C. Goldschmidt, The continuum limit of critical random graphs, preprint available from arXiv:0903.4730.

[2] R. Albert, H. Jeong and A.-L. Barabási, Diameter of the world-wide web, *Nature* **401** (1999), 130–131.

[3] K.B. Athreya and P.E. Ney, *Branching processes*, Springer, Berlin, 1972.

[4] A.-L. Barabási, R. Albert and H. Jeong, Scale-free characteristics of random networks: the topology of the world-wide web, *Physica A* **281** (2000), 69–77.

[5] J. van den Berg and H. Kesten, Inequalities with applications to percolation and reliability, *J. Appl. Probab.* **22** (1985), 556–569.

[6] B. Bollobás, The diameter of random graphs, *Trans. Amer. Math. Soc.* **267** (1981), 41–52.

[7] B. Bollobás, The evolution of random graphs. *Trans. Amer. Math. Soc.* **286** (1984), 257–274.

[8] B. Bollobás, *Random Graphs*, Academic Press Inc., London (1985).

[9] B. Bollobás and F.R.K. Chung, The diameter of a cycle plus a random matching, *SIAM J. Discrete Math.* **1** (1988), 328–333.

[10] B. Bollobás and W. Fernandez de la Vega, The diameter of random regular graphs, *Combinatorica* **2** (1982), 125–134.

[11] B. Bollobás, S. Janson and O. Riordan, The phase transition in inhomogeneous random graphs, *Random Structures and Algorithms* **31** (2007), 3–122.

[12] B. Bollobás and O. Riordan, The diameter of a scale-free random graph, *Combinatorica* **24** (2004), 5–34.

[13] B. Bollobás and O. Riordan, Random graphs and branching processes, in *Handbook of large-scale random networks*, *Bolyai Soc. Math. Stud* **18**, B. Bollobás, R. Kozma and D. Miklós eds (2009), pp. 15–115.

[14] Ju.D. Burtin, Asymptotic estimates of the diameter and the independence and domination numbers of a random graph, *Dokl. Akad. Nauk SSSR* **209** (1973), 765–768, translated in *Soviet Math. Dokl.* **14** (1973), 497–501.

[15] Ju.D. Burtin, Extremal metric characteristics of a random graph. I, *Teor. Verojatnost. i Primenen.* **19** (1974), 740–754.

[16] F. Chung and L. Lu, The diameter of sparse random graphs, *Adv. in Appl. Math.* **26** (2001), 257–279.

[17] F. Chung and L. Lu, The average distances in random graphs with given expected degrees, *Proc. Natl. Acad. Sci. USA* **99** (2002), 15879–15882 (electronic).

[18] F. Chung and L. Lu, The average distance in a random graph with given expected degrees, *Internet Math.* **1** (2003), 91–113.

[19] J. Ding, J.H. Kim, E. Lubetzky and Y. Peres, Anatomy of a young giant component in the random graph, preprint available from arXiv:0906.1839.

[20] J. Ding, J.H. Kim, E. Lubetzky and Y. Peres, Diameters in supercritical random graphs via first passage percolation, preprint available from arXiv:0906.1840.

[21] P. Erdős and A. Rényi, On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 17–61.

[22] H. van den Esker, R. van der Hofstad, G. Hooghiemstra and D. Znamenski, Distances in random graphs with infinite mean degrees, *Extremes* **8** (2006), 111–141 (2006).

[23] D. Fernholz and V. Ramachandran, The diameter of sparse random graphs, *Random Structures and Algorithms* **31** (2007), 482–516.

[24] T.E. Harris, Branching processes, *Ann. Math. Statistics* **19** (1948), 474–494.

[25] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem, Distances in random graphs with finite variance degrees, *Random Structures Algorithms* **27** (2005), 76–123.

[26] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski, Distances in random graphs with finite mean and infinite variance degrees, *Electron. J. Probab.* **12** (2007), 703–766 (electronic).

[27] S. Janson, On concentration of probability, in *Contemporary combinatorics*, Bolyai Soc. Math. Stud. **10**, János Bolyai Math. Soc. (2002), pp. 289–301.

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

[29] A.N. Kolmogorov, Zur Lösung einer biologischen Aufgabe, *Comm. Math. Mech. Chebyshev Univ. Tomsk* **2** (1938) 1–6.

[30] T. Łuczak, Component behavior near the critical point of the random graph process, *Random Structures Algorithms* **1** (1990), 287–310.

[31] T. Łuczak, Cycles in a random graph near the critical point, *Random Structures and Algorithms* **2** (1991), 421–440.

[32] T. Łuczak, Random trees and random graphs, *Random Struct. Alg.* **13** (1998), 485–500.

[33] T. Łuczak and T.G. Seierstad, The diameter behavior in the random graph process, Mittag–Leffler preprint Report no 5, 2008/2009 spring.

[34] A. Nachmias and Y. Peres, Critical random graphs: diameter and mixing time, *Annals of Probability* **36** (2008), 1267–1286.

[35] M.E.J. Newman, S.H. Strogatz and D.J. Watts, Random graphs with arbitrary degree distribution and their applications, *Physical Review E* **64** (2001), 026118.

[36] B. Pittel and C. Wormald, Counting connected graphs inside-out, *J. Combinatorial Theory B* **93** (2005), 127–172.

[37] D. Reimer, Proof of the van den Berg-Kesten conjecture, *Combin. Probab. Comput.* **9** (2000), 27–32.

[38] V. Sanwalani and N. Wormald, The diameter of random regular graphs (in preparation).

[39] N.C. Wormald, Models of random regular graphs, *Surveys in Combinatorics, 1999*, London Mathematical Society Lecture Note Series **267** (J.D. Lamb and D.A. Preece, eds) Cambridge University Press, Cambridge, pp. 239–298, 1999.