Component structure of the configuration model: Barely supercritical case

Remco van der Hofstad1 | Svante Janson2 | Malwina Luczak3

1Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands
2Department of Mathematics, Uppsala University, Uppsala, Sweden
3School of Mathematics and Statistics, University of Melbourne, Parkville, Victoria, Australia

Correspondence
Svante Janson, Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden.
Email: svante.janson@math.uu.se

Funding information
This research was supported by the EPSRC; EP/K032208/1. Netherlands Organisation for Scientific Research (NWO); 639.033.806; 024.002.003. Knut and Alice Wallenberg Foundation. Simons Foundation. EPSRC Leadership Fellowship; EP/J004022/2. ARC Future Fellowship; FT170100409.

Abstract
We study near-critical behavior in the configuration model. Let $D_n$ be the degree of a random vertex and $\nu_n = \mathbb{E}[D_n(D_n - 1)]/\mathbb{E}[D_n]$; we consider the barely supercritical regime, where $\nu_n \to 1$ as $n \to \infty$, but $\nu_n - 1 \gg n^{-1/3}(\mathbb{E}[D_n^3])^{2/3}$. Let $D_n^*$ denote the size-biased version of $D_n$. We prove that there is a unique giant component of size $n\rho_n \mathbb{E} D_n(1 + o(1))$, where $\rho_n$ denotes the survival probability of a branching process with offspring distribution $D_n^* - 1$. This extends earlier results of Janson and Luczak, as well as those of Janson, Luczak, Win-dridge, and House, to the case where the third moment of $D_n$ is unbounded. We further study the size of the largest component in the critical regime, where $\nu_n - 1 = O(n^{-1/3}(\mathbb{E}[D_n^3])^{2/3})$, extending and complementing results of Hatami and Molloy.

KEYWORDS
percolation, phase transition, random graphs, scaling window

1 | INTRODUCTION

In recent years, the critical and near-critical behavior of random graphs has received considerable attention. Here we study random graphs with given vertex degrees. (See Section 2.1 for precise definitions and assumptions.) In a random graph with given degrees on $n$ vertices, we let $D_n$ denote the degree of a random vertex; we consider asymptotics as $n \to \infty$. The fundamental theorem by Molloy and Reed [46] (see also [10, 36, 40, 43, 47], and Section 2) says that, under suitable technical assumptions, there exists w.h.p. (meaning “with high probability,” i.e., with probability tending to 1 as $n \to \infty$) a giant component of size $\Theta(n)$ if and only if $\lim_{n \to \infty} \mathbb{E} D_n(D_n - 2) > 0$.

The purpose of the present paper is to study near-critical behavior in greater detail; we assume $\mathbb{E} D_n(D_n - 2) \to 0$ so we know that the order $v(C_1)$ of the largest component is $o_p(n)$, and we want to find more precise asymptotics of $v(C_1)$. 

Hatami and Molloy [24] identified the critical window; they showed that (under weak technical conditions) if \( E D_n(D_n - 2) = O(n^{-1/3}(E D_n^3)^{2/3}) \), then \( v(C_1) \) is of the order \( n^{2/3}(E D_n^3)^{-1/3} \), while \( v(C_1) \) is larger if \( E D_n(D_n - 2) \gg n^{-1/3}(E D_n^3)^{2/3} \), and smaller if \( E D_n(D_n - 2) < 0 \) with \( |E D_n(D_n - 2)| \gg n^{-1/3}(E D_n^3)^{2/3} \). (See also Remark 2.16 for related work identifying the scaling limits of clusters in the critical window.) This parallels the well-known critical behavior of the random graph \( G(n, p) \) with \( p = (1 + \varepsilon_n)/n \), or \( G(n, M) \) with \( M = (1 + \varepsilon_n)n/2 \), where it was shown by Bollobás [7] and Łuczak [44] that the critical window is characterized by \( \varepsilon_n = O(n^{-1/3}) \); see also 
[8, 39].

Here we are mainly concerned with the barely supercritical regime, where \( E D_n(D_n - 2) \to 0 \), with \( E D_n(D_n - 2) > 0 \) and outside the critical window just defined. We find (under weak technical conditions) precise asymptotics of \( v(C_1) \), up to a factor \( 1 + o_p(1) \), in this regime. In the case when the degree distribution \( D_n \) has a bounded \( (4 + \varepsilon) \)th moment, these asymptotics were found by Janson and Łuczak [36]; this result was extended to the case when the third power \( D_n^3 \) is uniformly integrable by Janson, Łuczak, and Windridge [38]. In this paper, we only assume that the second moment \( E D_n^2 \) exists and is uniformly bounded. Our study reveals that there is a kind of phase transition. Roughly speaking, as long as the asymptotic degree distribution has a finite third moment (to be precise, as long as \( D_n^3 \) is uniformly integrable, the case studied in [36, 38]), the size of the largest component is proportional to \( n E(D_n(D_n - 2)) \). However, when the degree distribution has heavier tails, then the largest component is smaller; typically (but not always) of the order \( n E(D_n(D_n - 2))/E D_n^3 \). Precise results are given in Theorems 2.6-2.9, where Theorem 2.8 corresponds to the important example when the third moment of the degree distribution converges. Also, Example 2.15 discusses power-law degree sequences with possibly unbounded third moment of the degree distribution. (The same difference between the cases \( E D_n^3 = O(1) \) and \( E D_n^3 \to \infty \) is also evident in the result on the critical window by Hatami and Molloy [24] cited above.)

As said above, our results (Theorem 2.6 in particular) show that in the barely supercritical phase, the size of the largest component is concentrated within a factor \( 1 + o_p(1) \), that is, normalized by dividing by a suitable constant, the size converges in probability to 1. As a complement, we also show (Theorem 2.12) that this is not true in the critical window identified by Hatami and Molloy [24], and further investigated in [18, 19, 41, 51]. Inside the critical window, the size after normalization will converge in distribution, at least along subsequences, but the limit will not be constant; in fact any such limit will be unbounded. Again, this is precisely as in the well-known case of \( G(n, p) \), see [1, 45], so this provides another reason to regard the window defined above as the critical window, at least on the supercritical side. (We conjecture that the size of the largest component is concentrated also in the subcritical case, but, as far as we know, this has not yet been proved.)

It is well known that the process of exploration of the component containing a given vertex can be approximated by a Galton-Watson branching process; this gives, for example, a heuristic argument for the condition \( \lim_{n \to \infty} E D_n(D_n - 2) > 0 \) above. (See further Remark 2.5.) Indeed, in our main theorem (Theorem 2.6), we express the size of the largest component in terms of the survival probability of the approximating Galton-Watson process. In our case, with \( E D_n(D_n - 2) \to 0 \), we have to consider one Galton-Watson process for each \( n \), so the question of asymptotics of the survival probability of an asymptotically critical sequence of branching processes arises. This was studied by for example, [3, 27]; we give some further general results (needed to prove our results for random graphs) in Section 3.

Our proofs, however, do not use the branching process approximation directly; instead, they are based on extending the method of [36], where the exploration process is considered one vertex at a time, yielding a kind of random walk with drift (closely related to the branching process), which is then analyzed. Molloy and Reed [46, 47] and Hatami and Molloy [24] use similar methods, but there are several differences; for example, we use a continuous-time version of the exploration process, which gives us additional independence, and we use a different method to obtain bounds for the random fluctuations.
2 | MODEL, ASSUMPTIONS, AND MAIN RESULTS

2.1 | The configuration model

Given a positive integer \( n \) and a degree sequence \((d_1, d_2, \ldots, d_n)\), we let \( G(n, (d_i)_{i \in [n]}) \) be a simple graph (i.e., without loops or multiple edges) with the set \([n] = \{1, \ldots, n\}\) of vertices, chosen uniformly at random subject to vertex \( i \) having degree \( d_i \), for \( i \in [n] \). We tacitly assume that there is any such graph at all, so, for example, \( \sum_{i \in [n]} d_i \) must be even.

We follow the standard path of studying \( G(n, (d_i)_{i \in [n]}) \) using the configuration model, defined as follows, see for example [8, 25]. Given a degree sequence \((d_i)_{i \in [n]}\) with \( \sum_{i \in [n]} d_i \) even, we start with \( d_j \) free half-edges adjacent to vertex \( j \), for \( j \in [n] \). The random multigraph \( G^*(n, (d_i)_n) \) is constructed by successively pairing, uniformly at random, free half-edges into edges, until no free half-edges remain. (In other words, we create a uniformly random matching of the half-edges.) Loops and multiple edges may occur in \( G^*(n, (d_i)_n) \), but we can obtain \( G(n, (d_i)_{i \in [n]}) \) by conditioning \( G^*(n, (d_i)_{i \in [n]}) \) on being simple (i.e., without loops or multiple edges). Moreover, our condition (A2) implies that the probability of obtaining a simple graph is bounded away from 0 as \( n \to \infty \); see [2, 29, 33].

We assume that we are given such a degree sequence \((d_i)_{i \in [n]}\) for each \( n \) (at least in a subsequence), and we consider asymptotics as \( n \to \infty \). The degrees \( d_i = d_i^n \) may depend on \( n \), but for simplicity we do not show this in the notation.

2.2 | Basic assumptions and notation

All unspecified limits are as \( n \to \infty \). We use standard notation for asymptotics. In particular, \( a_n \asymp b_n \), where \( a_n \) and \( b_n \) are sequences of positive numbers, means that \( a_n/b_n \) is bounded above and below by positive constants; equivalently, \( a_n = O(b_n) \) and \( b_n = O(a_n) \). In contrast, \( a_n \sim b_n \) means the stronger \( a_n/b_n \to 1 \). Furthermore, \( a_n \gg b_n \) means \( a_n/b_n \to \infty \). Also, given two real numbers \( x, y \), \( x \wedge y \) will denote \( \min \{x, y\} \), and \( x \vee y \) will denote \( \max \{x, y\} \).

For random variables \( X_n \), and positive numbers \( a_n \), \( X_n = o_p(a_n) \) means \( X_n/a_n \overset{p}{\to} 0 \), that is, \( \mathbb{P}(|X_n| > \epsilon a_n) \to 0 \) for every \( \epsilon > 0 \). Also, \( X_n = O_p(a_n) \) means that \( X_n/a_n \) is bounded in probability, that is, for every \( \epsilon > 0 \) there exists \( C < \infty \) such that \( \mathbb{P}(|X_n| > Ca_n) < \epsilon \) for all \( n \) (or, equivalently, for all large \( n \)).

We let \( \Delta_n := \max_{i \in [n]} d_i \) denote the maximum degree in \( G(n, (d_i)_{i \in [n]}) \) and \( G^*(n, (d_i)_{i \in [n]}) \).

For \( k \in \mathbb{Z} \), we denote by

\[
n_k := \# \{ i : d_i = k \},
\]  
(2.1)

the number of vertices of degree \( k \), so that \( n = \sum_{k=1}^{\infty} n_k \). Furthermore, let

\[
\ell_n := \sum_{i \in [n]} d_i = \sum_{k=1}^{\infty} kn_k,
\]  
(2.2)

be the total number of half-edges; thus the number of edges is \( \ell_n/2 \).

Let \( D_n \) be the degree of a randomly chosen vertex in \( G(n, (d_i)_{i \in [n]}) \) or \( G^*(n, (d_i)_{i \in [n]}) \); the distribution of \( D_n \) is given by

\[
\mathbb{P}(D_n = k) = n_k/n. 
\]  
(2.3)
Let
\[ \mu_n := \mathbb{E}D_n = \sum_{k=1}^{\infty} k n_k / n = \ell_n / n, \tag{2.4} \]
\[ \nu_n := \frac{\mathbb{E}D_n(D_n - 1)}{\mathbb{E}D_n} = \frac{\sum_{k=1}^{\infty} k(k-1)n_k}{\sum_{k=1}^{\infty} kn_k} = \frac{\sum_{k=1}^{\infty} k(k-1)n_k}{\ell_n}. \tag{2.5} \]

Thus \( \mu_n \) is the average degree; \( \nu_n \) can be interpreted as the expected number of new half-edges found when the endpoint of a random half-edge is explored, see (2.16) and Remark 2.5.

As stated in Section 1, we will study near-critical behavior; we assume \( \nu_n \to 1 \) and, for the most part, also that \( \nu_n > 1 \) (and not too small); this is thus a subcase of the critical case so \( v(C_1) = o_p(n) \).

We define
\[ \epsilon_n := \nu_n - 1 = \frac{\mathbb{E}D_n(D_n - 2)}{\mathbb{E}D_n}. \tag{2.6} \]

Our basic assumptions are as follows (See also the remarks below, and additional conditions in the theorems):

(A1) \( D_n \), the degree of a randomly chosen vertex, converges in distribution to a random variable \( D \) with a finite and positive mean \( \mu := \mathbb{E}D \). In other words, there exists a probability distribution \( (p_k)_{k=0}^{\infty} \) such that
\[ \frac{n_k}{n} \to p_k, \quad k \geq 0, \tag{2.7} \]
and \( \mu = \sum_{k=0}^{\infty} kp_k \in (0, \infty) \). (Thus \( p_k = \mathbb{P}(D = k) \).)

(A2) The second moment \( \mathbb{E}D_n^2 \) is uniformly bounded: \( \mathbb{E}D_n^2 = O(1) \).

(A3) We have \( \mathbb{P}(D \notin \{0, 2\}) > 0 \). Equivalently, \( p_0 + p_2 < 1 \).

(A4) \( \nu_n \to 1 \). Equivalently, see (2.6),
\[ \epsilon_n \to 0. \tag{2.8} \]

Assuming (A1), this is also equivalent to
\[ \mathbb{E}D_n(D_n - 2) \to 0. \tag{2.9} \]

Remark 2.1 The assumption (A1) that \( D_n \) converges in distribution is mainly for convenience. By (A2), the sequence \( D_n \) is always tight, so every subsequence has a further subsequence that converges in distribution to some \( D \); moreover \( \mathbb{E}D < \infty \) follows from (A2) and \( \mathbb{E}D > 0 \) follows from (A3), provided the latter is reformulated as \( \liminf_{n \to \infty} \mathbb{P}(D_n \notin \{0, 2\}) > 0 \). It follows, using standard subsequence arguments, that results such as Theorem 2.6 that do not use \( D \) (explicitly or implicitly) in the statement hold also without (A1).

Remark 2.2 Condition (A2) implies uniform integrability of \( D_n \) and thus, together with (A1),
\[ \mu_n \to \mu. \tag{2.10} \]
Furthermore, it is easy to see that, assuming (A1) and (A2) is equivalent to $\nu_n = O(1)$. In particular, (A2) is implied by (A4); however, we list (A2) separately for emphasis and for easier comparison with conditions in other papers.

By Fatou’s lemma, (A2) also implies $\mathbb{E} D^2 < \infty$.

**Remark 2.3**  Condition (A2) is weaker than the condition

(A2’) $D^2_n$ are uniformly integrable.

As is well known, (A2′) is, assuming (A1), equivalent to $\mathbb{E} D_n^2 \rightarrow \mathbb{E} D^2 < \infty$, and thus also to $\mathbb{E} D^2 < \infty$ and

$$\nu_n \rightarrow \nu := \frac{\mathbb{E} D(D-1)}{\mathbb{E} D}. \quad (2.11)$$

In this case, (A4) is thus equivalent to $\nu = 1$, or, equivalently, $\mathbb{E} D(D - 2) = 0$, or $\mathbb{E} D^2 = 2\mu$.

On the other hand, if (A1), (A2) and (A4) are satisfied but (A2′) is not, then (by Fatou’s lemma) $\mathbb{E} D^2 < 2\mu$, $\mathbb{E} D(D - 2) < 0$ and $\nu < 1$.

We will not need (A2′) in the present paper, except when explicitly stated; it is satisfied in most examples.

**Remark 2.4**  Condition (A3) rules out the degenerate case when $D \in \{0, 2\}$ a.s.; for examples of exceptional behavior in this case, see [36, Remark 2.7].

Since $\mathbb{E} D(D - 2) \leq 0$, see Remark 2.3, (A3) is equivalent to $\mathbb{P}(D = 1) > 0$. Furthermore, if $D^2_n$ are uniformly integrable, so $\mathbb{E} D(D - 2) = 0$, see Remark 2.3, then (A3) is also equivalent to $\mathbb{P}(D > 2) > 0$.

### 2.3 The size-biased distribution

Let $D^*_n$ denote the size-biased distribution of $D_n$, i.e.,

$$P(D^*_n = k) = \frac{k}{\mathbb{E}[D_n]} P(D_n = k), \quad (2.12)$$

and let $\tilde{D}_n := D^*_n - 1$, that is,

$$P(\tilde{D}_n = k - 1) = P(D^*_n = k) = \frac{k P(D_n = k)}{\mathbb{E} D_n} = \frac{k n k}{n \mu}, \quad k \geq 1. \quad (2.13)$$

For any nonnegative function $f$,

$$\mathbb{E} f(D^*_n) = \frac{\mathbb{E} D_n f(D_n)}{\mathbb{E} D_n}; \quad (2.14)$$

and thus

$$\mathbb{E} f(\tilde{D}_n) = \frac{\mathbb{E} D_n f(D_n - 1)}{\mathbb{E} D_n}; \quad (2.15)$$

in particular

$$\mathbb{E} \tilde{D}_n = \mathbb{E}(D^*_n - 1) = \frac{\mathbb{E}(D_n(D_n - 1))}{\mathbb{E} D_n} = \nu_n = 1 + \epsilon_n. \quad (2.16)$$
Similarly, let $D^*$ have the size-biased distribution of $D$, and let $\tilde{D} := D^* - 1$. Thus $\mathbb{E}\tilde{D} = v = 1$. Since $D_n \xrightarrow{d} D$ by (A1) and $\mathbb{E}D_n \rightarrow \mathbb{E}D$ by (2.10), it follows that $D^*_n \xrightarrow{d} D^*$ and $\tilde{D}_n \xrightarrow{d} \tilde{D}$.

Note that (A3) implies that (and, given (A1), is equivalent to)

$$
\lim \inf_{n \to \infty} \mathbb{P}(\tilde{D}_n \neq 1) = \mathbb{P}(\tilde{D} \neq 1) = \mathbb{P}(D \neq 2) > 0. \tag{2.17}
$$

Let $\rho_n$ be the survival probability of a Galton-Watson process with offspring distribution $\tilde{D}_n$, starting from one individual. By (2.16) and basic branching process theory, $\rho_n > 0 \iff \epsilon_n > 0$, and, in this case $\rho_n$ is the unique solution in $(0, 1)$ to

$$
1 - \rho_n = \mathbb{E}(1 - \rho_n)^{\tilde{D}_n} = \sum_{k=1}^{\infty} \frac{kn_k}{n \mu_n} (1 - \rho_n)^{-1}. \tag{2.18}
$$

We study the asymptotics of $\rho_n$ in Section 3.

**Remark 2.5** We can interpret $D^*_n$ as the degree of a vertex chosen randomly by choosing a uniformly random half-edge, and $\tilde{D}_n$ as the number of additional half-edges at that vertex. Consequently, the initial stages of the exploration of a component of $G(n, (d_i)_{i \in [n]}), starting from a random vertex, can be approximated by a Galton-Watson process with offspring distribution $\tilde{D}_n$, except that the first generation has distribution $D_n$. The survival probability $\rho_n$ is thus closely connected to the probability that this modified Galton-Watson process is infinite, which approximates the probability that the chosen vertex lies in a large component. (In the supercritical case, this is asymptotically the same as the probability of the chosen vertex lying in the largest component.) To be precise, the modified Galton-Watson process has survival probability $\mathbb{E}(1 - (1 - \rho_n)^{\tilde{D}_n}) \sim \mu_n \rho_n$, which agrees with the factor $\mu_n \rho_n$ in Theorem 2.6, giving the proportion of vertices in the largest component.

### 2.4 Main results

Our results in this section hold for both the random simple graph $G_n := G(n, (d_i)_{i \in [n]})$ and the random multigraph $G^*_n := G^*(n, (d_i)_{i \in [n]})$. We first prove our theorems for $G^*_n$; they then hold for $G_n$, as is standard, by conditioning on $G^*_n$ being simple. To be precise, (A2) implies that $\lim \inf_{n \to \infty} \mathbb{P}(G^*_n$ is simple) $> 0$, see [29, 33], and thus the results below (which all say that certain events have small probabilities) transfer immediately from $G^*_n$ to $G_n$, except Theorem 2.12(ii), which is of a different kind and requires a special argument (given in Section 6.3).

In order to state our results, choose either $G_n$ or $G^*_n$; let $C_1$ denote the largest connected component, and let $C_2$ denote the second largest component. (For definiteness, we choose the component at random if there is a tie, and we define $C_2 := \emptyset$ if there is only one component.)

For a component $C$, we write $v(C)$ and $e(C)$ to denote the number of vertices and edges in $C$, respectively. Our main theorem is the following precise and general result concerning the supercritical case:

**Theorem 2.6** Suppose that (A1)–(A4) are satisfied, in particular $\epsilon_n = o(1)$. Suppose also that $\epsilon_n \gg n^{-1/3}(\mathbb{E}D^3)^{2/3}$. Then

$$
v(C_1) = \mu_n \rho_n n(1 + o_p(1)), \tag{2.19}
v(C_2) = o_p(\rho_n n). \tag{2.20}
$$
Furthermore, \( e(C_1) = (1 + o_p(1))v(C_1) = \mu_n \rho_n n(1 + o_p(1)) \) and \( e(C_2) = o_p(\rho_n n) \).

**Remark 2.7** Let \( v_k(C_1) \) denote the number of vertices of degree \( k \) in \( C_1 \). It can be seen from our proof of Theorem 2.6 that \( v_k(C_1) = \mu_n \rho_n \mathbb{P}(D_n^k = k)n(1 + o_p(1)) \).

In particular, Theorem 2.6 leads to the following special cases.

Define, recalling Remark 2.4

\[
\kappa := \mathbb{E}[\tilde{D}(\tilde{D} - 1)] = \frac{\mathbb{E}[D(D - 1)(D - 2)]}{\mathbb{E}[D]} \geq 0. \tag{2.21}
\]

Note that \( \kappa = \infty \) if and only if \( \mathbb{E}[D^3] = \infty \). Furthermore, if \( D_n^2 \) are uniformly integrable (ie, (A2') holds), then \( \mathbb{P}(D > 2) > 0 \) by Remark 2.4, and thus \( \kappa > 0 \). In this case, we also have \( \mathbb{E}[D(D - 2)] = 0 \), see Remark 2.3, and thus we also have the alternative formula

\[
\kappa = \frac{\mathbb{E}[D^3 - 3 \mathbb{E}[D^2] + 2 \mathbb{E}[D]]}{\mathbb{E}[D]} = \frac{\mathbb{E}[D^3] - 3 \mathbb{E}[D(D - 2)] - 4 \mathbb{E}[D]}{\mu} - 4. \tag{2.22}
\]

The next three theorems are easy consequences of Theorem 2.6, under our assumptions.

**Theorem 2.8** Suppose that (A1)--(A4) are satisfied, and that \( D_n^3 \) is uniformly integrable. (Thus, \( \mathbb{E}[D_n^3] \to \mathbb{E}[D^3] < \infty \).) Suppose further that \( \varepsilon_n n^{1/3} \to \infty \). Then

\[
v(C_1) = \frac{2\mu}{\kappa} \varepsilon_n n(1 + o_p(1)) = \frac{2n \mathbb{E}(D_n(D_n - 2)) \varepsilon_n}{\kappa}(1 + o_p(1)), \tag{2.23}
\]

\[
v(C_2) = o_p(\varepsilon_n n), \tag{2.24}
\]

where \( \kappa \in (0, \infty) \) is given by (2.21). Furthermore, \( e(C_1) = (1 + o_p(1))v(C_1) \) and \( e(C_2) = o_p(\varepsilon_n n) \).

**Theorem 2.9** Suppose that (A1)--(A4) are satisfied, and that \( \mathbb{E}[D^3] = \infty \). (Thus \( \mathbb{E}[D_n^3] \to \infty \).) Suppose further that \( \varepsilon_n \gg n^{-1/3}(\mathbb{E}[D^3])^{2/3} \). Then

\[
v(C_1) = o_p(\varepsilon_n n). \tag{2.25}
\]

Furthermore, \( e(C_1) = (1 + o_p(1))v(C_1) = o_p(\varepsilon_n n) \).

The results in Theorems 2.8 and 2.9 are more or less best possible of this type: in intermediate cases, where \( \mathbb{E}[D^3] < \infty \) but \( \limsup \mathbb{E}[D_n^3] > \mathbb{E}[D^3] \), neither (2.23) nor (2.25) holds in general, see Remark 3.3. To be precise, it follows from Examples 3.7 and 3.9 that \( \mathbb{E}[D_n^3] = O(1) \) is neither necessary nor sufficient for (2.23). Similarly, it follows from Examples 3.7 and 3.8 that \( \mathbb{E}[D_n^3] \to \infty \) is not sufficient for (2.25) and \( \mathbb{E}[D^3] = \infty \) is not necessary for (2.25). In such intermediate cases, partial answers are given by the following inequalities. Define, in analogy with (2.21),

\[
\kappa_n := \mathbb{E}[\tilde{D}_n(\tilde{D}_n - 1)] = \frac{\mathbb{E}[D_n(D_n - 1)(D_n - 2)]}{\mathbb{E}[D_n]} \tag{2.26}
\]
Note that, since $\varepsilon_n > 0$, by (2.6) we have $\mathbb{E}[D_n(D_n - 2)] > 0$, which in turn implies $\kappa_n > 0$. Furthermore, by Fatou’s lemma and (2.9),

$$\liminf_{n \to \infty} \kappa_n = \liminf_{n \to \infty} \frac{\mathbb{E}[D_n(D_n - 2)^2]}{\mathbb{E}D} + \lim_{n \to \infty} \frac{\mathbb{E}[D_n(D_n - 2)]}{\mathbb{E}D} > 0.$$  

(2.27)

Thus $\kappa_n$ is bounded away from 0, and it follows that

$$\kappa_n \asymp \mathbb{E}D_n^3.$$  

(2.28)

**Theorem 2.10** Suppose that (A1)–(A4) are satisfied. Suppose also that $\varepsilon_n \gg n^{-1/3}(\mathbb{E}D_n^3)^{2/3}$.

(i) Then

$$\nu(C_1) \geq 2\mu_n\varepsilon_n n(1 + o_p(1)).$$  

(2.29)

(ii) If $\mathbb{E}D_n^3 = O(1)$, then there exists constant $c, C > 0$ such that w.h.p.

$$c\varepsilon_n n \leq \nu(C_1) \leq C\varepsilon_n n.$$  

(2.30)

(iii) If $\varepsilon_n \Delta_n = o(\mathbb{E}D_n^3)$, then there exists constants $c', C, C' > 0$ such that w.h.p.

$$c' \frac{\varepsilon_n n}{\mathbb{E}D_n^3} \leq \frac{\nu(C_1)}{\kappa_n} \leq C' \frac{\varepsilon_n n}{\mathbb{E}D_n^3} \leq C' \frac{\nu(C_1)}{\kappa_n}.$$  

(2.31)

The lower bounds in (iii) are clearly less precise than the more general (2.29), but are given as companions to the upper bounds. A weaker and less precise version of the lower bound (2.29) was given by Hatami and Molloy [24, Theorem 1.3].

**Remark 2.11** We see from Theorems 2.8–2.10 that in the barely supercritical regime, for a given sequence $\varepsilon_n$, the giant component is smaller in cases where $\mathbb{E}D^3 = \infty$ than in cases where $\mathbb{E}D_n^3$ is bounded. (In both cases, the size of the giant component is by Theorem 2.6 roughly $n\rho_n$.) The barely supercritical behavior of the largest connected component when $\mathbb{E}[D_n^3] = O(1)$ is similar to that in the Erdős-Rényi random graph.

The condition $\varepsilon_n \gg n^{-1/3}(\mathbb{E}D_n^3)^{2/3}$ in the theorems above is best possible and characterizes supercritical behavior in the sense that, if $\varepsilon_n$ is smaller, then, unlike (2.19), $\nu(C_1)$ is not concentrated, as is shown by the following theorem for the critical window. Part (i) is proved by Hatami and Molloy [24, Theorem 1.1] under very similar conditions, including a slightly stronger assumption than (2.32).

**Theorem 2.12** Suppose that (A1)–(A4) hold and $\varepsilon_n = O(n^{-1/3}(\mathbb{E}D_n^3)^{2/3})$. Suppose further that

$$\Delta_n = o\left(\frac{n \mathbb{E}D_n^3}{1/3}\right).$$  

(2.32)
Then the following hold:

(i) \( v(C_1) = O_p\left(n^{2/3}(\mathbb{E} D_n^3)^{-1/3}\right). \) In other words, for any \( \delta > 0 \) there exists \( K = K(\delta) \) such that

\[
\mathbb{P}\left(v(C_1) > Kn^{2/3}(\mathbb{E} D_n^3)^{-1/3}\right) < \delta.
\] (2.33)

(ii) Moreover, for any \( K < \infty \),

\[
\liminf_{n \to \infty} \mathbb{P}\left(v(C_1) > Kn^{2/3}(\mathbb{E} D_n^3)^{-1/3}\right) > 0.
\] (2.34)

Both (i) and (ii) hold with \( v(C_1) \) replaced by \( e(C_1) \).

Theorem 2.12 says that \( v(C_1)/(n^{2/3}(\mathbb{E} D_n^3)^{-1/3}) \) is bounded in probability, but not w.h.p. bounded by any fixed constant. In particular, \( v(C_1) \) normalizes in this way converges in distribution, at least along suitable subsequences, but it does not converge to a constant along any subsequence; hence the limit in distribution (along a subsequence) is really random and not deterministic. Moreover, Theorem 2.12(ii) shows that any subsequential limit has unbounded support. (The result by Hatami and Molloy [24, Theorem 1.1(a)] shows that any subsequential limit is strictly positive a.s.) This is in contrast to the supercritical case in Theorem 2.6. (This contrast is well known in the classical Erdős-Rényi case \( G(n,p) \), see e.g. Aldous [1], who describes the limit distribution explicitly.)

Remark 2.13 Condition (2.32) can be written as

\[
\max_i d_i^3 = o\left(\sum_{i \in [n]} d_i^3\right).
\] (2.35)

It thus says that no single vertex gives a significant contribution to \( \sum_{i \in [n]} d_i^3 \). See [24, Section 1.2] and Example 6.3 for counterexamples in the case when (2.32) does not hold. Note also that always \( n\mathbb{E} D_n^3 = \sum_{i \in [n]} d_i^3 \geq \Delta_n^3 \), so \( \Delta_n \leq (n\mathbb{E} D_n^3)^{1/3} \). Hence, (2.32) is only a weak restriction. (Hatami and Molloy [24] use a slightly stronger assumption, which, roughly, amounts to assuming \( \Delta_n = O((n\mathbb{E} D_n^3)^{1/3} / \log n) \).)

Remark 2.14 If \( \epsilon_n \approx n^{-1/3}(\mathbb{E} D_n^3)^{2/3} \), so we are on the upper boundary of the critical window in Theorem 2.12, then, using (2.32), \( \epsilon_n \Delta_n = o(\mathbb{E} D_n^3) \) and thus Theorem 3.1(iv) applies to a Galton-Watson process with offspring distribution \( \hat{D}_n \) starting with one individual (as in the proof of Theorem 2.10(iii)), and yields \( \rho_n \approx \epsilon_n / \mathbb{E} D_n^3 \approx n^{-1/3}(\mathbb{E} D_n^3)^{-1/3} \). Thus Theorem 2.12 shows that the giant component is of order \( O_p(np_n) \) in this case too, although \( v(C_1)/(np_n) \) does not converge to a constant.

Example 2.15 (Power-law degrees) Many real-world networks are claimed to have power-law degree sequences, see for example, [25, Chapter 1] and the references therein. As a result, various random graph models have been proposed that can yield such graphs, the configuration model being one of the most popular. Let \( \gamma > 1 \) and assume that, in addition to the assumptions above, for some constants \( C, c > 0 \),

\[
\mathbb{P}(D_n > k) \leq Ck^{-\gamma}, \quad k \geq 1,
\] (2.36)

\[
\mathbb{P}(D_n > k) \geq ck^{-\gamma}, \quad 1 \leq k < \epsilon_n^{-1/(\gamma-2)}.
\] (2.37)
The upper limit \( \varepsilon_n^{-1/(\gamma-1)} \) in (2.37) could be reduced by any fixed constant factor. Note that some limit is required, since \( D_n \) is discrete and (2.36) implies \( \Delta_n = O(n^{1/\gamma}) \). Then, in Theorem 3.1(i) and Example 3.4, we show that \( \rho_n \approx \varepsilon_n \) when \( \gamma > 3 \) (so \( \mathbb{E} D_n^3 = O(1) \)), while

\[
\rho_n \approx \varepsilon_n^{1/(\gamma-2)}
\]

when \( \gamma \in (2, 3) \). Theorem 2.6 applies and yields that, for \( \gamma > 3 \), and using the form in Theorem 2.8,

\[
v(C_1) = \frac{2n}{\kappa} \mathbb{E}(D_n(D_n - 2))(1 + o_p(1)), \text{ while, for } \gamma \in (2, 3), \ v(C_1) \approx n\varepsilon_n^{1/(\gamma-2)}.
\]

**Remark 2.16** The critical regime as in Theorem 2.12 has attracted considerable attention, see for example, [18, 19, 41, 51] for results on the sizes of the largest connected components. Riordan [51] investigates the scaling behavior of near-critical clusters under the assumption that all degrees are uniformly bounded. Dhara and coworker [18] perform an analysis under conditions that are close to ours when \( \mathbb{E} D_n^3 \to \mathbb{E} D^3 \), but focus on the scaling limit of critical clusters when \( n = 1 + \lambda n^{-1/3} + o(n^{-1/3}) \) (also for percolation on the configuration model, where the dependence on \( \lambda \) is identified as the multiplicative coalescent, see also Aldous [1] for the Erdős-Rényi setting and [22, 26, 30] for percolation on random graphs with given degrees).

In the case where \( \mathbb{E} D^3 = \infty \), and in the same vein as Example 2.15, often stronger assumptions are made and our results in Theorem 2.12 in this case are closest in spirit to those in [24] in that they only depend on the scaling of \( \varepsilon_n \) and \( \mathbb{E}[D_n^3] \). Order the degrees such that \( d_1 \geq d_2 \geq \cdots \geq d_n \). Joseph [41] assumes that \( (d_i)_{i \in [n]} \) are an i.i.d. sample from a distribution whose distribution function satisfies

\[ F(x) = cx^{-\gamma}(1 + o(1)) \text{ for } x \text{ large.} \]

In this case, \( (d_in^{-1/\gamma})_{i \geq 1} \) jointly converge in distribution to \( (\gamma_i^{-1/\gamma})_{i \geq 1} \), where \( \Gamma_i \) form a Poisson point process. Dhara and coworker [18] instead take \( d_i \) such that \( d_in^{-1/\gamma} \to c_i \), and, in particular, \( \mathbb{E} D_n^3 \sim n^{3/\gamma-1} \sum_{i \geq 1} c_i^3 \), where it is assumed that \( \sum_{i \geq 1} c_i^3 < \infty \), while \( \sum_{i \geq 1} c_i^2 = \infty \) (as is the case when \( c_i \approx i^{-1/\gamma} \) with \( \gamma \in (2, 3) \)). In this case, Theorem 2.12 suggests that the largest critical components should scale like

\[
n^{2/3}(\mathbb{E} D_n^3)^{-1/3} \approx n^{2/3}(n^{3/\gamma-1})^{-1/3} = n^{(\gamma-1)/\gamma}.
\]

The results in [19, 41] confirm this scaling, and show that the sizes of the largest connected components, rescaled by \( n^{-(\gamma-1)/\gamma} \), converge to a limiting sequence, while the critical window is of order \( n^{-(\gamma-2)/\gamma} \). Interestingly, the description of this limit looks quite different in [41] compared to [19], which is probably due to the fact that Joseph [41] also averages out over the randomness in the degrees. Interestingly, our results are also used in Dhara and coworker [17] to study the barely supercritical regime of percolation on the configuration model for \( \gamma \in (1, 2) \), where the percolation parameter tends to zero with the graph size to observe near-critical behavior.

### 2.5 Complexity of large components

The structure of components has received substantial attention in the literature, in particular, the existence of multicyclic components, that is, components \( C \) with \( e(C) > \nu(C) \). The detailed scaling limit results in [18, 19, 41, 51] resolve this question completely in the critical case. We investigate this question in the barely supercritical setting in Section 7 and find the asymptotic complexity of the largest component \( C_1 \), see Theorems 7.1 and 7.4–7.5. Here, for power-law degrees as in Example 2.15, the width of the critical window is tightly related to the growth of the complexity of the barely supercritical clusters. As can be expected, the complexity of \( C_1 \) interpolates between tight, as in the critical case, and linear in \( n \) as in the strictly supercritical regime (as shown in [47]).
2.6 Discussion

In this section, we discuss our results and pose further questions.

2.6.1 CLT for the giant component

It would be of interest to extend Theorem 2.6 to a statement about the fluctuations of $v(C_1)$ around $\mu_n \rho_n n$. In the light of central limit results for the processes that characterize the component sizes (see, e.g., Lemma 6.4), it is tempting to conjecture that a CLT holds for $v(C_1)$. From our methodology, however, this does not follow easily. A related question involves proving a CLT for the complexity $k(C_1)$ in the barely supercritical regime. (Cf. [49] for the Erdős-Rényi case.)

2.6.2 Related random graphs

Often, one can deduce results for rank-1 inhomogeneous random graphs (see [9] for the definition) from those derived for the configuration model conditioned on simplicity. Examples of such graphs are the Poissonian or Norros-Reittu random graph [48], the generalized random graph model [11], and the expected degree or Chung-Lu random graph [13–16]. In each of these models, edges are present independently: an edge between $i, j \in [n]$ is present with probability $p_{ij}$, where $p_{ij}$ is close to $w_i w_j / \ell_n$ for appropriately chosen vertex weights $(w_i)_{i \in [n]}$, and $\ell_n = \sum_{i \in [n]} w_i$ denotes the total weight. When the weight sequence satisfies conditions similar to (A1)–(A4), then also the random vertex degrees do, and thus results carry over rather easily from the configuration model to these models.

In slightly more detail, by [31], in the case where $\mathbb{E}[D_n^2] \to \mathbb{E}[D^2]$, the above three random graph models are asymptotically equivalent, so that proving a result for one immediately establishes it for any of the others as well. Furthermore, when conditioned on the degree sequence, the generalized random graph is a uniform random graph with that degree sequence [11]. We already know that Theorem 2.6 holds for uniform random graphs whose vertex degrees obey conditions (A1)–(A4), so that, by conditioning on the degree sequence, in order to deduce the same for rank-1 inhomogeneous random graphs, it suffices to prove that (A1)–(A4) indeed hold (with convergence in probability) for the degrees for the generalized random graph in the critical case. This proof is standard, and can, for example, be found in [4] or [25, Section 7.7]. The critical case of these models was studied in [5, 6].

3 THE BRANCHING PROCESS SURVIVAL PROBABILITY

Our proofs of Theorems 2.6 and 2.8-2.10 will use some estimates of the survival probability of barely supercritical Galton-Watson processes. In this section, we state and prove these estimates in a general form, for general Galton-Watson processes with offspring distribution $X_n$. We will return to the setting of the configuration model in the later sections, where we apply the results stated below with offspring distribution $X_n = \tilde{D}_n$. We will write $\rho_n$ for the survival probability of a branching process with offspring distribution $X_n$, starting with one individual. We also define $\alpha_n := -\log(1 - \rho_n)$.

Relation (3.5) below was conjectured and supported by a heuristic argument by Ewens [21]; Eshel [20] gave counter-examples but also a proof of (3.5) under some conditions. More general sufficient conditions were given by Hoppe [27] and Athreya [3]; both also gave a necessary and sufficient condition for (3.5) in terms of the probability generating function of the offspring distribution $X_n$. (The necessary and sufficient conditions in [3,27] are stated differently, but they can be seen to be equivalent, using integration by parts.) Here we give further results, stated in a form more suitable for our purposes, but note that there are overlaps with earlier ones in the literature. In particular, Theorem 3.1(ii)
follows easily from results in both [3, 27]. Furthermore, (3.3) was given by [27, Corollary 3.3] (in an equivalent formulation).

**Theorem 3.1** (Survival probability of a near-critical branching process.) Let \( X_n \) be a sequence of nonnegative integer-valued random variables such that \( \mathbb{E}[X_n] = 1 + \varepsilon_n \), where \( \varepsilon_n > 0 \) and \( \varepsilon_n \to 0 \) as \( n \to \infty \). Suppose also that \( \lim \inf_n \mathbb{P}(X_n \neq 1) > 0 \). Let \( \rho_n \) be the survival probability of a branching process with offspring distribution \( X_n \), starting with one individual, that is, the unique solution in \((0, 1)\) to

\[
1 - \rho_n = \mathbb{E}[(1 - \rho_n)^{X_n}].
\]

Then \( \rho_n \to 0 \) and, more precisely,

\[
\rho_n = O(\varepsilon_n).
\]

Furthermore,

\[
\rho_n \geq \frac{2\varepsilon_n}{\mathbb{E}X_n(X_n - 1)},
\]

and

\[
\varepsilon_n \geq \mathbb{E}(X_n \wedge (\rho_n X_n^2)).
\]

Moreover:

(i) If \( \mathbb{E}X_n^2 = O(1) \), then \( \rho_n \asymp \varepsilon_n \).

(ii) If \( X_n \overset{d}{\to} X \) for some random variable \( X \) and \( \mathbb{E}[X_n^2] \to \mathbb{E}[X^2] < \infty \), then

\[
\rho_n \sim \frac{2\varepsilon_n}{\mathbb{E}[X(X - 1)]}.
\]

(iii) If \( X_n \overset{d}{\to} X \) for some random variable \( X \) with \( \mathbb{E}[X^2] = \infty \), then

\[
\rho_n = o(\varepsilon_n).
\]

(iv) If \( \Delta_n \) are numbers such that \( X_n \leq \Delta_n \) a.s. and \( \varepsilon_n \Delta_n = o(\mathbb{E}X_n^2) \), then

\[
\rho_n \asymp \frac{\varepsilon_n}{\mathbb{E}[X_n(X_n - 1)]} \asymp \frac{\varepsilon_n}{\mathbb{E}X_n^2}.
\]

**Proof** We first show that \( \rho_n = o(1) \) as \( n \to \infty \). (For a more general result on continuity of the survival probability as a functional of the offspring distribution, see [12, Lemma 4.1].) To see this, assume, for a contradiction, that there exists a subsequence \( n_l \) such that \( \rho_{n_l} \to \rho > 0 \). Since \( \mathbb{E}X_n = O(1) \), the sequence \( X_n \) is tight, so there exists a further subsequence with \( X_n \overset{d}{\to} X \) along the subsequence, for some nonnegative integer-valued random variable \( X \). Furthermore, by the Skorohod coupling theorem [42, Theorem 4.30], we may assume that the variables \( X_n \) are defined on a probability space where the convergence is almost sure. Then, by dominated convergence, along the subsequence, \( \mathbb{E}[(1 - \rho_n)^{X_n}] \to \mathbb{E}[(1 - \rho)^X] \), and so, by (3.1),

\[
1 - \rho = \mathbb{E}[(1 - \rho)^X].
\]
In other words, \( \rho \) is the survival probability of a branching process with offspring distribution \( X \). On the other hand, by Fatou’s lemma, \( \mathbb{E} X \leq \lim \mathbb{E} X_n = 1 \), so this branching process is critical or subcritical; furthermore, \( \mathbb{P}(X \neq 1) \geq \lim \inf_n \mathbb{P}(X_n \neq 1) > 0 \) which excludes the case \( X = 1 \) a.s. Consequently, the survival probability \( \rho = 0 \), a contradiction. Hence \( \rho_n \to 0 \) as \( n \to \infty \).

Note that \( \alpha_n = -\log(1-\rho_n) > 0 \), and that \( \rho_n \to 0 \) implies
\[
\alpha_n \sim \rho_n. \tag{3.9}
\]

Also let
\[
F(x) := e^{-x} - 1 + x; \tag{3.10}
\]
note that \( F(x) \sim x^2/2 \) as \( x \to 0 \). Then (3.1) can be written
\[
\mathbb{E} e^{-\alpha_n X_n} = \mathbb{E}(1 - \rho_n)^X_n = 1 - \rho_n = e^{-\alpha_n}, \tag{3.11}
\]
and thus
\[
\mathbb{E} F(\alpha_n X_n) = \mathbb{E}(e^{-\alpha_n X_n} - 1 + \alpha_n X_n) = e^{-\alpha_n} - 1 + \alpha_n(1 + \varepsilon_n) = F(\alpha_n) + \alpha_n \varepsilon_n. \tag{3.12}
\]

Hence,
\[
\mathbb{E} \frac{F(\alpha_n X_n)}{\alpha_n^2} = \frac{F(\alpha_n)}{\alpha_n^2} + \frac{\varepsilon_n}{\alpha_n} = \frac{1}{2} + o(1) + \frac{\varepsilon_n}{\alpha_n}. \tag{3.13}
\]

Suppose now that (3.2) fails. Then there exists a subsequence with \( \varepsilon_n/\rho_n \to 0 \) and thus, by (3.9) and (3.13),
\[
\mathbb{E} \frac{F(\alpha_n X_n)}{\alpha_n^2} \to \frac{1}{2}. \tag{3.14}
\]

As above, by considering a subsubsequence, we may also assume that \( X_n \to X \) a.s. for some random variable \( X \), and then a.s., since \( \alpha_n \to 0 \),
\[
\frac{F(\alpha_n X_n)}{\alpha_n^2} \to \frac{X^2}{2}. \tag{3.15}
\]

By Fatou’s lemma, (3.15) and (3.14) yield
\[
\frac{1}{2} \mathbb{E} X^2 \leq \lim \inf_{n \to \infty} \mathbb{E} \frac{F(\alpha_n X_n)}{\alpha_n^2} = \frac{1}{2}. \tag{3.16}
\]

Furthermore, since the function \( F(x)/x \) is increasing on \( [0, \infty) \), (3.14) implies that, for any \( K > 0 \),
\[
\lim \sup_{n \to \infty} \mathbb{E} (X_n 1_{\{X_n \geq K\}}) \leq \lim \sup_{n \to \infty} \mathbb{E} \frac{KF(\alpha_n X_n)}{F(\alpha_n K)} = \lim_{n \to \infty} \frac{K\alpha_n^2/2}{F(\alpha_n K)} = \frac{1}{K}. \tag{3.17}
\]

Hence, still along the subsequence, the random variables \( X_n \) are uniformly integrable, and, since \( \mathbb{E} X_n \to 1 \) and \( X_n \to X \) a.s., we have \( \mathbb{E} X = 1 \). However, this together with (3.16) yields \( \text{Var}(X) = 0 \), so \( X = 1 \) a.s., which as above is excluded by our assumption \( \lim \inf_n \mathbb{P}(X_n \neq 1) > 0 \). This contradiction shows that (3.2) holds.
Next, for any integer \( m \geq 0 \) and \( \rho \in [0, 1] \), \((1 - \rho)^m \leq 1 - m\rho + \binom{m}{2} \rho^2 \). Hence,

\[
1 - \rho_n = \mathbb{E}(1 - \rho_n)^{X_n} \leq \mathbb{E}\left(1 - X_n\rho_n + \frac{X_n(X_n - 1)}{2} \rho_n^2\right),
\]

\[
= 1 - (1 + \epsilon_n)\rho_n + \frac{\mathbb{E}(X_n(X_n - 1))}{2} \rho_n^2,
\]

(3.18)

and (3.3) follows, recalling \( \rho_n > 0 \).

To show (3.4), note that (3.2) and (3.9) show that \( \alpha_n = O(\epsilon_n) \) and thus \( \epsilon_n/\alpha_n \) is bounded below. Furthermore, \( F(x) \approx x^2 \) for \( x \geq 0 \), and thus, by (3.13),

\[
\frac{\epsilon_n}{\alpha_n} \approx \frac{\epsilon_n}{\alpha_n} + \frac{1}{2} \approx \mathbb{E} \left( \frac{F(\alpha_nX_n)}{\alpha_n^2} \right) \approx \mathbb{E} \left( (\alpha_n^{-1}X_n) \wedge X_n^2 \right).
\]

(3.19)

Hence, using (3.9) again,

\[
\epsilon_n \approx \mathbb{E} \left( X_n \wedge (\alpha_nX_n^2) \right) \approx \mathbb{E} \left( X_n \wedge (\rho_nX_n^2) \right).
\]

(3.20)

(i): An immediate consequence of (3.2) and (3.3).

(ii): As above, we may assume \( X_n \rightarrow X \) a.s. (now for the full sequence), and thus (3.15). Since \( X_n \rightarrow X \) and \( \mathbb{E}X_n^2 \rightarrow \mathbb{E}X^2 < \infty \), the sequence \( X_n^2 \) is uniformly integrable. Furthermore, \( 0 \leq F(x) \leq x^2/2 \) for \( x \geq 0 \) and thus \( 0 \leq F(\alpha_nX_n)/\alpha_n^2 \leq X_n^2/2 \), so the sequence \( F(\alpha_nX_n)/\alpha_n^2 \) is also uniformly integrable, which together with (3.15) implies

\[
\mathbb{E} \left( \frac{F(\alpha_nX_n)}{\alpha_n^2} \right) \rightarrow \frac{1}{2} \mathbb{E}X^2.
\]

(3.21)

Moreover, the uniform integrability of \( X_n^2 \) also implies \( \mathbb{E}X = \lim_{n \to \infty} \mathbb{E}X_n = 1 \). Using (3.21) in (3.13), we thus find

\[
\frac{\epsilon_n}{\alpha_n} = \mathbb{E} \left( \frac{F(\alpha_nX_n)}{\alpha_n^2} \right) - \frac{1}{2} + o(1) = \frac{1}{2} (\mathbb{E}X^2 - 1) + o(1) = \frac{1}{2} \mathbb{E}(X(X - 1)) + o(1).
\]

(3.22)

As noted above, \( \mathbb{E}(X(X - 1)) = \text{Var } X > 0 \) and thus (3.22) yields, recalling (3.9),

\[
\frac{\epsilon_n}{\rho_n} \sim \frac{\epsilon_n}{\alpha_n} \sim \frac{1}{2} \mathbb{E}(X(X - 1)).
\]

(3.23)

A rearrangement yields (3.5).

(iii): We may again assume that (3.15) holds a.s., which now by Fatou’s lemma implies (cf. (3.16))

\[
\mathbb{E} \left( \frac{F(\alpha_nX_n)}{\alpha_n^2} \right) \rightarrow \infty.
\]

(3.24)

Thus \( \epsilon_n/\alpha_n \to \infty \) by (3.13). This yield (3.6), again using (3.9).

(iv): Note first that (3.2) and (3.3) imply that \( 1/\mathbb{E}[X_n(X_n - 1)] = O(1) \), that is, that \( \mathbb{E}[X_n(X_n - 1)] \) is bounded below. Since \( X_n(X_n - 1) \leq X_n^2 \leq 1 + 2X_n(X_n - 1) \), it follows that \( \mathbb{E}[X_n(X_n - 1)] \approx \mathbb{E}X_n^2 \), and thus the final “≥” in (3.7) holds.

A lower bound for \( \rho_n \) is given by (3.3), and it remains only to show a matching upper bound. By (3.4), there exists a constant \( C \) such that \( \mathbb{E}(X_n \wedge (\rho_nX_n^2)) < C\epsilon_n \). Let \( \beta_n := C\epsilon_n/\mathbb{E}X_n^2 \). Then
that integration by parts (or an equivalent Fubini argument), for any \( r \) simple calculation (we have equality in (3.18) and thus in (3.3)), \( \rho \) fails. Moreover, Remark 3.2 The assumption \( \lim \inf \mathbb{P}(X_n \neq 1) > 0 \) is essential: if \( X_n \xrightarrow{d} X = 1 \), almost anything can happen. For a simple example, let \( X_n \in \{0, 1, 2\} \) with \( \mathbb{P}(X_n = 0) = q_n \), \( \mathbb{P}(X_n = 2) = p_n \) and \( \mathbb{P}(X_n = 1) = 1 - p_n - q_n \) where \( p_n > q_n > 0 \) and \( p_n \to 0 \). Then \( \epsilon_n = p_n - q_n \) and, by (3.1) and a simple calculation (we have equality in (3.18) and thus in (3.3)), \( \rho_n = 1 - q_n/p_n = \epsilon_n/p_n \). Thus (3.2) fails. Moreover, \( \rho_n = 1 - q_n/p_n \) may converge to any number in \([0, 1]\), or may oscillate. (See also the examples in [27].)

Remark 3.3 If \( \mathbb{E}X_n^2 \to \infty \) but \( X_n \xrightarrow{d} X \) with \( \mathbb{E}X^2 < \infty \), it is not necessarily the case that (3.6) holds, but it is still possible; see Examples 3.7 and 3.8. Similarly, if \( \mathbb{E}X_n^2 \to C < \infty \) and \( X_n \xrightarrow{d} X \) but \( \mathbb{E}X^2 < C \), then (3.5) may or may not hold; see Examples 3.7 and 3.9.

We consider several examples illustrating various possible behaviors. See also the examples by Hoppe [27].

Example 3.4 (Power laws) Let \( 1 < \beta < 2 \) and assume that for some constants \( C, c > 0 \),

\[
\mathbb{P}(X_n > x) \leq Cx^{-\beta}, \quad x > 0, \tag{3.26}
\]

\[
\mathbb{P}(X_n > x) \geq cx^{-\beta}, \quad 1 \leq x < \epsilon_n^{-1/(\beta-1)}. \tag{3.27}
\]

Here, due to the size-biasing in (2.13), \( \beta \) is related to \( \gamma \) in Example 2.15 by \( \beta = \gamma - 1 \). Then, by an integration by parts (or an equivalent Fubini argument), for any \( r > 0 \),

\[
\mathbb{E}(X_n \wedge (rX_n^2)) = \int_0^{1/r} 2rx\mathbb{P}(X_n > x)\,dx + \int_{1/r}^{\infty} \mathbb{P}(X_n > x)\,dx,
\]

\[
\leq 2C r \int_0^{1/r} x^{1-\beta}\,dx + C \int_{1/r}^{\infty} x^{-\beta}\,dx,
\]

\[
= \left( \frac{2}{2-\beta} + \frac{1}{\beta-1} \right) Cr^{\beta-1}.
\]

Taking \( r = \rho_n \), this and (3.4) yield

\[
\epsilon_n = O(\rho_n^{\beta-1}). \tag{3.29}
\]

On the other hand, taking \( r = A\epsilon_n^{1/(\beta-1)} \) for a (large) constant \( A > 1 \), and assuming that \( n \) is so large that \( r < 1 \), by (3.27),

\[
\mathbb{E}(X_n \wedge (rX_n^2)) \geq \frac{1}{r} \mathbb{P}(X_n \geq \frac{1}{r}) \geq cr^{\beta-1} = cA^{\beta-1}\epsilon_n. \tag{3.30}
\]

Choosing \( A \) sufficiently large, this and (3.4) yield (for large \( n \))

\[
\mathbb{E}(X_n \wedge (rX_n^2)) > \mathbb{E}(X_n \wedge (\rho_nX_n^2)), \tag{3.31}
\]
and thus \( r > \rho_n \). Consequently, \( \rho_n = O(\varepsilon_n^{1/(\beta - 1)}) \), which together with (3.29) yield

\[
\rho_n \sim \varepsilon_n^{1/(\beta - 1)}. \tag{3.32}
\]

This example shows that \( \rho_n \) may decrease as an arbitrarily large power of \( \varepsilon_n \). (Choose \( \beta \) close to 1.)

**Example 3.5** For an instance of Example 3.4, let \( 1 < \beta < 2 \), and let \( X \) be a nonnegative integer-valued random variable with \( \mathbb{E} X = 1 \) and \( \mathbb{P}(X > x) \propto x^{-\beta} \) as \( x \to \infty \). Fix a sequence \( \varepsilon_n \to 0 \) (with \( \varepsilon_n > 0 \)) and a sequence \( M_n \) of integers with \( M_n \geq \varepsilon_n^{1/(\beta - 1)} \). Let \( X_n' \) := \( X \land M_n \), and define \( X_n \) by

\[
\mathbb{P}(X_n = k) = \begin{cases} 
\mathbb{P}(X_n' = 0) - \delta_n, & k = 0, \\
\mathbb{P}(X_n' = 1) + \delta_n, & k = 1, \\
\mathbb{P}(X_n' = k), & k \geq 2,
\end{cases} \tag{3.33}
\]

where \( \delta_n := \varepsilon_n + \mathbb{E}(X - X_n') \). Then \( \mathbb{E} X_n = \mathbb{E} X_n' + \delta_n = 1 + \varepsilon_n \) as required. Note that \( \mathbb{E}(X - X_n') \propto M_n^{-(\beta - 1)} = O(\varepsilon_n) \), so \( \delta_n \sim \varepsilon_n \); in particular \( \delta_n \to 0 \) and the definition (3.33) is valid at least for large \( n \) (since \( \mathbb{P}(X = 0) > 0 \) by \( \mathbb{E} X = 1 \)). Clearly, \( X_n \overset{d}{\to} X \).

Furthermore, (3.26) and (3.27) hold, and thus (3.32) holds.

Moreover, we may choose \( M_n \) arbitrarily large, and thus \( \mathbb{E} X_n^2 \approx M_n^{2 - \beta} \) can be made arbitrarily large; this shows that there is no formula similar to (3.5) giving \( \rho_n \), even within a constant factor, in terms of \( \varepsilon_n \) and \( \mathbb{E} X_n^2 \) (or \( \mathbb{E} X(X_n - 1) \)).

We may also take \( M_n = \infty \); then (3.32) still holds and \( \mathbb{E} X_n^2 = \infty \).

**Example 3.6** Choose \( \varepsilon_n \in (0, 1] \) with \( \varepsilon_n \to 0 \) and \( p_n \in (0, 1/n] \) with \( n p_n \to 0 \) and define (for \( n \geq 3 \)) \( X_n \) by

\[
\mathbb{P}(X_n = k) = \begin{cases} 
\frac{1 - \varepsilon_n + (n - 2)p_n}{2}, & k = 0, \\
\frac{1 + \varepsilon_n - 4p_n}{2}, & k = 2, \\
p_n, & k = n.
\end{cases} \tag{3.34}
\]

Then \( \mathbb{E} X_n = 1 + \varepsilon_n \) as required, \( X_n \overset{d}{\to} X \) with \( \mathbb{P}(X = 0) = \mathbb{P}(X = 2) = \frac{1}{2} \), and thus \( \mathbb{E} X = 1 \), \( \mathbb{E} X^2 = 2 \) and \( \mathbb{E} X(X - 1) = 1 \), and

\[
\mathbb{E} X_n^2 = 2 + n^2 p_n + o(1). \tag{3.35}
\]

In particular, \( \mathbb{E} X_n^2 \to \mathbb{E} X^2 \) if and only if \( n^2 p_n \to 0 \).

Furthermore,

\[
\mathbb{E} F(\alpha_n X_n) = \frac{1 + o(1)}{2} F(2\alpha_n) + p_n F(n\alpha_n) = \alpha_n^2 (1 + o(1)) + p_n F(n\alpha_n), \tag{3.36}
\]

and thus (3.12) implies

\[
\alpha_n \varepsilon_n = \alpha_n^2 (\frac{1}{2} + o(1)) + p_n F(n\alpha_n). \tag{3.37}
\]

We consider several cases of this in the following examples.
**Example 3.7** Choose $\epsilon_n$ and $p_n$ in Example 3.6 such that $np_n = o(\epsilon_n)$. Then $pnF(n\alpha_n) = O(p_n n\alpha_n) = o(\epsilon_n \alpha_n)$, and thus (3.37) yields $\alpha_n \epsilon_n \sim \frac{1}{2} \alpha_n^2$ and thus

$$\rho_n \sim \alpha_n \sim 2\epsilon_n,$$  

(3.38)

just as given by (3.5). This includes cases with $n^2 p_n \to 0$, when Theorem 3.1(ii) applies by (3.35), but also cases with $n^2 p_n \to \infty$, when $E X_n^2 \to \infty$ by (3.35). (For example, take $\epsilon_n = n^{-1/4}$ and $p_n = n^{-3/2}$.)

If we instead take $p_n = n^{-2}$ and $\epsilon_n = n^{-1/2}$, then $E X_n^2 \to 3 > E X^2$ by (3.35), while (3.5) nevertheless holds by (3.38).

**Example 3.8** Choose $\epsilon_n \leq n^{-1}$ in Example 3.6, so $\rho_n = O(n^{-1})$ by (3.2). Then $\rho_n X_n = O(1)$, so (3.4) yields

$$\epsilon_n \sim E(\rho_n X_n^2) = \rho_n E X_n^2.$$  

(3.39)

If we further choose $p_n$ with $np_n \to 0$ and $n^2 p_n \to \infty$, then $E X_n^2 \to \infty$ by (3.35), and thus $\rho_n = O(\epsilon_n)$ by (3.39). (For example, take $\epsilon_n = n^{-1}$ and $p_n = n^{-3/2}$.)

**Example 3.9** Choose $\epsilon_n = n^{-1}$ and $p_n = An^{-2}$ in Example 3.6, for some constant $A > 0$. Thus $E X_n^2 \to 2 + A$ by (3.35), and $E[X_n(X_n - 1)] \to 1 + A$. In particular, $E X_n^2 = O(1)$ and thus Theorem 3.1(i) yields $\alpha_n \sim \rho_n \sim n^{-1}$. More precisely, (3.37) yields, after multiplication by $n^2$,

$$n\alpha_n \sim \frac{1}{2}(n\alpha_n)^2 + AF(n\alpha_n).$$  

(3.40)

As just said, $n\alpha_n / \epsilon_n$ is bounded above and below, and (3.40) shows that, if $n\alpha_n \to a$ along some subsequence, then $a = \frac{1}{2} a^2 + AF(a)$, or

$$\frac{a - \frac{1}{2} a^2}{F(a)} = A.$$  

(3.41)

Hence $0 < a < 2$. Furthermore, it is easy to see (by differentiating) that $F(a)/(a - \frac{1}{2} a^2)$ is strictly increasing on $(0,2)$. Hence (3.41) has a unique solution $a = a(A) \in (0,2)$ for any $A > 0$, and thus $n\alpha_n \to a(A)$. Consequently, also $\rho_n / \epsilon_n = np_n \to a(A)$, given by (3.41).

It is easily verified that $2 > a(A) > 2/(1 + A)$. Hence, (3.5) does not hold, and neither does (3.5) with $E[X(X - 1)]$ replaced by $E[X_n(X_n - 1)]$.

## 4 | FURTHER PRELIMINARIES

### 4.1 More on $\rho_n$ and $\alpha_n$ in the barely supercritical case

Suppose that (A1)–(A4) are satisfied, and furthermore $\epsilon_n > 0$. (Note that the assumptions of Theorems 2.6 and 2.8–2.10 imply that $\epsilon_n > 0$, except possibly for some small $n$ that we may ignore.)

In what follows, $\rho_n$ will denote the survival probability of a Galton-Watson process with offspring distribution $D_n$, see Section 2.3 and (2.18). As in Section 3, it will often be convenient to work with

$$\alpha_n := -\log(1 - \rho_n).$$  

(4.1)
Lemma 4.1 If (A1)–(A4) are satisfied and \(\varepsilon_n > 0\), then \(\rho_n \to 0\),
\[
\alpha_n \sim \rho_n
\]  
and
\[
\varepsilon_n \leq \mathbb{E}(D_n^2 \wedge (\rho_n D_n^3)) \leq \mathbb{E}(D_n^2 \wedge (\alpha_n D_n^3)).
\]  

Proof  Theorem 3.1 applies to \(X_n = \hat{D}_n\), with \(X = \hat{D}\) and with \(\varepsilon_n\) as in Section 2.1 by (2.16). In particular, \(\rho_n = O(\varepsilon_n) \to 0\) and thus, by (4.1), \(\alpha_n \sim \rho_n\). Furthermore, by (3.4),
\[
\varepsilon_n \leq \mathbb{E}\left(\hat{D}_n \wedge (\rho_n \hat{D}_n^2)\right) \leq \mathbb{E}\left(D_n \wedge (\rho_n \hat{D}_n^2)\right).
\]  
Moreover, if \(D_n^* > 1\) then \(D_n^* \leq 2\hat{D}_n\). Thus, using (3.2) and (3.4),
\[
\mathbb{E}\left(D_n^* \wedge (\rho_n (D_n^*)^2)\right) \leq \rho_n + 4 \mathbb{E}\left(\hat{D}_n \wedge (\rho_n \hat{D}_n^2)\right) = O(\varepsilon_n).
\]  
Combining (4.4)–(4.5) and using (2.14), we find
\[
\varepsilon_n \leq \mathbb{E}(D_n^* \wedge (\rho_n (D_n^*)^2)) = \frac{1}{\mathbb{E}D_n} \mathbb{E}(D_n (1 - \rho_n)^D) = \frac{\mathbb{E}(D_n (1 - \rho_n)^D)}{\mu_n},
\]  
proving the first part of (4.3); the second follows from \(\alpha_n \sim \rho_n\). \[\Box\]

Note also that (2.18) implies, by (2.14),
\[
(1 - \rho_n)^2 = \mathbb{E}(1 - \rho_n)^D = \frac{\mathbb{E}(D_n(1 - \rho_n)^D)}{\mathbb{E}D_n} = \frac{\mathbb{E}(D_n(1 - \rho_n)^D)}{\mu_n},
\]  
which can be rewritten as
\[
\mathbb{E}(D_n e^{-\alpha_n D_n}) = \mu_n e^{-2\alpha_n}. \tag{4.8}
\]

In the case \(\mathbb{E}D_n^3 \to \mathbb{E}D^3 < \infty\), that is, when \(D_n^3\) are uniformly integrable, we have \(\mathbb{E}\hat{D}_n^3 \to \mathbb{E}\hat{D}^2\) by (2.15); hence (3.5) applies and yields, using (2.15) again and the notation (2.21), where now \(\kappa > 0\) by (3.5) or Remark 2.4,
\[
\alpha_n \sim \rho_n \sim \frac{2\varepsilon_n}{\mathbb{E}(\hat{D}(-1))} = \frac{2\varepsilon_n\mu}{\mathbb{E}(D(D - 1)(D - 2))} = \frac{2\varepsilon_n}{\kappa}. \tag{4.9}
\]

4.2  The Skorohod coupling theorem

We assume in (A1) that \(D_n \overset{d}{\longrightarrow} D\). By the Skorohod coupling theorem [42, Theorem 4.30], we may without loss of generality assume the stronger \(D_n \overset{a.s.}{\longrightarrow} D\); this will be convenient (although not really necessary) in some proofs. (We have already used the Skorohod coupling theorem in a similar way for \(X_n\) in Section 3, and will use it for a third set of variables in the proof of Lemma 5.7.)
4.3 | A semimartingale inequality

Our proofs below will use a semimartingale inequality to control the deviations of various random processes.

We say that a stochastic process $X(t)$, defined on an interval $[0, T]$, is a **semimartingale with drift** $\xi(t)$ (with respect to a filtration $(\mathcal{F}_t)$) if $X(t)$ is adapted and

$$X(t) = M(t) + \int_0^t \xi(u) \, du,$$

for some martingale $M(t)$. It is proved in [28, Lemma 2.2] that, if $X(t)$ is a bounded semimartingale with drift $\xi(t)$, then

$$\mathbb{E} \sup_{s \leq t \leq u} |X(t)|^2 \leq 13 \mathbb{E} |X(u)|^2 + 13 \left( \int_s^u \sqrt{\mathbb{E} \xi(t)^2} \, dt \right)^2 \leq 13 \mathbb{E} |X(u)|^2 + 13(u-s) \int_s^u \mathbb{E} [\xi(t)^2] \, dt. \quad (4.11)$$

We will be using the following modification of (4.11).

**Lemma 4.2**  Let $X(t)$ be a semimartingale with drift $\xi(t)$, defined on $[0, u]$. Then

$$\mathbb{E} \sup_{0 \leq t \leq u} |X(t)|^2 \leq 13 \sum_{j=0}^\infty \mathbb{E} |X(2^{-j}u)|^2 + 13 \int_0^u t \mathbb{E} [\xi(t)^2] \, dt. \quad (4.12)$$

**Proof**  Let $u_j := 2^{-j}u$. We have

$$\sup_{0 \leq t \leq u} |X(t)|^2 \leq \sum_{j=0}^\infty \sup_{u_{j+1} \leq t \leq u_j} |X(t)|^2,$$

since $X(t)$ is a.s. right-continuous at 0 (and everywhere) by (4.10). We take the expectation, and note that by (4.11),

$$\mathbb{E} \sup_{u_{j+1} \leq t \leq u_j} |X(t)|^2 \leq 13 \mathbb{E} |X(u_j)|^2 + 13(u_j-u_{j+1}) \int_{u_{j+1}}^{u_j} t \mathbb{E} [\xi(t)^2] \, dt \leq 13 \mathbb{E} |X(u_j)|^2 + 13 \int_{u_{j+1}}^{u_j} t \mathbb{E} [\xi(t)^2] \, dt. \quad (4.14)$$

The result follows by (4.13) and (4.14).

Inequality (4.12) will yield better estimates than inequality (4.11) in cases when process $(X(t))$ takes relatively small values near time 0 (so that $\sum_{j=0}^\infty \mathbb{E} |X(2^{-j}u)|^2$ is finite and not too large) but has quite significant drift (so that $\int_0^u t \mathbb{E} [\xi(t)^2] \, dt$ is significantly smaller than $u \int_0^u \mathbb{E} [\xi(t)^2] \, dt$).

5 | THE SUPERCRITICAL CASE

As explained in Section 2.4, it suffices to prove Theorems 2.6 and 2.8–2.10 for the multigraph $G^*_n$, since the simple graph case follows by conditioning on simplicity. We thus consider the random multigraph $G^*_n := G^*(n, (d_i)_{i=1}^n)$ constructed by the configuration model.
5.1 A more general theorem

We follow the structure of proof in [36]. We explore the clusters of the multigraph given by the configuration model one by one, using the cluster exploration strategy introduced in [36, Section 4]. We regard each edge as consisting of two half-edges, each half-edge having one endpoint. We label the vertices as *sleeping* or *awake*, and the half-edges as *sleeping*, *active* or *dead*. The sleeping and active half-edges are called *living* half-edges. (During the exploration, the endpoint of a sleeping half-edge is sleeping, while the endpoint of an active or dead half-edge is awake.)

We start with all vertices and half-edges sleeping. We pick a vertex, make it awake and label its half-edges as active. We then take any active half-edge, say \( x \), and find its partner half-edge \( y \) in the graph; we label these two half-edges as dead; furthermore, if the endpoint of \( y \) is sleeping, we label it awake and all the other half-edges at this vertex active. We repeat the above steps as long as there is an active half-edge available. When there is no active half-edge left, then we have obtained the first component. We then pick another vertex, and reveal its component, and so on, until all the components have been found.

We apply this procedure to \( G^*_n \), revealing its edges during the process. This means that, initially, we only observe the vertex degrees and the half-edges, but not how they are joined into edges. Hence, each time we need a partner of an edge, it is uniformly distributed over all living half-edges, and the dead half-edges correspond to the half-edges that have already been paired. We choose our pairings by giving the half-edges i.i.d. random maximal lifetimes with distribution Exp\((1)\). In other words, each half-edge dies spontaneously at rate 1 (unless killed earlier), and the probability that, if not killed, it survives until time \( t \) is \( e^{-t} \). Each time we need to find the partner of a half-edge \( x \), we then wait until the next living half-edge \( \neq x \) dies, and take that one. This gives the following algorithm for simultaneously constructing and exploring the components of \( G^*(n, (d_i)_n^1) \):

\textbf{C1} Select a sleeping vertex and declare it awake and all of its half-edges active. To be precise, we choose the vertex by choosing a half-edge uniformly at random among all sleeping half-edges.

The process stops when there is no sleeping half-edge left; the remaining sleeping vertices are all isolated and we have explored all other components.

\textbf{C2} Pick an active half-edge (which one does not matter) and kill it, that is, change its status to dead.

\textbf{C3} Wait until the next half-edge dies (spontaneously). This half-edge is paired to the one killed in the previous step \textbf{C2} to form an edge of the graph. If the vertex it belongs to is sleeping, then we declare this vertex awake and all of its other half-edges active. Repeat from \textbf{C2} if there is any active half-edge; otherwise from \textbf{C1}.

The components are created between the successive times \textbf{C1} is performed: the vertices in the component created between two successive such times are the vertices awakened during the corresponding interval.

We let \( S_n(t) \) and \( A_n(t) \) be the numbers of sleeping and active half-edges, respectively, at time \( t \geq 0 \), and let \( L_n(t) = S_n(t) + A_n(t) \) denote the number of living half-edges. Furthermore, we let \( V_{n,k}(t) \) denote the number of sleeping vertices of degree \( k \) at time \( t \), and let \( V_n(t) \) be the number of sleeping vertices at time \( t \); thus

\[ V_n(t) = \sum_{k=0}^{\infty} V_{n,k}(t), \quad S_n(t) = \sum_{k=0}^{\infty} kV_{n,k}(t). \quad (5.1) \]

These (random) functions are right-continuous by definition. We denote left limits by, for example, \( S_n(t-) \).
Let $T_1 < T_2$ be random times when C1 is performed. Then the exploration starts on new components at times $T_1$ and $T_2$, and the components found between $T_1$ and $T_2$ in total have $V_n(T_1) - V_n(T_2)$ vertices and $S_n(T_1) - S_n(T_2)$ half-edges, and hence $[S_n(T_1) - S_n(T_2)]/2$ edges. Note also that $A_n(t) = 0$ when C1 is performed, and $A_n(t) \geq 0$ for every $t$.

We also introduce variants of $(S_n(t), A_n(t), V_n(t))_{t \geq 0}$ obtained by ignoring the effect of C1. Let $\tilde{V}_{n,k}(t)$ denote the number of vertices of degree $k$ such that all of their $k$ half-edges have their exponential maximal life times greater than $t$. Then $\tilde{V}_{n,k}(t)$ has a $\text{Bin}(n, e^{-kt})$ distribution, and the $(\tilde{V}_{n,k}(t))_{k=1}^{\infty}$ are independent random variables for any fixed $t$. Let

$$\tilde{V}_n(t) := \sum_{k=0}^{\infty} \tilde{V}_{n,k}(t), \quad \tilde{S}_n(t) := \sum_{k=0}^{\infty} k \tilde{V}_{n,k}(t), \quad (5.2)$$

and

$$\tilde{A}_n(t) := L_n(t) - \tilde{S}_n(t) = A_n(t) - (\tilde{S}_n(t) - S_n(t)). \quad (5.3)$$

It is obvious that $\tilde{S}_n(t) \geq S_n(t)$; moreover, $\tilde{S}_n(t) - S_n(t)$ increases only when C1 is performed, and it is not difficult to show that, see [36, Lemma 5.3 and (5.7)],

$$0 \leq \tilde{S}_n(t) - S_n(t) = A_n(t) - \tilde{A}_n(t) < - \inf_{s \leq t} \tilde{A}_n(s) + \Delta_n, \quad (5.4)$$

where, as before, $\Delta_n := \max_{1 \leq i \leq n} d_i$ is the maximum vertex degree.

In order to explain the argument used to prove Theorem 2.6 more clearly, and to explain the connections to the previous versions of this argument used in [36], we give the argument in a general form (that includes the two versions in [36]), using certain parameters and functions, $\tau, \beta_n, \gamma_n, \hat{g}(t), \hat{h}(t), \psi_n(t)$. We assume that these satisfy certain regularity and asymptotic conditions (B1)–(B8), and then prove a general result, Theorem 5.4. The sequences $\beta_n, \gamma_n$ are near-critical scaling parameters, while $\hat{g}(t), \hat{h}(t), \psi_n(t)$ are asymptotic approximations for the processes $\tilde{V}_n(t), \tilde{S}_n(t), \tilde{A}_n(t)$ introduced above to study the exploration process. The choices of these parameters and functions used in the proofs of [36, Theorems 2.3 and 2.4] are described in Remarks 5.5 and 5.6. In order to prove Theorem 2.6, we instead make the choices in (5.17)–(5.21) below. (The reader who only wants a proof of Theorem 2.6 can thus assume these choices throughout.) We verify in Section 5.2 that the choices in (5.17)–(5.21) actually satisfy the assumptions (B1)–(B8).

Assumptions (B1)–(B8) are as follows.

(B1) $\tau > 0$ is fixed.

(B2) $(\beta_n)$ and $(\gamma_n)$ are sequences of positive numbers such that $\gamma_n = O(\beta_n)$.

(B3) $\hat{g}, \hat{h} : [0, \infty) \to \mathbb{R}$ are continuous functions; $\hat{g}$ is strictly positive on $(0, \infty)$ and $\hat{h}$ is strictly increasing on $(0, \infty)$.

(B4) $(\psi_n)$ is a sequence of continuous functions on $[0, 2\tau]$ such that:

(a) $\psi_n(0) = 0$;
(b) $\psi_n(\tau) = o(1)$;
(c) for some $\tau' > 0$, $\psi_n(t) \geq 0$ on $[0, \tau']$;
(d) for any compact interval $[a, b] \subset (0, \tau)$, $\lim_{n \to \infty} \inf_{a \leq t \leq b} \psi_n(t) > 0$;
(e) for every $t > \tau$, $\lim_{n \to \infty} \sup_{t \leq s \leq \tau} \psi_n(t) < 0$;
(f) $(\psi_n)$ is equicontinuous at $\tau$, that is, if $t_n \to \tau$, then $\psi_n(t_n) \to 0$. 
(B5) \[
\sup_{t \leq 2\tau} \left| \frac{1}{n\gamma_n} \tilde{A}_n(\beta_n t) - \psi_n(t) \right| \overset{p}{\to} 0.
\]

(B6) \[
\sup_{t \leq 2\tau} \left| \frac{1}{n\beta_n} (\tilde{V}_n(0) - \tilde{V}_n(\beta_n t)) - \tilde{g}(t) \right| \overset{p}{\to} 0.
\]

(B7) \[
\sup_{t \leq 2\tau} \left| \frac{1}{n\beta_n} (\tilde{S}_n(0) - \tilde{S}_n(\beta_n t)) - \tilde{h}(t) \right| \overset{p}{\to} 0.
\]

(B8) \[
\frac{\Delta_n}{n\gamma_n} \to 0.
\]

Note that (B6) and (B7) imply that necessarily \( \tilde{g}(0) = \tilde{h}(0) = 0 \).

**Remark 5.1** (Some intuition behind (B1)-(B8)) In all our applications, we will take \( \beta_n = \alpha_n \sim \rho_n \).

We see that \( \beta_n \) arises in two ways in our conditions. The first is the time scale on which the giant is found as all our processes are evaluated at time \( \beta_n t \). The second as the scaling of \( \tilde{S}_n \) and \( \tilde{V}_n \), which scale like \( n\beta_n \). The fact that these are the same is a sign that necessarily \( \gamma_n \) is close to linear for small \( t \). Furthermore, \( n\gamma_n \) is the size of \( \tilde{A}_n \), which will be proved to be close to \( A_n \). Since \( \tilde{A}_n \) is the difference of two processes that both run on scale \( n\beta_n \) and are positive, it follows that \( \gamma_n = O(\beta_n) \) should hold due to possible cancellations. In Remark 5.9, we will intuitively explain how \( \gamma_n \), which is the scale of the number of active vertices, arises and how our conditions on \( \epsilon_n \) can be interpreted in terms of the concentration of the process \( \big( \tilde{A}_n(\beta_n t) \big)_{t \geq 0} \).

**Remark 5.2** In the case when \( \psi_n = \psi \) does not depend on \( n \), (B4) says simply that \( \psi \) is continuous with \( \psi(0) = \psi(\tau) = 0 \), \( \psi > 0 \) on \( (0, \tau) \) and \( \psi < 0 \) on \( (\tau, 2\tau) \). In general, (B4) should be interpreted as an asymptotic version of this. In particular, for any \( \epsilon > 0 \) with \( \epsilon < \tau \), we have \( \psi_n(\tau - \epsilon) > 0 \) and \( \psi_n(\tau + \epsilon) < 0 \) for all large \( n \); it follows that, at least for large \( n \), \( \psi_n \) has a zero \( t_n > 0 \) such that \( t_n \to \tau \). Furthermore, every zero of \( \psi_n \) is \( o(1) \), \( \tau + o(1) \) or \( \geq \tau \).

**Remark 5.3** If, at least for all large \( n \), \( \psi_n \) is concave on \( [0, 2\tau] \) (which is the case in our main application), then (B4) can be replaced by the simpler

(B4') \( \psi_n \) is continuous and concave on \( [0, 2\tau] \) and such that \( \psi_n(0) = 0 \), \( \psi_n(\tau) = o(1) \), \( \psi_n(2\tau) = O(1) \) and \( \liminf_{n \to \infty} \psi_n(\tau/2) > 0 \);

in fact, (B4') is easily seen to imply (B4) (with, eg, \( t' = \tau/2 \)), at least for large \( n \), which suffices.

We now state a general theorem concerning the largest and second largest component sizes under assumptions (B1)-(B8). Recall that, for a component \( C \), we write \( v(C) \) and \( e(C) \) to denote the number of vertices and edges in the component, respectively. (In Lemma 5.8 we extend this notation to the case where \( C \) is a union of several components.)
**Theorem 5.4** Under assumptions (B1)-(B8),

\[
\begin{align*}
\nu(C_1) &= n\beta_n\hat{g}(\tau) + o_p(n\beta_n), \\
\varepsilon(C_1) &= n\beta_n\hat{h}(\tau)/2 + o_p(n\beta_n).
\end{align*}
\]

Furthermore, \(\nu(C_2), \varepsilon(C_2) = o_p(n\beta_n)\).

The proof of Theorem 5.4 follows [36, Sections 5 and 6] with minor modifications, omitting some details (and repeating others). Before giving the details, we offer some intuition behind its statement. Suppose that we are able to show (as we will later) that \((S_n(t), A_n(t), V_n(t))\) are close to \((\tilde{S}_n(t), \tilde{A}_n(t), \tilde{V}_n(t))\). By Remark 5.2 and (B5), there is a large component whose exploration commences within time \(o_p(\beta_n\tau)\) and ends at time \(\beta_n\tau(1 + o_p(1))\); this turns out to be the largest component. Moreover, by (B6), the number of vertices in this component is \(n\beta_n\hat{g}(\tau)(1 + o_p(1))\); and, by (B7), the number of half-edges is close to \(n\beta_n\hat{h}(\tau)(1 + o_p(1))\).

**Remark 5.5** We note that [36, Theorem 2.3] is one example of Theorem 5.4, with \(\tau = -\ln\xi, \beta_n = \gamma_n = 1, \psi_n(t) = \psi(t) = H(e^{-t}), \hat{g}(t) = 1 - g(e^{-t}), \hat{h}(t) = h(1 - h(e^{-t}) = \mu(1 - e^{-2t}) + \psi(t); in this case, (B5), (B6), (B7) are [36, (5.6), (5.2), (5.3)]. (Here, \(\psi_n(t)\) is not always concave.)

**Remark 5.6** Similarly, [36, Theorem 2.4] is another instance of Theorem 5.4, now with \(\beta_n = \mathbb{E}D_n(D_n - 2) \to 0\) as in [36], \(\gamma_n = \beta_n^2, \psi_n(t) = \psi(t) = t - \beta t^2/2, \tau = 2/\beta, \hat{g}(t) = \mu t, \hat{h}(t) = 2\mu t; for (B5), (B6), (B7), see [36, (6.7)], Lemma 6.3 and the Taylor expansions in the proof of Lemma 6.4]. Note that (B8) holds, since \(n^{2/3}\gamma_n \to \infty\) and \(\Delta_n = o(n^{1/3})\), because \(D_n^3\) is uniformly integrable. (Warning: \(\beta_n\) here has a different meaning than \(\beta\) and \(\beta_n\) in [36, Theorem 2.4 and (2.11)].)

We will see later that also Theorem 2.6 follows from Theorem 5.4.

The proof of Theorem 5.4 will use the following lemmas; the second generalizes [36, Lemmas 5.6 and 6.4].

**Lemma 5.7** Assume (B1)–(B8) and let \(T_n\) be random times such that \(T_n \xrightarrow[p]{} \tau\). Then

\[
\sup_{0 \leq t \leq T_n} \frac{1}{n\gamma_n} \left| \tilde{S}_n(\beta_n t) - S_n(\beta_n t) \right| = \sup_{0 \leq t \leq T_n} \frac{1}{n\gamma_n} \left| \tilde{A}_n(\beta_n t) - A_n(\beta_n t) \right| \xrightarrow[p]{} 0.
\]

**Proof** We may replace \(T_n\) by \(T_n \wedge (2\tau)\), since w.h.p. \(T_n \wedge (2\tau) = T_n\); hence we may assume that \(T_n \leq 2\tau\). Furthermore, using the Skorohod coupling theorem, we may assume that \(T_n \xrightarrow[a.s.]{} \tau\). We note next that this implies

\[
\inf_{0 \leq t \leq T_n} \psi_n(t) \to 0
\]

a.s., and thus in probability. In fact, if (5.8) fails at some point in the probability space, and \(T_n \to \tau\), then there exists \(t_n\), at least for some subsequence of \(n\), with \(0 \leq t_n \leq T_n = \tau + o(1)\) and \(\psi_n(t_n) < -\epsilon\), for some \(\epsilon < 0\). (Recall that \(\psi_n(0) = 0\), so the infimum is never positive.) We may select a further subsequence with \(t_n \to t' \in [0, \tau]\); this contradicts (B4). (Consider the cases \(t' = 0, t' = \tau\) and \(0 < t' < \tau\) separately, and use (B4)(c), (B4)(f), (B4)(d).)

By (5.8) and (B5),

\[
\inf_{0 \leq t \leq T_n} \frac{1}{n\gamma_n} \tilde{A}_n(\beta_n t) \xrightarrow[p]{} 0;
\]

(5.9)
and thus, by (5.4) and (B8),

$$
\sup_{0 \leq s \leq T_n} \frac{1}{n\gamma_n} |S_n(\beta_n t) - S_n(\beta_n t)| = \sup_{0 \leq s \leq T_n} \frac{1}{n\gamma_n} |A_n(\beta_n t) - \tilde{A}_n(\beta_n t)|
$$

$$
\leq \inf_{0 \leq s \leq T_n} \frac{1}{n\gamma_n} \tilde{A}_n(\beta_n t) + \frac{\Delta_n}{n\gamma_n} \overset{p}{\rightarrow} 0. \tag{5.10}
$$

In what follows we consider several random times. They generally depend on $n$ but we simplify the notation and denote them by $T_1, T'_1, \ldots$ as an abbreviation of $T_{1n}, \ldots$

**Lemma 5.8** Let $T'_1$ and $T'_2$ be two (random) times when C1 are performed, with $T'_1 \leq T'_2$, and assume that $T'_1 / \beta_n \overset{p}{\rightarrow} t_1$ and $T'_2 / \beta_n \overset{p}{\rightarrow} t_2$ where $0 \leq t_1 \leq t_2 \leq \tau$. If $\tilde{C}$ is the union of all components explored between $T'_1$ and $T'_2$, then, under assumptions (B1)-(B8),

$$
v(\tilde{C}) = n\beta_n (\check{g}(t_2) - \check{g}(t_1)) + o_p(n\beta_n),
$$

$$
e(\tilde{C}) = \frac{1}{2} n\beta_n (\check{h}(t_2) - \check{h}(t_1)) + o_p(n\beta_n).
$$

In particular, if $t_1 = t_2$, then $v(\tilde{C}) = o_p(n\beta_n)$ and $e(\tilde{C}) = o_p(n\beta_n)$.

**Proof** Taking, for $j = 1, 2$, $T_n = T'_j / \beta_n + \tau - t_j$ in (5.7), we see that

$$
\sup_{0 \leq s \leq T'_j} |\check{S}_n(t) - S_n(t)| = o_p(n\gamma_n). \tag{5.11}
$$

Since further $0 \leq \check{V}_n(t) - V_n(t) \leq \check{S}_n(t) - S_n(t)$, see (5.1), we have also

$$
\sup_{0 \leq s \leq T'_j} |\check{V}_n(t) - V_n(t)| = o_p(n\gamma_n). \tag{5.12}
$$

Since $\tilde{C}$ consists of the vertices awakened in the interval $[T'_1, T'_2)$, by (5.12), (B6) and (B3), as well as $\gamma_n = O(\beta_n)$,

$$
v(\tilde{C}) = V_n(T'_1) - V_n(T'_2) = \check{V}_n(T'_1) - \check{V}_n(T'_2) + o_p(n\gamma_n).
$$

$$
= n\beta_n (\check{g}(T'_2 / \beta_n) - \check{g}(T'_1 / \beta_n) + o_p(1))
$$

$$
= n\beta_n (\check{g}(t_2) - \check{g}(t_1) + o_p(1)).
$$

Similarly, using (5.11) and (B7),

$$
2e(\tilde{C}) = S_n(T'_1) - S_n(T'_2) = \check{S}_n(T'_1) - \check{S}_n(T'_2) + o_p(n\gamma_n).
$$

$$
= n\beta_n (\check{h}(T'_2 / \beta_n) - \check{h}(T'_1 / \beta_n) + o_p(1))
$$

$$
= n\beta_n (\check{h}(t_2) - \check{h}(t_1) + o_p(1)).
$$

\[\square\]
Proof of Theorem 5.4 Note that (5.7) (with $T_n = \tau$) and (B5) show that

$$\sup_{t \leq \tau} \left| \frac{1}{\gamma_n} A_n(\beta_n t) - \psi_n(t) \right| \overset{p}{\to} 0. \quad (5.13)$$

Hence, using (B4)(d), for every $\epsilon > 0$, w.h.p. $A_n(t) > 0$ on $[\beta_n \epsilon, \beta_n (\tau - \epsilon)]$, so no new components are started during that interval. On the other hand, if $0 < \epsilon < \tau$, then by (5.3), (B5) and (5.7),

$$\frac{1}{\gamma_n} \left[ \left( \tilde{S}_n(\beta_n (\tau + \epsilon)) - S_n(\beta_n (\tau + \epsilon)) \right) - \left( \tilde{S}_n(\beta_n \tau) - S_n(\beta_n \tau) \right) \right]$$

$$= \frac{1}{\gamma_n} \left[ \left( A_n(\beta_n (\tau + \epsilon)) - \tilde{A}_n(\beta_n (\tau + \epsilon)) \right) - \left( A_n(\beta_n \tau) - \tilde{A}_n(\beta_n \tau) \right) \right]$$

$$\geq -\frac{1}{\gamma_n} \tilde{A}_n(\beta_n (\tau + \epsilon)) - \frac{1}{\gamma_n} A_n(\beta_n \tau) - \tilde{A}_n(\beta_n \tau)$$

$$= -\psi_n(\tau + \epsilon) + o_p(1).$$

This is w.h.p. positive, since $\limsup_{n \to \infty} \psi_n(\tau + \epsilon) < 0$ by (B4)(e), and then C1 is performed at least once between $\beta_n \tau$ and $\beta_n (\tau + \epsilon)$.

Consequently, if $T_1$ is the last time C1 is performed before $\beta_n \tau/2$ and $T_2$ is the next time, then w.h.p. $0 \leq T_1 \leq \beta_n \epsilon$ and $\beta_n (\tau - \epsilon) \leq T_2 \leq \beta_n (\tau + \epsilon)$. Since $\epsilon$ can be chosen arbitrarily small, this shows that $T_1/\beta_n \overset{p}{\to} 0$ and $T_2/\beta_n \overset{p}{\to} \tau$.

Let $C'$ be the component explored between $T_1$ and $T_2$. By Lemma 5.8 (with $t_1 = 0$ and $t_2 = \tau$), $C'$ has

$$v(C') = n \beta_n (\hat{g}(\tau) + o_p(1)) \quad (5.14)$$
v
$$e(C') = \frac{1}{2} n \beta_n (\hat{h}(\tau) + o_p(1)) \quad (5.15)$$
edges.

It remains to prove that all other components have only $o_p(n \beta_n)$ edges (and thus vertices) each. (This implies $C_1 = C'$.) We argue as in [36, pp. 213-214 (end of Section 6)]. We fix a small $\epsilon > 0$ and say that a component is large if it has at least $\epsilon n \beta_n$ edges, and thus at least $2 \epsilon n \beta_n$ half-edges. If $\epsilon$ is small enough, then w.h.p. $C'$ is large by (5.15), and further $(\hat{h}(\tau) - \epsilon) n \beta_n < 2 \epsilon (C') < (\hat{h}(\tau) + \epsilon) n \beta_n$. Let $E_\epsilon$ be the event that $2 \epsilon (C') < (\hat{h}(\tau) + \epsilon) n \beta_n$ and that the total number of half-edges in large components is at least $(\hat{h}(\tau) + 2 \epsilon) n \beta_n$.

It follows by Lemma 5.8 applied to $T_0 = 0$ and $T_1$ that the total number of vertices and half-edges in components found before $C'$ is discarding the edge $o_p(\beta_n)$. Thus there exists a sequence $\beta'_n$ of constants such that $\beta'_n = o(\beta_n)$ and w.h.p. at most $n \beta'_n$ vertices are awakened and at most $n \beta'_n$ half-edges are made active before $T_1$, when the first large component is found.

Let us now condition on the final graph obtained through our component-finding algorithm. It follows from our specification of $C_1$ that, given $G^*(n, (d_i)_1)$, the components appear in our process in size-biased order (with respect to the number of edges), obtained by picking half-edges uniformly at random (with replacement, for simplicity) and taking the corresponding components, ignoring every component that already has been taken. We have seen that w.h.p. this finds components containing at most $n \beta'_n$ vertices and half-edges before a half-edge in a large component is picked. Therefore, starting again at $T_2$, w.h.p. we find at most $n \beta'_n$ half-edges in new components before a half-edge is chosen in some large component; this half-edge may belong to $C'$, but if $E_\epsilon$ holds, then with probability at least
\[ \varepsilon_1 := 1 - \left( \hat{h}(\tau) + \varepsilon \right) / (\hat{h}(\tau) + 2\varepsilon) > 0 \] does not, and therefore it belongs to a new large component. Consequently, with probability at least \( \varepsilon_1 \mathbb{P}(\mathcal{E}_\varepsilon) + o(1) \), the algorithm finds a second large component at a time \( T_3 \), and less than \( n\beta'_n \) vertices and half-edges between \( T_2 \) and \( T_3 \). In this case, let \( T_4 \) be the time this second large component is completed. If no such second large component is found, let for definiteness \( T_3 = T_4 = T_2 \).

The number of half-edges found between \( T_2 \) and \( T_3 \) is, using \( \tilde{S}_n(t) \geq S_n(t) \), (5.7) with \( T_n = T_2/\beta_n \), (B2) and (B7) together with the fact that \( T \)
\[ \tilde{S}_n(T_2) - S_n(T_2) \geq \tilde{S}_n(T_2) - S_n(T_2) - \tilde{S}_n(T_3) \]
\[ = \tilde{S}_n(T_2) - \tilde{S}_n(T_3) + o_p(\gamma_n) \]
\[ \geq \tilde{S}_n(T_2) - \tilde{S}_n((2\beta_n \tau) \land T_3) + o_p(\gamma_n) \]
\[ = n\beta_n \left( \hat{h}(2\tau) \land (T_3/\beta_n) \right) - \tilde{h}(T_2/\beta_n) + o_p(n\beta_n). \]

Since, by the definitions above, this is at most \( n\beta'_n = o(n\beta_n) \), it follows that \( \hat{h}(2\tau) \land (T_3/\beta_n) \leq \tilde{h}(T_2/\beta_n) \leq \tilde{h}(T_2/\beta_n) \leq o_p(1) \). Furthermore, \( T_2 \leq T_3 \) and \( T_2/\beta_n \to \tau \), and thus w.h.p. \( T_2/\beta_n \leq 2\tau \). Hence, using (B3), it follows that \( (2\tau) \land (T_3/\beta_n) = \tau = o_p(1) \), and thus \( T_3/\beta_n \to \tau \). Consequently, (5.7) applies to \( T_n = T_3/\beta_n \), and, since no C1 is performed between \( T_3 \) and \( T_4 \), using also (B8) again,
\[ \sup_{t \leq T_3} \left| \tilde{S}_n(t) - S_n(t) \right| \leq \sup_{t \leq T_3} \left| \tilde{S}_n(t) - S_n(t) \right| + \Delta_n = o_p(\gamma_n). \] (5.16)

Let \( t_0 \in (\tau, 2\tau) \); then by (B4)(e), for some \( \delta > 0 \), \( \psi_n(t_0) < -2\delta \) for all large \( n \), and thus (B5) shows that w.h.p. \( \tilde{A}_n(\beta_n t_0) \leq -n\gamma_n \delta \), and thus
\[ \tilde{S}_n(\beta_n t_0) - S_n(\beta_n t_0) = A_n(\beta_n t_0) - A_n(\beta_n t_0) \geq -A_n(\beta_n t_0) \geq n\gamma_n \delta. \]

Hence (5.16) shows that w.h.p. \( T_4 < \beta_n t_0 \). Since \( t_0 - \tau \) can be chosen arbitrarily small, and further \( T_2 \leq T_3 \leq T_4 \) and \( T_2/\beta_n \to \tau \), it follows that \( T_4/\beta_n \to \tau \).

Finally, by Lemma 5.8 again, this time applied to \( T_3 \) and \( T_4 \), the number of edges found between \( T_3 \) and \( T_4 \) is \( o_p(n\beta_n) \). Hence, w.h.p. there is no large component found there, although the construction gave a large component with probability at least \( \varepsilon_1 \mathbb{P}(\mathcal{E}_\varepsilon) + o(1) \). Consequently, \( \varepsilon_1 \mathbb{P}(\mathcal{E}_\varepsilon) = o(1) \) and thus \( \mathbb{P}(\mathcal{E}_\varepsilon) = o(1) \).

Recalling the definition of \( \mathcal{E}_\varepsilon \), we see that w.h.p. the total number of half-edges in large components is less than \((\hat{h}(\tau) + 2\varepsilon)n\beta_n \); since w.h.p. at least \((\hat{h}(\tau) - \varepsilon)n\beta_n \) of these belong to \( C' \), see (5.15), there are at most \( 3\varepsilon n\beta_n \) half-edges, and therefore at most \( \frac{3}{2} \varepsilon n\beta_n + 1 \) vertices, in any other component.

Choosing \( \varepsilon \) small enough, this shows that w.h.p. \( C_1 = C' \), and further \( v(C_2) \leq e(C_2) + 1 \leq \frac{3}{2} \varepsilon n\beta_n + 1 \). This completes the proof of Theorem 5.4.

5.2 Proof of Theorems 2.6–2.10

Now suppose that we are given a sequence of degree distributions \( D_n \) satisfying the conditions (A1)–(A4). We choose the parameters in (B1)-(B8) as follows, where \( \rho_n \) as before is the survival probability of a Galton-Watson process with offspring distribution \( \tilde{D}_n \); see (2.18); recall that \( \rho_n > 0 \) since \( \varepsilon_n > 0 \). (Note that \( \alpha_n \) in (5.18) is the same as in (4.1).) Also recall that \( \mu = \mathbb{E} D \). Define
\[ \tau := 1, \] (5.17)
\[ \beta_n := \alpha_n = -\log(1 - \rho_n), \]  
(5.18)

\[ \tilde{g}(t) := \mu t, \quad \tilde{h}(t) := 2\mu t, \]  
(5.19)

\[ \gamma_n := \mathbb{E}(D_n(1 \land \alpha DD_n)^2), \]  
(5.20)

\[ \psi_n(t) := \gamma_n^{-1}(\mu_n e^{-2\alpha_n t} - \mathbb{E}(D_n e^{-\alpha DD_n})). \]  
(5.21)

Recall that, by Lemma 4.1, \( \rho_n \to 0 \) and \( \beta_n = \alpha_n \to 0 \).

**Remark 5.9** (Intuition behind (B1)-(B8) continued) Recall by (B5) that \( n\gamma_n \) is the size of \( \tilde{A}_n(\alpha_n t) \). See (5.51), where we show that \( \mathbb{E}[\tilde{A}_n(\alpha_n t)] = \epsilon_n e^{-\alpha_n t} + O(1) - \sum_{k=0}^{\infty} k n_k e^{-\alpha_n k} \), which by Taylor expansion is indeed of the order \( n\gamma_n = n \mathbb{E}(D_n(1 \land \alpha DD_n)^2) \). This explains how \( \gamma_n \) in (5.20) arises.

Let us next relate this to the condition \( \epsilon_n \gg n^{-1/3}(\mathbb{E} D_n^2)^{-2/3} \). Every time when \( \tilde{A}_n \) hits zero, a connected component is explored. Since \( A_n(\alpha_n t) \approx \tilde{A}_n(\alpha_n t) \) by Lemma 5.7, one can therefore expect that the size of the barely supercritical component is well concentrated precisely when the hitting time of zero of \( \tilde{A}_n \) is. This follows when the process \( t \mapsto \tilde{A}_n(\alpha_n t) \) is well concentrated (and its limit has a unique first zero). Now, \( \tilde{A}_n(\alpha_n t) = L_n(\alpha_n t) - \tilde{S}_n(\alpha_n t) \), and both processes turn out to have similar variances, the one for \( \tilde{S}_n(\alpha_n t) \) being easier to compute since \( \tilde{S}_n(t) = \sum_{k=0}^{\infty} k \tilde{V}_{n,k}(t) \) with \( \tilde{V}_{n,k}(t) \) independent Bin\((n_k, e^{-\gamma_n t}) \) random variables. Thus,

\[ \text{Var}(\tilde{S}_n(\alpha_n t)) = \sum_{k \geq 0} k^2 n_k e^{-\gamma_n k t}(1 - e^{-\gamma_n k}) \leq n \mathbb{E}[D_n^2(1 \land (\alpha_n D_n))] \approx n\epsilon_n, \]  
(5.22)

where we crucially rely on (4.3). This suggests that the process \( t \mapsto \tilde{A}_n(\alpha_n t) \) is well concentrated precisely when \( n\epsilon_n \ll (n\gamma_n)^2 \). The latter turns out to be the case when \( \epsilon_n \gg n^{-1/3}(\mathbb{E} D_n^2)^{-2/3} \). Indeed, by Cauchy-Schwarz, \( \epsilon_n^2 = O(\gamma_n^2 \mathbb{E} D_n^3) \) (see also Lemma 5.19), so that \( n\epsilon_n/(n\gamma_n^2) = \epsilon_n^4/(\gamma_n^2 n\epsilon_n^3) = O((\mathbb{E} D_n^3)^{-2}/(n\epsilon_n^3)) \). This explains the barely supercriticality condition \( \epsilon_n \gg n^{-1/3}(\mathbb{E} D_n^2)^{-2/3} \) that we assume throughout this paper. While the above arguments only prove the one-way bounds that we need in the proof, the fact that we observe critical behavior when \( \epsilon_n = O(n^{-1/3}(\mathbb{E} D_n^2)^{-2/3}) \) (see Theorem 2.12) suggests that the above inequalities are in fact asymptotically sharp.

We next show that under the conditions of Theorem 2.6, these parameters satisfy (B1)–(B8) (possibly except for some small \( n \) that we may ignore). This will take a series of lemmas.

**Lemma 5.10** Assume (A1)–(A4). Then the parameters defined in (5.17)–(5.21) satisfy (B1), (B2), (B3) and (B4'), and thus also (B4), at least for \( n \) large. Furthermore,

\[ \beta_n^2 = \alpha_n^2 = O(\gamma_n^2). \]  
(5.23)

**Proof** (B1): Trivial.

(B2): Since \( \epsilon_n > 0 \), we have \( \rho_n > 0 \) and thus \( \alpha_n > 0 \) and \( \gamma_n > 0 \). Furthermore, by (5.20),

\[ \gamma_n \leq \mathbb{E}(D_n(\alpha_n D_n)) = \alpha_n \mathbb{E} D_n^2 = O(\alpha_n). \]  
(5.24)

(B3): Trivial by (5.19).

(B4'): By the definition (5.21) and (4.8), \( \psi_n(0) = 0 \) and \( \psi_n(\tau) = \psi_n(1) = 0 \). \( \psi_n(t) \) is trivially continuous. (Recall that each \( D_n \) is a discrete random variable taking only a finite number of different values.)
We next show that $\psi_n$ is concave on $[0, 2\pi] = [0, 2]$ for large $n$. (It is not always concave on $(0, \infty)$, nor does it have to be concave on $[0, 2]$ for small $n$, as can be seen by simple counterexamples with $D_n \in \{1, 3\}$.) By (5.21),
\[
\gamma_n \alpha_n^{-2} \psi_n''(t) = 4\mu_n e^{-2\alpha_n t} - \mathbb{E}(D_n^3 e^{-\alpha_n t D_n}).
\]
For every $t \in [0, 2]$ we have
\[
\mathbb{E}(D_n^3 e^{-\alpha_n t D_n}) = \mathbb{E}(D_n(D_n - 2)^2 e^{-\alpha_n t D_n}) + 4 \mathbb{E}(D_n^2 e^{-\alpha_n t D_n}) - 4 \mathbb{E}(D_n e^{-\alpha_n t D_n}) \geq \mathbb{E}(D_n(D_n - 2)^2 e^{-\alpha_n t D_n}) + 4 \mathbb{E}(D_n^2 e^{-\alpha_n t D_n}) - 4 \mathbb{E} D_n.
\]
For the first term on the right-hand side of (5.26) we may assume, by the Skorohod coupling (see Section 4.2), that $D_n \xrightarrow{a.s.} D$ and thus $D_n(D_n - 2)^2 e^{-\alpha_n D_n} \xrightarrow{a.s.} D(D(D - 2)^2); thus Fatou’s lemma yields
\[
\liminf_{n \to \infty} \mathbb{E}(D_n(D_n - 2)^2 e^{-\alpha_n D_n}) \geq \mathbb{E}(D(D - 2)^2).
\]
Next, using (4.3),
\[
\mathbb{E}(D_n^2(1 - e^{-\alpha_n D_n})) \leq 2 \mathbb{E}(D_n^2(1 - \alpha_n D_n)) = O(\varepsilon_n) = o(1)
\]
and thus, using also (2.9),
\[
\mathbb{E}(D_n^2 e^{-\alpha_n D_n}) = \mathbb{E}(D_n^2) + O(\varepsilon_n) = \mathbb{E}(D_n(D_n - 2)) + 2 \mathbb{E} D_n + O(\varepsilon_n) = 2\mu + o(1).
\]
Combining (5.26)-(5.29) and $\mathbb{E} D_n = \mu_n \to \mu$, we obtain
\[
\liminf_{n \to \infty} \inf_{t \in [0, 2]} \mathbb{E}(D_n^3 e^{-\alpha_n t D_n}) \geq \mathbb{E}(D(D - 2)^2) + 8\mu - 4\mu
\]
and thus by (5.25) and (A3),
\[
\limsup_{n \to \infty} \sup_{t \in [0, 2]} \gamma_n \alpha_n^{-2} \psi_n''(t) \leq -\mathbb{E}(D(D - 2)^2) < 0.
\]
Consequently, for $n$ large, $\psi_n''(t) < 0$ on $[0, 2]$, and thus $\psi_n$ is concave in this interval.

Next we verify (5.23). In fact, if $D_n \neq 0$, then $1 \wedge \alpha_n D_n \geq \alpha_n$ and thus the definition (5.20) implies
\[
\gamma_n \geq \mathbb{E}(D_n \alpha_n^{-2}) = \alpha_n^2 \mu_n.
\]
Thus $\alpha_n^2 / \gamma_n \leq 1 / \mu_n = O(1)$, since $\mu_n \to \mu > 0$.

We now complete the proof of (B4’). We can write the definition (5.21) as
\[
\gamma_n \psi_n(t) = \mathbb{E}(D_n(1 - e^{-\alpha_n t D_n})) - \mu_n (1 - e^{-2\alpha_n t}).
\]
Since $\psi_n(1) = 0$, we thus have
\[
\gamma_n \psi_n(2) = \gamma_n \psi_n(2) - 2 \gamma_n \psi_n(1)
\]
\[
= -\mathbb{E}(D_n(1 - 2e^{-\alpha_n D_n} + e^{-2\alpha_n D_n})) + \mu_n (1 - 2e^{-2\alpha_n} + e^{-4\alpha_n})
\]
\[
= -\mathbb{E}(D_n(1 - e^{-\alpha_n D_n})^2) + \mu_n (1 - e^{-2\alpha_n})^2.
\]
Consequently, using (5.32),
\[
\gamma_n \psi_n(2) \leq \mu_n \left(1 - e^{-2a_n}\right)^2 \leq 4 \mu_n \alpha_n^2 \leq 4 \gamma_n,
\] (5.35)
and, by (5.20),
\[
-\gamma_n \psi_n(2) \leq \mathbb{E}(D_n(1 - e^{-\alpha_n D_n}))^2 \leq \mathbb{E}(D_n(1 \land \alpha_n D_n)^2) = \gamma_n.
\] (5.36)
Consequently, \(-1 \leq \psi_n(2) \leq 4\) and thus \(|\psi_n(2)| \leq 4\).

Similarly,
\[
2 \gamma_n \psi_n(\frac{1}{2}) = 2 \gamma_n \psi_n(\frac{1}{2}) - \gamma_n \psi_n(1) = \mathbb{E}(D_n(1 - 2e^{-\alpha_n D_n/2} + e^{-\alpha_n D_n})) - \mu_n(1 - 2e^{-\alpha_n} + e^{-2\alpha_n})
\]
\[
= \mathbb{E}(D_n(1 - e^{-\alpha_n D_n/2})^2) - \mu_n(1 - e^{-\alpha_n})^2.
\] (5.37)

Denote the two terms on the right-hand side of (5.37) by \(A_1\) and \(A_2\). Since \(1 - e^{-x} \approx 1 \land x\),
\[
A_1 \approx \mathbb{E}(D_n(1 \land (\alpha_n D_n)^2)) = \gamma_n.
\] (5.38)

In order to show that \(\liminf_{n \to \infty} \psi_n(\frac{1}{2}) > 0\), it thus remains only to show that \(A_1\) is not cancelled by \(A_2\). First, \(\alpha_n \to 0\) and thus \(A_2 \sim \mu_n \alpha_n^2 \sim \mu \alpha_n^2\). Furthermore, since \(1 - e^{-x} \geq xe^{-x}\) for \(x \geq 0\),
\[
A_1 \geq \frac{\alpha_n^2}{4} \mathbb{E}(D_n e^{-\alpha_n D_n}).
\] (5.39)

Thus, using (5.30) and (A3),
\[
\liminf_{n \to \infty} \frac{A_1}{A_2} \geq \liminf_{n \to \infty} \frac{\mathbb{E}(D_n e^{-\alpha_n D_n})}{4\mu} \geq \frac{\mathbb{E}(D(D - 2)^2) + 4\mu}{4\mu} > 1.
\] (5.40)

Since \(A_1, A_2 \geq 0\), it follows that \(A_1 - A_2 \approx A_1\), and thus (5.37) and (5.38) yield
\[
2 \gamma_n \psi_n(\frac{1}{2}) \approx A_1 \approx \gamma_n,
\] (5.41)
which verifies \(\liminf_{n \to \infty} \psi_n(\frac{1}{2}) > 0\). This completes the proof of (B4').

\[\blacktriangleleft\]

Remark 5.11 Note, for later use, that we have shown that, for large \(n\) at least, \(\psi_n\) is concave on \([0, 2]\) with \(\psi_n(0) = \psi_n(1) = 0\) and, by (5.36), \(\psi_n(2) \geq -1\); hence 0 \(\geq \psi'_n(1) \geq -1\), and thus 0 \(\leq \psi_n(t) \leq 1 - t\) for \(t \in [0, 1]\) and 1 \(\leq \psi_n(t) \leq 0\) for \(t \in [0, 2]\), so \(|\psi_n(t)| \leq 1\) for \(t \in [0, 2]\).

We next show that (B5)-(B7) hold if we replace the random processes \(\tilde{A}_n, \tilde{V}_n\) and \(\tilde{S}_n\) by their expectations, at least under the extra assumption that \(\eta\gamma_n \to \infty\).

Lemma 5.12 (Asymptotics of means of \(\tilde{S}_n(t), \tilde{A}_n(t), \tilde{V}_n(t)\)) Assume (A1)-(A4), \(\varepsilon_n > 0\), and additionally that \(\eta\gamma_n \to \infty\). Then, with parameter values as in (5.17)-(5.21), for any fixed \(t_0\),
\[
\sup_{t \leq t_0} \left| \frac{1}{n\hat{p}_n} (\mathbb{E}[\tilde{S}_n(0)] - \mathbb{E}[\tilde{S}_n(\beta_n t)]) - \hat{h}(t) \right| = o(1),
\] (5.42)
\[ \sup_{t \leq 0} \left| \frac{1}{n} \left( \mathbb{E} \tilde{V}_n(0) - \mathbb{E} \tilde{V}_n(\beta_n t) \right) - \psi(t) \right| = o(1). \]  
\[ (5.43) \]

\[ \sup_{t \leq 0} \left| \frac{1}{n} \mathbb{E} \tilde{A}_n(\beta_n t) - \psi_n(t) \right| = o(1). \]
\[ (5.44) \]

**Proof** We have, using

\[ \mathbb{E} D_n^2 = \frac{1}{n} \sum k^2 n_k = \mathbb{E} D_n(D_n - 1) + \mathbb{E} D_n = \mu_n v_n + \mu_n = \mu_n(2 + \epsilon_n), \]
\[ (5.45) \]

\( \beta_n = \alpha_n \), and the definition (3.10),

\[ \frac{1}{n \beta_n} \left( \mathbb{E} \tilde{S}_n(0) - \mathbb{E} \tilde{S}_n(\beta_n t) \right) = \frac{1}{n \alpha_n} \sum_{k=1}^{\infty} k \left( \mathbb{E} \tilde{V}_{n,k}(0) - \mathbb{E} \tilde{V}_{n,k}(\alpha_n t) \right) 
= \frac{1}{n \alpha_n} \sum_{k=1}^{\infty} kn_k (1 - e^{-\alpha_n tk}) = \frac{1}{\alpha_n} \mathbb{E} (D_n(1 - e^{-\alpha_n tD_n})) 
= t \mathbb{E} D_n^2 + \frac{1}{\alpha_n} \mathbb{E} (D_n(1 - e^{-\alpha_n tD_n} - \alpha_n tD_n)) 
= t \mu_n (2 + \epsilon_n) - \frac{1}{\alpha_n} \mathbb{E} (D_n F(\alpha_n tD_n)). \]
\[ (5.46) \]

We now estimate the last term, noting that

\[ 0 \leq F(x) \leq x \wedge x^2. \]
\[ (5.47) \]

Thus, for all \( t \in [0, t_0] \),

\[ 0 \leq \frac{1}{\alpha_n} \mathbb{E} (D_n F(\alpha_n tD_n)) \leq \mathbb{E} (D_n (t_0 D_n \wedge (\alpha_n t_0^2 D_n^2))) \leq (t_0 + t_0^3) \mathbb{E} (D_n^2 \wedge (\alpha_n D_n^3)). \]
\[ (5.48) \]

By (4.3), this is \( O(\epsilon_n) = o(1), \) and (5.42) follows from (5.46) by the definition (5.19) of \( \hat{h}(t) \).

The proof of (5.43) is similar, and easier, as there is one fewer power of \( k \) involved.

To prove (5.44), note first that \( L_n(t) \) is a death process where individuals die at rate 1, except that when someone dies, another is immediately killed (by C2), so the number of living individuals drops by 2, except when the last is killed; moreover \( L_n(0) = \epsilon_n - 1 \), where we recall from (2.2) that \( \epsilon_n = n \mu_n \) is the total number of half-edges. We can couple \( L_n(t) \) with a similar process \( \tilde{L}_n(t) \) starting at \( \tilde{L}_n(0) = \epsilon_n \) so that both processes jump whenever the smaller jumps, and then

\[ |L_n(t) - \tilde{L}_n(t)| \leq 1 \]
\[ (5.49) \]

for all \( t \), cf. [36, Proof of Lemma 6.1]. Then \( \frac{1}{2} \tilde{L}_n(t) \) is a standard death process with intensity 2, starting at \( \epsilon_n/2 \), and thus \( \mathbb{E} \tilde{L}_n(t) = \epsilon_n e^{-2t} \). Hence,

\[ \left| \mathbb{E} L_n(t) - \epsilon_n e^{-2t} \right| = \left| \mathbb{E} L_n(t) - \mathbb{E} \tilde{L}_n(t) \right| \leq 1 \]
\[ (5.50) \]

for all \( t \geq 0 \). Consequently, uniformly in all \( t \geq 0 \),

\[ \mathbb{E} [\tilde{A}_n(\alpha_n t)] = \mathbb{E} [L_n(\alpha_n t)] - \mathbb{E} [\tilde{S}_n(\alpha_n t)] = \epsilon_n e^{-2\alpha_n t} + O(1) - \sum_{k=0}^{\infty} kn_k e^{-\alpha_n tk} \]
\[ (5.51) \]
and thus, by (5.21) and the assumption \(n\gamma_n \to \infty\),

\[
\frac{1}{n} \mathbb{E} \widetilde{A}_n(\alpha_t) = \mu_n e^{-2\alpha_nt} - \mathbb{E} \left( D_n e^{-\alpha_ntD_n} \right) + O\left(n^{-1}\right) = \gamma_n \psi_n(t) + o(\gamma_n),
\]

which proves (5.44).

**Remark 5.13** In the case when \(D_n^3\) is uniformly integrable, or equivalently \(\mathbb{E} D_n^3 \to \mathbb{E} D^3 < \infty\), the sequence \((\alpha_n^{-2}D_n) \land D_n^3\) is uniformly integrable (since \(D_n^3\) is), and converges a.s. to \(D^3\) if we assume \(D_n \overset{\text{a.s.}}{\to} D\), as we may by Section 4.2; consequently, using (5.20),

\[
\frac{\gamma_n}{\alpha_n^2} = \mathbb{E}((\alpha_n^{-2}D_n) \land D_n^3) \to \mathbb{E} D^3 < \infty.
\]

Thus, in this case, \(\gamma_n \asymp \alpha_n^2\). In other words, we could have defined \(\gamma_n\) as \(\alpha_n^2\) or, for example, \(\mu_n\alpha_n^2\) in this case, instead of by (5.20) (provided we modify \(\psi_n\) accordingly). Moreover, a simple calculation using (4.9), which we omit, shows that, with \(\kappa\) given by (2.21) and (2.22),

\[
\psi_n(t) := \frac{\kappa \mu}{2 \mathbb{E} D^3} (t - t^2) + o(1),
\]

uniformly on each compact interval; thus we may in this case as an alternative take \(\psi_n(t) := \frac{\kappa \mu}{2 \mathbb{E} D^3} (t - t^2)\), independently of \(n\). (Cf. Remark 5.2 and, with a simple change of time scale, Remark 5.6.)

On the other hand, if \(\mathbb{E} D^3 = \infty\), then, assuming again \(D_n \overset{\text{a.s.}}{\to} D\), we have \(\alpha_n^{-2}D_n \land D_n^3 \overset{\text{a.s.}}{\to} D^3\) since \(\alpha_n \to 0\). Thus Fatou’s lemma yields, instead of (5.53), \(\gamma_n/\alpha_n^2 \to \mathbb{E} D^3 = \infty\), that is,

\[
\alpha_n^2 = o(\gamma_n).
\]

Moreover, in this case it is, using (5.55), easy to see that if we define

\[
\varphi_n(t) := \mathbb{E} \left( D_n(1 - e^{-\alpha_nD_n}) \right) - 2\alpha_n\mu_n t,
\]

then

\[
\psi_n(t) := \frac{\varphi_n(t)}{\gamma_n} + o(1)
\]

uniformly on each compact interval; thus we may in this case as an alternative take \(\psi_n(t) := \varphi_n(t)/\gamma_n\).

In both these cases we can thus use simpler versions of \(\gamma_n\) and \(\psi_n\); however, we prefer not to do so; instead we use definitions (5.20) and (5.21), which work in all cases.

**Remark 5.14** Typically, as in Example 2.15, \(\mathbb{E}(D_n(\alpha_nD_n) \land (\alpha_nD_n)^2) \asymp \mathbb{E}(D_n(1 \land (\alpha_nD_n)^2))\) and then, by (5.20) and (4.3),

\[
\gamma_n = \mathbb{E}(D_n(1 \land (\alpha_nD_n)^2)) \asymp \mathbb{E}(D_n(\alpha_nD_n \land (\alpha_nD_n)^2)) \asymp \alpha_n \varepsilon_n.
\]

In this case, we could have used \(\gamma_n := \alpha_n \varepsilon_n\) instead of the choice (5.20) (provided we modify \(\psi_n\) accordingly).

We next show that the random variables \(\widetilde{A}_n(t), \widetilde{V}_n(t)\) and \(\widetilde{S}_n(t)\) are so well concentrated for all \(t\) that we may replace them in conditions (B5), (B6), (B7) by their expectations. For later use, we state the next estimates in a more general form than needed here; we then give simpler consequences in Lemma 5.16.
Lemma 5.15 (Concentration of $\tilde{S}_n(t), \tilde{A}_n(t), \tilde{V}_n(t)$) Assume (A1)-(A4). Then there exists a constant $C$ such that, for any $u \geq 0$,
\[
\mathbb{E} \left[ \sup_{t \leq u} |\tilde{S}_n(t) - \mathbb{E} \tilde{S}_n(t)|^2 \right] \leq Cn \mathbb{E} (D_n^2(1 \wedge uD_n)), \tag{5.59}
\]
\[
\mathbb{E} \left[ \sup_{t \leq u} |\tilde{V}_n(t) - \mathbb{E} \tilde{V}_n(t)|^2 \right] \leq Cn \mathbb{E} (D_n^2(1 \wedge uD_n)), \tag{5.60}
\]
\[
\mathbb{E} \left[ \sup_{t \leq u} |\tilde{A}_n(t) - \mathbb{E} \tilde{A}_n(t)|^2 \right] \leq Cn \mathbb{E} (D_n^2(1 \wedge uD_n) + C. \tag{5.61}
\]

The final “$+ C$” in (5.61) is probably an artefact of our proof, but it is harmless for our purposes.

Proof The process $\tilde{V}_{n,k}(t)$ is a simple death process where each individual dies with rate $k$; it follows that $\tilde{V}_{n,k}(t)$ is a semimartingale with drift $-k\tilde{V}_{n,k}(t)$. Consequently, $\tilde{S}_n(t) = \sum_{k=0}^{\infty} k\tilde{V}_{n,k}(t)$ is a semimartingale with drift $-\sum_{k=0}^{\infty} k^2\tilde{V}_{n,k}(t)$, and $\tilde{S}_n(t) - \mathbb{E} \tilde{S}_n(t)$ is a semimartingale with drift $\tilde{\xi}(t) := -\sum_{k=0}^{\infty} k^2(\tilde{V}_{n,k}(t) - \mathbb{E} \tilde{V}_{n,k}(t))$.

We have, noting that $\tilde{V}_{n,k}(t)$ are independent and $\tilde{V}_{n,k}(t) \sim \text{Bin}(n_k, e^{-kt})$ for each $k$,
\[
\mathbb{E} [\tilde{S}_n(t) - \mathbb{E} \tilde{S}_n(t)]^2 = \sum_{k=0}^{\infty} \text{Var} (k\tilde{V}_{n,k}(t)) = \sum_{k=0}^{\infty} k^2 \text{Var} (\tilde{V}_{n,k}(t))
\]
\[
= \sum_{k=0}^{\infty} k^2 n_k e^{-kt}(1 - e^{-kt}) \leq \sum_{k=0}^{\infty} n_k k^2 (kt \wedge (kt)^{-1}). \tag{5.62}
\]

Similarly
\[
\mathbb{E} [\tilde{\xi}(t)]^2 = \sum_{k=0}^{\infty} \text{Var} (k^2\tilde{V}_{n,k}(t)) = \sum_{k=0}^{\infty} k^4 \text{Var} (\tilde{V}_{n,k}(t)) = \sum_{k=0}^{\infty} k^4 n_k e^{-kt}(1 - e^{-kt})
\]
\[
\leq \sum_{k=0}^{\infty} n_k k^4 e^{-kt}(1 \wedge kt). \tag{5.63}
\]

Hence, for some constant $C_1$,
\[
\sum_{j=0}^{\infty} \mathbb{E} [\tilde{S}_n(2^{-j}u) - \mathbb{E} \tilde{S}_n(2^{-j}u)]^2 \leq \sum_{k=0}^{\infty} n_k k^2 \sum_{j=0}^{\infty} (2^{-j}ku \wedge (2^{-j}ku)^{-1}). \tag{5.64}
\]
\[
\leq C_1 \sum_{k=0}^{\infty} n_k k^2 (ku \wedge 1)
\]

and
\[
\int_0^u t \mathbb{E} [\tilde{\xi}(t)]^2 \, dt \leq \sum_{k=0}^{\infty} n_k k^4 \int_0^u e^{-kt} (t \wedge kt^2) \, dt \leq \sum_{k=0}^{\infty} n_k k^2 (1 \wedge (ku)^3).
\]

Consequently, Lemma 4.2 yields
\[
\mathbb{E} \left[ \sup_{t \leq u} |\tilde{S}_n(t) - \mathbb{E} \tilde{S}_n(t)|^2 \right] \leq C_2 \sum_{k=0}^{\infty} n_k k^2 (1 \wedge ku) + C_3 \sum_{k=0}^{\infty} n_k k^2 (1 \wedge (ku)^3) \tag{5.65}
\]
\[
\leq C_4 \sum_{k=0}^{\infty} n_k k^2 (1 \wedge ku) = C_4n \mathbb{E} (D_n^2(1 \wedge (uD_n))).
\]

This yields (5.59).
We obtain (5.60) similarly; the estimates are the same, but with smaller powers of $k$, which can only help us.

Moreover, by a similar argument (but without having to sum over $k$), or by [36, Lemma 6.1] (with a modification for $u > 1$),

$$
\mathbb{E} \left[ \sup_{t \leq u} |\tilde{L}_n(t) - \mathbb{E} \tilde{L}_n(t)|^2 \right] \leq C_5 n(u \wedge 1),
$$

(5.66)

and thus, by (5.49),

$$
\mathbb{E} \left[ \sup_{t \leq u} |L_n(t) - \mathbb{E} L_n(t)|^2 \right] \leq C_6 n(u \wedge 1) + C_7.
$$

(5.67)

By definition, $\tilde{A}_n(t) = \tilde{L}_n(t) - \tilde{S}_n(t)$, and thus (5.61) follows by combining (5.67) and (5.59), noting that $\mathbb{E} D_n^2(uD_n \wedge 1) \geq \mathbb{P}(D_n = 1)(u \wedge 1)$ and $\mathbb{P}(D_n = 1) \to \mathbb{P}(D = 1) > 0$ by Remark 2.4. \hfill \blacksquare

**Lemma 5.16** (Concentration of $\tilde{S}_n(t), \tilde{A}_n(t), \tilde{V}_n(t)$) Assume (A1)–(A4). Let, as above, $\beta_n = \alpha_n = -\log(1 - \rho_n)$, as in (5.18). Then, for any fixed $t_0$,

$$
\sup_{t \leq t_0} \left| \tilde{S}_n(\beta_n t) - \mathbb{E} \tilde{S}_n(\beta_n t) \right| = O_p((n\epsilon_n)^{1/2}),
$$

(5.68)

$$
\sup_{t \leq t_0} \left| \tilde{V}_n(\beta_n t) - \mathbb{E} \tilde{V}_n(\beta_n t) \right| = O_p((n\epsilon_n)^{1/2}),
$$

(5.69)

$$
\sup_{t \leq t_0} \left| \tilde{A}_n(\beta_n t) - \mathbb{E} \tilde{A}_n(\beta_n t) \right| = O_p((n\epsilon_n)^{1/2} + 1).
$$

(5.70)

**Proof** Taking $u = \alpha_n t_0$, we obtain by (4.3),

$$
\mathbb{E} (D_n^2(1 \wedge uD_n)) \leq (1 \vee t_0) \mathbb{E} (D_n^2(1 \wedge \alpha_n D_n)) = O(\epsilon_n).
$$

(5.71)

Thus the right-hand sides of (5.59) and (5.60) and (5.61) are $O(n\epsilon_n)$ and $O(n\epsilon_n + 1)$, respectively; hence (5.68)–(5.70) follow using Markov's inequality. \hfill \blacksquare

The final three lemmas provide further estimates of the quantities $\beta_n = \alpha_n$ and $\gamma_n$ as set in (5.18) and (5.20).

**Lemma 5.17** Assume (A1)–(A4). If $\alpha_n \Delta_n = O(1)$, then

$$
\rho_n \sim \alpha_n \geq \frac{\epsilon_n}{\mathbb{E} D_n^2}.
$$

(5.72)

$$
\gamma_n \sim \alpha_n \epsilon_n \geq \frac{\epsilon_n^2}{\mathbb{E} D_n^3}.
$$

(5.73)

**Proof** We have $\alpha_n D_n \leq \alpha_n \Delta_n = O(1)$, and thus

$$
(1 \wedge \alpha_n D_n) \geq \alpha_n D_n.
$$

(5.74)

Hence (4.3) implies

$$
\epsilon_n \geq \mathbb{E}(\alpha_n D_n^2)
$$

(5.75)

and (5.72) follows, recalling (4.2).
Furthermore, (5.74) and (5.20) yield, using (5.75),
\[
\gamma_n \asymp \mathbb{E}(\alpha_n^2 D_n^3) = \alpha_n \mathbb{E}(\alpha_n D_n^3) \asymp \alpha_n \epsilon_n,
\]
showing (5.73).

Lemma 5.18  Assume (A1)-(A4) and $\epsilon_n > 0$. If
\[
(n\epsilon_n)^{1/2} = o(n\gamma_n),
\]
then (B8) holds, that is,
\[
\Delta_n = o(n\gamma_n).
\]

Proof  Suppose first that $\alpha_n \Delta_n \leq 1$. Then, using (5.73),
\[
\frac{\Delta_n}{n\gamma_n} \leq \frac{1}{n\gamma_n \alpha_n} = O\left(\frac{\epsilon_n}{n\gamma_n^2}\right) = O\left(\frac{n\epsilon_n}{(n\gamma_n)^2}\right),
\]
and thus (5.78) follows from (5.77) in this case.

Suppose next that $\alpha_n \Delta_n \geq 1$. Since $\mathbb{P}(D_n = \Delta_n) \geq 1/n$, we have by (4.3)
\[
\epsilon_n \asymp \mathbb{E}(D_n^2(1 \wedge (\alpha_n D_n))) \geq \frac{1}{n} \Delta_n^2(1 \wedge (\alpha_n \Delta_n)) = \frac{\Delta_n^2}{n}.
\]
Consequently, $\Delta_n = O\left((n\epsilon_n)^{1/2}\right)$, and thus (5.77) implies (5.78) in this case too.

Lemma 5.19  Assume (A1)-(A4) and $\epsilon_n > 0$. Then
\[
\epsilon_n^2 = O(\gamma_n \mathbb{E} D_n^3).
\]

Proof  The Cauchy-Schwarz inequality yields, using (5.20),
\[
\left(\mathbb{E}(D_n^2(1 \wedge \alpha_n D_n))\right)^2 \leq \mathbb{E}(D_n(1 \wedge \alpha_n D_n)^2) \mathbb{E}(D_n^3) = O(\gamma_n \mathbb{E} D_n^3).
\]
Hence the result follows by (4.3).

Proof of Theorem 2.6  First note that (B1)-(B4) hold for the parameter values in (5.17)-(5.21) by Lemma 5.10.

Next, by Lemma 5.19,
\[
\frac{n\epsilon_n}{(n\gamma_n)^2} = \frac{\epsilon_n^4}{\gamma_n^2 n\epsilon_n^3} = O\left(\frac{\mathbb{E} D_n^3)^2}{n\epsilon_n^3}\right),
\]
which is $o(1)$ by the assumption. Hence (5.77) holds. Consequently, Lemma 5.18 shows that (5.78) holds. In other words, (B8) holds.

Since $\Delta_n \geq 1$, (5.78) implies
\[
n\gamma_n \to \infty,
\]
and thus Lemma 5.12 applies and shows (5.42)–(5.44).
Moreover, (5.77) and (5.84) imply that the right-hand sides of (5.68)-(5.70) are \( o_p(n\gamma_n) \). Furthermore, \( \gamma_n = O(a_n) \) by (B2), see (5.24). Hence Lemmas 5.12 and 5.16 yield (B5), (B6) and (B7).

We have verified (B1)-(B8), so Theorem 5.4 applies and the result follows, recalling (5.17), (5.19), (2.10) and (4.2). Note that \( \hat{g}(\tau) = \hat{h}(\tau)/2 \), so the asymptotics for \( v(C_1) \) and \( e(C_1) \) are the same.

**Proof of Theorem 2.8** By assumption, \( E D_n^3 = O(1) \) and \( \epsilon_n n^{1/3} \to \infty \), so Theorem 2.6 applies; thus (2.19) holds. Furthermore, as said in Section 4.1, Theorem 3.1(ii) applies with \( X_n = \tilde{D}_n \) and yields (4.9), which together with (2.10) yields the first equality in (2.23); the second equality then follows by (2.6). Similarly, (2.20) and (4.9) (or (3.2)) yield (2.24).

**Proof of Theorem 2.9** Again, Theorem 2.6 applies. Moreover, by (2.15), we have \( E \tilde{D}_n^2 = E(D(D - 1)^2) / E D = \infty \), and so Theorem 3.1(iii) applies, yielding \( \rho_n = o(\epsilon_n) \).

**Proof of Theorem 2.10** Theorem 2.6 applies.
(i): Follows from (2.19), (3.3) for \( X_n = \tilde{D}_n \) and (2.26).
(ii): Now, by (2.15), \( E \tilde{D}_n^2 \leq E D_n^2 / E D_n = O(1) \). Hence Theorem 3.1(i) applies and yields \( \rho_n \asymp \epsilon_n \); consequently (2.19) implies (2.30).
(iii): By (2.28) and (2.26), \( E D_n^2 = O(\kappa_n) = O(\{\tilde{D}_n(\tilde{D}_n - 1)\}) = O(\{\tilde{D}_n^2\}) \). Thus, the assumption implies \( \epsilon_n \Delta_n = o(\{\tilde{D}_n^2\}) \). Hence, Theorem 3.1(iv) applies and (2.31) follows by (2.26) and (2.28).

### 6 | THE CRITICAL CASE

We define, for convenience and for comparison with Hatami and Molloy [24],

\[
R_n := E D_n^3. \tag{6.1}
\]

The basic condition for the critical case in Theorem 2.12 is thus, as in [24],

\[
\epsilon_n = O(n^{-1/3} R_n^{2/3}). \tag{6.2}
\]

**Remark 6.1** Our \( R_n \) is not exactly the same as \( R \) defined by Hatami and Molloy [24], which equals our \( E D_n(D_n - 2)^2 / E D_n = \kappa_n - \epsilon_n \), see (2.26) and (2.6), but the two values are equivalent in the sense \( R_n \asymp R_{\text{HatamiMolloy}} \), see (2.28) and (2.27); hence the two values are equivalent for our purposes.

Note that, as said in Remark 2.13, \( R_n \geq \frac{1}{n} \Delta_n^3 \) and hence always

\[
\Delta_n \leq (nR_n)^{1/3}. \tag{6.3}
\]

Note also that in Theorem 2.12 we impose the slightly stronger condition (2.32), that is,

\[
\Delta_n = o((nR_n)^{1/3}). \tag{6.4}
\]

Furthermore, by (A2),

\[
R_n = E D_n^3 \leq \Delta_n E D_n^2 = O(\Delta_n). \tag{6.5}
\]
Hence, (6.4) implies $\Delta_n^3 = o(nR_n) = o(n\Delta_n)$ and thus $\Delta_n^2 = o(n)$ and

$$\Delta_n = o(n^{1/2}), \quad (6.6)$$

and thus also, by (6.5),

$$R_n = o(n^{1/2}). \quad (6.7)$$

In Theorem 2.12 we assume both (6.2) and (6.4), and it follows from (6.2) and (6.7) that $\varepsilon_n = o(1)$, so (A4) follows from the other conditions. (However, for emphasis we keep it in the statements in Theorem 2.12 and below.)

Note also that, using (5.45) and (2.10),

$$R_n \geq E\hat{D}_n^3 \to 2\mu > 0, \quad \text{so } R_n \text{ is bounded below and } R_n^{-1} = O(1).$$

We continue to work with the configuration model and the multigraph $G^*_n$ as in the preceding section. In Section 6.3 we give additional arguments for the graph case.

### 6.1 Proof of Theorem 2.12(i)

The idea is to use Theorem 2.6 for the supercritical case and a kind of monotonicity in $\varepsilon_n$; it is intuitively clear that a larger $\varepsilon_n$ ought to result in a larger largest component, and thus the supercritical case will provide an upper bound for the critical case. The formal details are as follows.

**Proof of Theorem 2.12(i)**  Let $\omega(n) \to \infty$ slowly, so slowly that, cf. (6.7) and (6.4),

$$\omega(n)R_n = o\left(n^{1/2}\right), \quad (6.8)$$

$$\omega(n)\Delta_n \leq (nR_n)^{1/3}. \quad (6.9)$$

Let $m_n := \lfloor n^{2/3}R_n^{2/3}o(n)^{2/3} \rfloor$. Change the degree sequence $(d_i)_{i \in [n]}$ to $(\hat{d}_i)_{i \in [n]}$ by replacing $2m_n$ vertices of degree 1 by $m_n$ vertices of degree 0 and $m_n$ vertices of degree 2. This is possible (at least for large $n$) because $n_1/n = \mathbb{P}(D_n = 1) \to \mathbb{P}(D = 1) > 0$, see Remark 2.4, and thus, using (6.8),

$$m_n \leq n^{2/3}(R_n o(n))^{2/3} = o(n) = o(n_1). \quad (6.10)$$

We denote the variables for the modified degree sequence by $\hat{D}_n$ and so on. Note that the modification does not change the sum of vertex degrees, so $E\hat{D}_n = E\hat{D}_n = \mu_n$, but it increases $E[D_n(D_n - 1)]$ by $2m_n/n \sim 2n^{-1/3}R_n^{2/3}o(n)^{2/3}$. Thus, using (6.2) and $\omega(n) \to \infty$,

$$\hat{\varepsilon}_n = \varepsilon_n + 2m_n/n \sim 2n^{-1/3}R_n^{2/3}o(n)^{2/3}. \quad (6.11)$$

Similarly, $R_n = E\hat{D}_n^3$ is increased to

$$\hat{R}_n = E\hat{D}_n^3 = R_n + \frac{6m_n}{n} = R_n + o(1) \sim R_n, \quad (6.12)$$

where we have used (6.10) to see that the difference is insignificant. Furthermore, it is easily seen that (A1)-(A4) still hold (with the same $D$), using (6.10) and (6.11) for (A1) and (A4).
Since (6.11) and (6.12) imply $\varepsilon_n \gg n^{-1/3} \hat{R}_n^{2/3}$, and (6.11) and (6.9) imply $\varepsilon_n \Delta_n = o(R_n)$, Theorem 2.10(iii) applies to the modified degree sequence and yields, w.h.p.,

$$v(\hat{C}_1) \leq C \frac{\hat{\varepsilon}_n^{1/2}}{R_n} = o\left(n^{2/3} R_n^{-1/3} \omega(n)\right).$$ (6.13)

In particular, w.h.p.

$$v(\hat{C}_1) \leq n^{2/3} R_n^{-1/3} \omega(n).$$ (6.14)

We can obtain $G^*(n, (\hat{d}_i)_{i\in[n]})$ from $G^*(n, (d_i)_{i\in[n]})$ by merging $m_n$ pairs of vertices of degree 1 into vertices of degree 2, and adding $m_n$ vertices of degree 0 to keep the total number of vertices. Any connected set $C$ of vertices in $G^*(n, (\hat{d}_i)_{i\in[n]})$ then corresponds to a connected set of at least $v(C)/2$ vertices in $G^*(n, (d_i)_{i\in[n]})$. Consequently, $v(\hat{C}_1) \geq \frac{1}{2} v(C_1)$ and thus (6.14) implies, w.h.p.,

$$v(C_1) \leq 2v(\hat{C}_1) \leq 2n^{2/3} R_n^{-1/3} \omega(n).$$ (6.15)

Since $\omega(n) \to \infty$ arbitrarily slowly, (6.15) implies $v(C_1) = O_p(n^{2/3} R_n^{-1/3})$. (If not, we could find $\delta > 0$ and $K = K(n) \to \infty$ such that, at least along a subsequence, $\mathbb{P}(v(C_1) \geq K(n)n^{2/3} R_n^{-1/3}) \geq \delta$. We choose $\omega(n)$ with $\omega(n) \leq K(n)/2$ to obtain a contradiction. See also [32].) This completes our proof of (2.33).

\begin{remark}
In our proof we needed only the simple, deterministic bound $v(C_1) \leq 2v(\hat{C}_1)$. Actually, when Theorem 2.12(i) is proved, it implies together with Theorem 2.10(iii) that w.h.p. $v(C_1) \ll v(\hat{C}_1)$, that is, that the giant component $\hat{C}_1$ for the modified sequence w.h.p. is much larger than $C_1$ for the original sequence; the reason is that, in the merging described above, the giant component typically absorbs many small components.
\end{remark}

\begin{example}
Consider a critical example with $\varepsilon_n = O(n^{-1/3})$, $R_n = O(1)$ and $\Delta_n = o(n^{1/3})$. For example (as in [24]), we can let $3/4$ of all vertices have degree 1 and the rest degree 3. Alternatively, we can take the Erdős-Rényi graph $G(n, 1/n)$ and condition on the degree sequence, as described for general rank-1 inhomogeneous random graphs in Section 2.6. Then $v(C_1)$ is typically of order $n^{2/3}$, see [24, Theorem 2.12].

Let $m_n$ be integers with $n^{1/3} \ll m_n \ll n^{1/2}$. Modify the degree sequence $(d_i)_{i\in[n]}$ to $(\hat{d}_i)_{i\in[n]}$ by merging $m_n$ vertices of degree 1 to a single vertex of degree $m_n$, and adding $m_n - 1$ vertices of degree 0. Then it is easily seen that $\hat{\varepsilon}_n \approx m_n^2/n$, $\hat{R}_n \sim m_n^2/n$ and $\hat{\Delta}_n = m_n$. Thus (6.2) holds for the modified sequence but not (6.4). Furthermore,

$$v(\hat{C}_1) \geq v(C_1) - m_n$$ (6.16)

so $v(C_1)$ is typically also of order (at least) $n^{2/3}$. Hence, (2.33) fails.
\end{example}

\section{Proof of Theorem 2.12(ii) in the multigraph case}
In this section, we consider only the multigraph case. Unlike all other results in this paper, the graph case does not follow immediately by conditioning. We treat the graph case in the next section.
We use the cluster exploration process and notation from Section 5.1. Let
\[ t_1 := (nR_n)^{-1/3}, \]  
(6.17)
and note that \( t_1 = O(n^{-1/3}) = o(1) \) and, by (6.7), \( t_1 \gg n^{-1/2} \) and thus \( nt_1 \to \infty \) and \( nt_1^2 \to \infty \). Furthermore, let
\[ \sigma_n^2 := \text{Var} \tilde{S}_n(t_1). \]  
(6.18)

**Lemma 6.4** Assume (A1)-(A4) and (6.4).

(i) Then
\[ \sigma_n^2 \sim (nR_n)^{2/3}. \]  
(6.19)
Moreover, \( \tilde{S}_n(t_1) \) is asymptotically normal:
\[ \left( \tilde{S}_n(t_1) - \mathbb{E} \tilde{S}_n(t_1) \right) / \sigma_n \xrightarrow{d} N(0, 1). \]  
(6.20)

(ii) Let \( \sigma_{L,n}^2 := 4nt_1 \mu_n \). Then \( L_n(t_1) \) is asymptotically normal, with
\[ \left( L_n(t_1) - \mathbb{E} L_n(t_1) \right) / \sigma_{L,n} \xrightarrow{d} N(0, 1). \]  
(6.21)
Furthermore, \( \limsup \sigma_{L,n}^2 / \sigma_n^2 < 1 \).

(iii) For any \( b > 0 \), there exists \( c(b) > 0 \) such that
\[ \mathbb{P} \left( A_n(t_1) - \mathbb{E} A_n(t_1) > b \sigma_n \right) \geq c(b) + o(1). \]  
(6.22)

**Proof**

(i): We have, see Section 5.1 and in particular (5.2), \( \tilde{S}_n(t) = \sum_{i \in [n]} d_i I_i(t) \), where \( I_i(t) \) is the indicator that no half-edge at vertex \( i \) has died spontaneously up to time \( t \). These indicators are independent and \( I_i(t) \sim \text{Be}(e^{-d_i t}) \). Hence, as in (5.62) but written slightly differently, noting that \( t_1 d_i \leq t_1 \Delta_n = o(1) \) by (6.17) and (6.4),
\[ \text{Var} \tilde{S}_n(t_1) = \sum_{i \in [n]} d_i^2 \text{Var} I_i(t_1) = \sum_{i \in [n]} d_i^2 e^{-d_i t_1} (1 - e^{-d_i t_1}) \sim \sum_{i \in [n]} d_i^3 t_1 = t_1 nR_n = (nR_n)^{2/3}, \]
which is (6.19). Similarly, with \( Y_i := d_i I_i(t_1) \) and using (6.4),
\[ \sum_{i \in [n]} \mathbb{E} |Y_i - \mathbb{E} Y_i|^3 = \sum_{i \in [n]} d_i^3 \mathbb{E} |I_i(t_1) - \mathbb{E} I_i(t_1)|^3 \leq \sum_{i \in [n]} d_i^3 \text{Var} I_i(t_1) \leq \sum_{i \in [n]} d_i^4 t_1 \]
\[ = t_1 n \mathbb{E} D_i^4 \leq t_1 n \Delta_n R_n = o(nR_n) = o(\sigma_n^2). \]  
(6.23)

Consequently, the central limit theorem with Lyapounov’s condition [23, Theorem 7.2.2] applies and yields (6.20).

(ii): We use the modified process \( \tilde{L}_n(t) \) defined just before (5.50). Then \( \frac{1}{2} \tilde{L}_n(t) \sim \text{Bin} \left( \frac{1}{2} \ell_n, e^{-2t} \right) \) for every \( t \geq 0 \). In particular, recalling from (2.2) and (2.4) that \( \ell_n = n \mu_n \),
\[ \text{Var} \tilde{L}_n(t_1) = 4 \cdot \frac{1}{2} \ell_n e^{-2t_1} (1 - e^{-2t_1}) \sim 4 \ell_n t_1 = \sigma_{L,n}^2. \]  
(6.24)
Since \( nt_1 \to \infty \), we have \( \sigma_{L,n}^2 \to \infty \), and the central limit theorem for the binomial distribution yields \((\bar{L}_n(t_1) - \mathbb{E}\bar{L}_n(t_1)) / \sigma_{L,n} d \to N(0, 1)\). Since \(|L_n(t_1) - \bar{L}_n(t_1)| \leq 1\) by (5.49), (6.21) follows. Furthermore,

\[
\frac{\sigma_n^2}{\sigma_{L,n}^2} = \frac{\sigma_n^2}{4nt_1 \mu_n} \sim \frac{(nR_n)^{2/3}}{4n(nR_n)^{-1/3} \mu_n} = \frac{R_n}{4n} = \frac{\mathbb{E} D_n^3}{4 \mathbb{E} D_n}. \tag{6.25}
\]

Consequently, using (5.30) (with \( t = 0 \)),

\[
\lim\inf_{n \to \infty} \frac{\sigma_n^2}{\sigma_{L,n}^2} = \lim\inf_{n \to \infty} \frac{\mathbb{E} D_n^3}{4 \mathbb{E} D} \geq \frac{\mathbb{E} (D(D - 2)^2)}{4 \mu} > 1. \tag{6.26}
\]

(iii): By (ii), there exists \( \delta > 0 \) such that, for large \( n \), \( \sigma_{L,n} < (1 - 2\delta) \sigma_n \). Let \( a := \delta^{-1} b \) and let \( \Phi \) be the usual standard normal distribution function. Then, by (6.20) and (6.21),

\[
\mathbb{P}(\bar{S}_n(t_1) - \mathbb{E}\bar{S}_n(t_1) < -a \sigma_n) \to \Phi(-a), \tag{6.27}
\]

\[
\mathbb{P}(L_n(t_1) - \mathbb{E} L_n(t_1) < -(1 + \delta) a \sigma_{L,n}) \to \Phi(-(1 + \delta) a). \tag{6.28}
\]

Hence, with probability at least \( c + o(1) \), where \( c := \Phi(-a) - \Phi(-(1 + \delta) a) > 0 \), we have \( \bar{S}_n(t_1) - \mathbb{E}\bar{S}_n(t_1) < -a \sigma_n \) and \( L_n(t_1) - \mathbb{E} L_n(t_1) \geq -(1 + \delta) a \sigma_{L,n} \), and thus, recalling (5.3),

\[
\tilde{A}_n(t_1) - \mathbb{E}\tilde{A}_n(t_1) > a \sigma_n - (1 + \delta) a \sigma_{L,n} > a \sigma_n - (1 + \delta)(1 - 2\delta) a \sigma_n > \delta a \sigma_n = b \sigma_n. \tag{6.29}
\]

**Remark 6.5** Presumably, \( \bar{S}_n(t_1) \) and \( L_n(t_1) \) are asymptotically jointly normal, which would imply that \( \tilde{A}_n(t_1) \) is asymptotically normal and yield a more direct proof of (6.22). However, it seems more technical to prove joint asymptotic normality here, so instead we prefer the more elementary argument above.

**Lemma 6.6** Assume (A1)-(A4) and (6.2) and (6.4). Then, uniformly for \( t \leq t_1 \),

\[
\mathbb{E}\bar{S}_n(t) = n \mu_n - 2m \mu_n + O(\sigma_n), \tag{6.30}
\]

\[
\mathbb{E} L_n(t) = n \mu_n e^{-\nu t} + O(1) = n \mu_n - 2m \mu_n + O(\sigma_n), \tag{6.31}
\]

\[
\mathbb{E}\tilde{A}_n(t) = O(\sigma_n). \tag{6.32}
\]

**Proof** Similarly to the proof of Lemma 5.12, \( V_{n,k}(t) \sim \text{Bin}(n_k, e^{-\nu t}) \) and thus, using (5.45),

\[
\mathbb{E}\bar{S}_n(t) = \sum_{k=0}^{\infty} k \mathbb{E} V_{n,k}(t) = \sum_{k=0}^{\infty} k n_k e^{-\nu t} = \sum_{k=0}^{\infty} kn_k \left(1 - k \mu + \mathcal{O}(k^2 \mu^2)\right)
= n \mathbb{E} D_n - tn \mathbb{E} D_n^2 + O(\nu^2 \mathbb{E} D_n^3)
= n \mu_n - tm \mu_n (2 + \varepsilon_n) + O(\nu^2 n R_n), \tag{6.33}
\]

which yields (6.30) by (6.2), (6.17) and (6.19).

Furthermore, by (5.50),

\[
\mathbb{E} L_n(t) = n \mu_n e^{-\nu t} + O(1) = n \mu_n - 2m \mu_n + O(n \mu_n^2 + 1), \tag{6.34}
\]
and (6.31) follows because, by (6.17) and (6.19),
\[ nt_1^2 + 1 \sim nt_1^2 = n^{1/3}R_n^{-2/3} \sim \sigma_nR_n^{-1} = O(\sigma_n). \]  
(6.35)

Finally, (6.32) follows from (6.30) and (6.31).

\[ \square \]

**Lemma 6.7**  
Assume (A1)-(A4) and (6.2) and (6.4). Then,
\[ \mathbb{E} \left[ \sup_{t \leq t_1} |\tilde{A}_n(t)|^2 \right] = O(\sigma_n^2). \]  
(6.36)

**Proof**  
By Lemma 5.15, together with (6.17) and (6.19),
\[ \mathbb{E} \left[ \sup_{t \leq t_1} |\tilde{A}_n(t) - \mathbb{E} \tilde{A}_n(t)|^2 \right] \leq Cn \mathbb{E} \left[ D_n^2(1 \wedge t_1D_n) \right] + C \]
\[ \leq Cnt_1 \mathbb{E} \left[ D_n^2 \right] + C = Cnt_1R_n + C = O(\sigma_n^2). \]  
(6.37)

Furthermore, \( \sup_{t \leq t_1} |\mathbb{E} \tilde{A}_n(t)| = O(\sigma_n) \) by (6.32), and (6.36) follows.

\[ \square \]

For ease of notation, let \( N_k := \tilde{V}_{nk}(t_1) \), the (random) number of vertices of degree \( k \) such that none of their half-edges dies spontaneously by time \( t_1 \). Thus \( \tilde{S}_n(t_1) = \sum_k kN_k \), see (5.2). Let further
\[ Z_n := \sum_{k=0}^{\infty} k^2(n_k - N_k) \geq 0. \]  
(6.38)

**Lemma 6.8**  
Assume (A1)-(A4) and (6.2) and (6.4). Then, there exists a constant \( C_8 \) such that w.h.p.
\[ Z_n \leq C_8\sigma_n^2. \]  
(6.39)

**Proof**  
\( N_k \sim \text{Bin}(n_k, e^{-kt_1}) \) and thus, using (6.17) and (6.19),
\[ \mathbb{E} Z_n = \sum_{k=0}^{\infty} k^2 n_k \left( 1 - e^{-kt_1} \right) \leq \sum_{k=0}^{\infty} k^3 n_k t_1 = t_1 nR_n = O(\sigma_n^2). \]  
(6.40)

Furthermore, using also (6.4),
\[ \text{Var} Z_n = \sum_{k=0}^{\infty} k^4 \text{Var} N_k \leq \sum_{k=0}^{\infty} k^4 n_k \left( 1 - e^{-kt_1} \right) \leq \sum_{k=0}^{\infty} k^5 n_k t_1 \]
\[ = t_1 n \mathbb{E} D_n^2 \leq t_1 n \Delta_n^2 R_n = o\left( (nR_n)^{4/3} \right) = o(\sigma_n^3). \]  
(6.41)

Now (6.39) follows by (6.40) and (6.41) and Chebyshev’s inequality.

\[ \square \]

We condition on \( \mathcal{F}_{t_1} \), the \( \sigma \)-field generated by all events up to time \( t_1 \). Note that \( \mathcal{F}_{t_1} \) determines \( N_k \), and thus \( \tilde{S}_n(t_1) \) and \( Z_n \), and also \( L_n(t_1) \) and \( \tilde{A}_n(t_1) \).

**Lemma 6.9**  
Assume (A1)-(A4) and (6.2) and (6.4). For any fixed \( B < \infty \) and all \( t \in [0, Bt_1] \),
\[ \mathbb{E} \left( \tilde{S}_n(t_1 + t) \mid \mathcal{F}_{t_1} \right) = \tilde{S}_n(t_1) - 2tn\mu_n + tZ_n + O(\sigma_n), \]  
(6.42)
\[ \mathbb{E} \left( L_n(t_1 + t) \mid \mathcal{F}_{t_1} \right) \geq L_n(t_1) - 2tn\mu_n + O(\sigma_n), \]  
(6.43)
\[ \mathbb{E} \left( \tilde{A}_n(t_1 + t) \mid \mathcal{F}_{t_1} \right) \geq \tilde{A}_n(t_1) - tZ_n + O(\sigma_n). \]  
(6.44)
Proof. We have, in analogy with (6.33), using (6.38),

\[
\mathbb{E}(\tilde{S}_n(t_1 + t) \mid F_{t_1}) = \sum_{k=0}^{\infty} kN_k e^{-kt} = \sum_{k=0}^{\infty} kN_k (1 - kt + O(k^2 t^2))
\]

\[
= \tilde{S}_n(t_1) - t \left( \sum_{k=0}^{\infty} k^2 n_k - Z_n \right) + O(t^2 \sum_{k=0}^{\infty} k^3 n_k)
\]

\[
= \tilde{S}_n(t_1) - tn E D_n^2 + tZ_n + O(t^2 nR_n).
\]

Then (6.42) follows by (5.45) and estimates as in the proof of Lemma 6.6, using (6.2), (6.17), (6.19) and the assumption \( t = O(t_1) \).

For \( L_n \) we use again the coupling with \( \bar{L}_n \). As \( \frac{1}{2} \bar{L}_n(t) \) is a standard death process with intensity 2,

\[
\mathbb{E}(L_n(t_1 + t) \mid F_{t_1}) = \mathbb{E}(\bar{L}_n(t_1 + t) \mid F_{t_1}) + O(1) = \bar{L}_n(t_1)e^{-2t} + O(1)
\]

\[
= L_n(t_1) - 2nL_n(t_1) + O(1 + n^2).
\]

Then (6.43) follows, since \( L_n(t_1) < \xi_n = n\mu_n \), using again (6.35).

Finally, (6.44) follows from (6.42) and (6.43).

\[\square\]

**Lemma 6.10.** Assume (A1)-(A4). For any fixed \( B < \infty \) and all \( t \in [0, Bt_1] \),

\[
\mathbb{E}\left[ \sup_{t \leq Bt_1} |\tilde{S}_n(t_1 + t) - \mathbb{E}(\tilde{S}_n(t_1 + t) \mid F_{t_1})|^2 \mid F_{t_1} \right] = O(\sigma_n^2)\tag{6.47}
\]

\[
\mathbb{E}\left[ \sup_{t \leq Bt_1} |L_n(t_1 + t) - \mathbb{E}(L_n(t_1 + t) \mid F_{t_1})|^2 \mid F_{t_1} \right] = O(\sigma_n^2)\tag{6.48}
\]

\[
\mathbb{E}\left[ \sup_{t \leq Bt_1} |\tilde{A}_n(t_1 + t) - \mathbb{E}(\tilde{A}_n(t_1 + t) \mid F_{t_1})|^2 \mid F_{t_1} \right] = O(\sigma_n^2)\tag{6.49}
\]

Proof. Conditioned on \( F_{t_1} \), the process \( \tilde{S}_n(t_1 + t) \) is exactly as \( \tilde{S}_n(t) \), but starting with \( N_k \) vertices of degree \( k \) instead of \( n_k \). Hence the arguments in (5.62)-(5.65) in the proof of Lemma 5.15 hold in this case too and, since \( N_k \leq n_k \), we obtain, for any \( u \geq 0 \),

\[
\mathbb{E}\left[ \sup_{t \leq u} |\tilde{S}_n(t_1 + t) - \mathbb{E}(\tilde{S}_n(t_1 + t) \mid F_{t_1})|^2 \mid F_{t_1} \right] \leq C_4 n \mathbb{E}(D_n^2 uD_n \wedge 1) \leq C_4 nu \mathbb{E}(D_n^3).
\]

The result (6.47) follows by taking \( u = Bt_1 \), using again (6.1), (6.17) and (6.19).

Similarly, as in (5.67), or by [36, Lemma 6.1] after conditioning on \( F_{t_1} \), we obtain, since \( t_1 = o(1) \),

\[
\mathbb{E}\left[ \sup_{t \leq Bt_1} |L_n(t_1 + t) - \mathbb{E}(L_n(t_1 + t) \mid F_{t_1})|^2 \mid F_{t_1} \right] = O(nt_1 + 1)\tag{6.50}
\]

Furthermore, as said above, \( nt_1 \to \infty \) and \( R_n^{-1} = O(1) \), and thus, cf. (6.35),

\[
nt_1 + 1 \asymp nt_1 = n^{2/3} R_n^{-1/3} \asymp \sigma_n^2 R_n^{-1} = O(\sigma_n^2)\tag{6.51}
\]

Hence, (6.50) yields (6.48). Finally, (6.49) follows by combining (6.47) and (6.48).

\[\square\]
Lemma 6.11 Assume (A1)-(A4) and (6.2) and (6.4). For any fixed $B > 1$, there is some $p(B) > 0$ such that with probability at least $p(B) + o(1),$

$$\tilde{A}_n(t) > 0 \quad \text{for all } t \in [t_1, Bt_1]. \quad (6.52)$$

Proof Fix $B > 1$. Let $b > 0$ be another fixed number, to be determined later. Consider the event

$$\mathcal{E}(b) := \{\tilde{A}_n(t_1) - \mathbb{E}\tilde{A}_n(t_1) > b\sigma_n \text{ and } Z_n \leq C_8\sigma_n^2\}, \quad (6.53)$$

with $C_8$ as in Lemma 6.8. By (6.22) and (6.39), $\mathbb{P}(\mathcal{E}(b)) \geq c(b) + o(1)$, where $c(b) > 0$ is independent of $n$. Define also the family of events $\{\mathcal{E}_i(C) : C > 0\}$, with $\mathcal{E}_i(C)$ given by

$$\mathcal{E}_i(C) := \left\{\sup_{t \leq Bt_1}|\tilde{A}_n(t_1 + t) - \mathbb{E}(\tilde{A}_n(t_1 + t) | \mathcal{F}_{t_1})| \leq C\sigma_n\right\}. \quad (6.54)$$

Furthermore, let

$$\mathcal{E}(b, C) := \mathcal{E}(b) \cap \mathcal{E}_1(C). \quad (6.55)$$

Note that $\mathcal{E}(b) \in \mathcal{F}_{t_1}$. Hence, by Lemma 6.10 and Chebyshev’s inequality, there exists a constant $C_9$ such that

$$\mathbb{P}(\mathcal{E}_1(C) | \mathcal{E}(b)) \geq 1 - \frac{C_9\sigma_n^2}{(C\sigma_n)^2} = 1 - \frac{C_9}{C^2}. \quad (6.56)$$

Consequently, if we choose $C := 2C_9^{1/2}$, then

$$\mathbb{P}(\mathcal{E}(b, C)) = \mathbb{P}(\mathcal{E}_1(C) | \mathcal{E}(b)) \mathbb{P}(\mathcal{E}(b)) \geq \frac{3}{4} \mathbb{P}(\mathcal{E}(b)) \geq \frac{3}{4} c(b) + o(1). \quad (6.57)$$

On the event $\mathcal{E}(b, C)$, we have by (6.54), (6.44), (6.53), (6.32) and (6.17), for any $t \in [0, Bt_1],$

$$\tilde{A}_n(t_1 + t) \geq \mathbb{E}(\tilde{A}_n(t_1 + t) | \mathcal{F}_{t_1}) - C\sigma_n \geq \tilde{A}_n(t_1) - tZ_n + O(\sigma_n)$$

$$> b\sigma_n + \mathbb{E}\tilde{A}_n(t_1) - C_8\sigma_n^2 + O(\sigma_n)$$

$$= b\sigma_n + O(\sigma_n). \quad (6.58)$$

The implicit constants here depends on $B$ but not on $b$; thus the final error term $O(\sigma_n) \geq -C_{10}(B)\sigma_n$ for some $C_{10}(B)$. Hence we may for any $B$ choose $b = b(B) := C_{10}(B)$, and the result follows, with $p(B) = \frac{3}{4} c(b(B))$.

We can obtain results for $\tilde{V}_n$ similar to the results for $\tilde{S}_n$ above (in Lemmas 6.4, 6.6, 6.9, and 6.10) by the same arguments. However, we have no need for such results involving conditioning and uniform estimates; the following simple results are enough.

Lemma 6.12 Assume (A1)-(A4) and (6.4). Fix $B > 0$. For any $t \in [0, Bt_1],$

$$\tilde{V}_n(t) = n - n\mu_n t + O_p\left(nt_1^2 + \sqrt{nt_1}\right) = n - n\mu_n t + o_p(nt_1). \quad (6.59)$$
Proof. Recall that \( \widetilde{V}_n(t) = \sum k \widetilde{V}_{n,k}(t) \) where \( \widetilde{V}_{n,k}(t) \) are independent and \( \widetilde{V}_{n,k}(t) \sim \text{Bin}(n_k, e^{-k t}) \). Hence,

\[
\mathbb{E} \widetilde{V}_n(t) = \sum_{k=0}^{\infty} n_k e^{-k t} = \sum_{k=0}^{\infty} n_k \left( 1 - k t + O(k^2 t^2) \right) = n - n \mu_n t + O(n t^2)
\]

(6.60)

and

\[
\text{Var} \widetilde{V}_n(t) = \sum_{k=0}^{\infty} n_k e^{-k t} (1 - e^{-k t}) \leq \sum_{k=0}^{\infty} n_k k t = n \mu_n t = O(nt).
\]

(6.61)

The first equality in (6.59) follows from (6.60)-(6.61). The second follows because \( n t_1^2 = o(nt_1) \) and \( \sqrt{nt_1} = o(nt_1) \).

Lemma 6.13 Assume (A1)-(A4) and (6.4), and define \( V_n'(t) := \widetilde{V}_n(t) - V_n(t) \geq 0 \). Fix \( B > 1 \). Then

\[
V_n'(t_1) - V_n'(B t_1) \leq O_p(t_1 \sigma_n) = o_p(nt_1).
\]

(6.62)

Proof. \( V_{n,k}'(t) := \widetilde{V}_{n,k}(t) - V_{n,k}(t) \) is the number of vertices of degree \( k \) that are awake at time \( t \), but their \( k \) half-edges all have maximal life times larger than \( t \). This number may increase when C1 is performed, and it decreases when a half-edge at one of these vertices dies spontaneously (and C3 is performed). Consequently, conditioning of \( \mathcal{F}_{t_1} \), for any \( t \geq 0 \),

\[
\mathbb{E} \left( (V_{n,k}'(t_1) - V_{n,k}'(t_1 + t))_+ \mid \mathcal{F}_{t_1} \right) \leq k t V_{n,k}'(t_1).
\]

Summing over \( k \) yields, using (5.2),

\[
\mathbb{E} \left( (V_n'(t_1) - V_n'(t_1 + t))_+ \mid \mathcal{F}_{t_1} \right) \leq t (\widetilde{S}_n(t_1) - S_n(t_1)).
\]

(6.63)

By (5.4) and Lemma 6.7, noting that \( \Delta_n = O(\sigma_n) \) by (6.3) and (6.19),

\[
\widetilde{S}_n(t_1) - S_n(t_1) < \sup_{t \leq t_1} |A_n(t)| + \Delta_n = O_p(\sigma_n).
\]

(6.64)

In other words, for every \( \varepsilon > 0 \) there exist \( K(\varepsilon) \) independent of \( n \) such that

\[
\mathbb{P}(\widetilde{S}_n(t_1) - S_n(t_1)) > K(\varepsilon) \sigma_n \leq \varepsilon.
\]

(6.65)

Furthermore, for any fixed \( K \), (6.63) implies

\[
\mathbb{E} \left( (V_n'(t_1) - V_n'(t_1 + t))_+ \mid \widetilde{S}_n(t_1) - S_n(t_1) \leq K \sigma_n \right) = O(t \sigma_n).
\]

(6.66)

It follows by (6.66), Markov’s inequality and (6.65) that, for any \( t > 0 \),

\[
(V_n'(t_1) - V_n'(t_1 + t))_+ = O_p(t \sigma_n).
\]

(6.67)

Now take \( t = (B - 1)t_1 \). ■
Proof of Theorem 2.12(ii) Note that the assumptions include (6.2) and (6.4). Recall also that $A_n(t) \geq \tilde{A}_n(t)$ for all $t$, see (5.4). Hence by Lemma 6.11, for every $B > 1$, there exists $p(B) > 0$ such that with probability at least $p(B) + o(1)$, $A_n(t) \geq \tilde{A}_n(t) > 0$ for all $t \in [t_1, Bt_1]$. By the discussion in Section 5.1, this means that $C1$ is not performed during the interval $[t_1, Bt_1]$ and thus all vertices awakened during this interval belong to the same component, say $C$. The number of these vertices is $V_n(t_1) - V_n(Bt_1)$. Consequently, with probability at least $p(B) + o(1)$,

$$v(C_1) \geq v(C) \geq V_n(t_1) - V_n(Bt_1). \quad (6.68)$$

Furthermore, by Lemmas 6.13 and 6.12,

$$V_n(t_1) - V_n(Bt_1) = \tilde{V}_n(t_1) - \tilde{V}_n(Bt_1) + V_n^*(t_1) - V_n^*(t_1) \geq \tilde{V}_n(t_1) - \tilde{V}_n(Bt_1) + o_p(nt_1)$$

$$= n\mu(B-1)t_1 + o_p(nt_1) = n\mu(B-1)t_1 + o_p(nt_1). \quad (6.69)$$

Hence, $V_n(t_1) - V_n(Bt_1) > (\mu(B-1) - 1)nt_1$ w.h.p.

Finally, given any $K > 0$, choose $B$ such that $\mu(B-1) = K + 1$. Then (6.68) and (6.69) thus show that, with probability at least $p(B) + o(1)$, recalling (6.17),

$$v(C_1) \geq V_n(t_1) - V_n(Bt_1) > Kn_1 = Kn^{2/3}p_n^{-1/3}, \quad (6.70)$$

which completes the proof of (2.34).

6.3 Proof of Theorem 2.12(ii) in the graph case

Unlike the other results in this paper, Theorem 2.12(ii) says that a certain event asymptotically has a positive but possibly small probability. In order to obtain the same result for the simple random graph $G_n$, from the result for $G^*_n$, we have to show that this event has a large intersection with the event $E_s := \{G^*_n \text{ is simple}\}$. Recall that (A2) implies $\mathbb{P}(E_s) \geq c_s + o(1)$ for some $c_s > 0$. In fact, (6.6) and (A4) imply, see e.g. [29, Corollary 1.4] or [2, Theorem 1.1],

$$\mathbb{P}(E_s) = e^{-\nu_n/2 - \nu_n^4/4} + o(1) = e^{-3/4} + o(1), \quad (6.71)$$

so we take $c_s := e^{-3/4}$.

We claim the following:

Lemma 6.14 Assume (A1)-(A4) and (6.4). Then the asymptotic normality (6.20) and (6.21) hold also conditioned on $E_s$. (The expectations in (6.20) and (6.21) are still for the configuration model, without conditioning.)

We postpone the proof of the lemma.

Proof of Theorem 2.12(ii) in the graph case Note that, given Lemma 6.14, we obtain also (6.22) conditioned on $E_s$ by the argument in the proof of Lemma 6.4. That is, for any $b > 0$, there exists $c(b) > 0$ such that

$$\mathbb{P}(\tilde{A}_n(t_1) - \mathbb{E} \tilde{A}_n(t_1) > b\sigma_n \mid E_s) \geq c(b) + o(1). \quad (6.72)$$
Consider now Lemma 6.11. It follows, similarly to the first part of the proof of Lemma 6.11, that \( \mathbb{P}(\mathcal{E}(b) \cap \mathcal{E}_a) \geq c(b) + o(1) \). Hence, \( \mathbb{P}(\mathcal{E}(b) \cap \mathcal{E}_a) \geq c(b)c_s + o(1) \). Since \( \mathcal{E}_a \not\in F_{t_1} \), we modify the next part of the proof of Lemma 6.11. By (6.56),

\[
\mathbb{P}(\mathcal{E}_1(C) \cap \mathcal{E}_a | \mathcal{E}(b)) \geq \frac{1}{2} c_s c(b) + \mathbb{P}(\mathcal{E}_a | \mathcal{E}(b)) - 1 \geq \frac{1}{2} c_s c(b) + \frac{C_9}{C^2} + \mathbb{P}(\mathcal{E}_a | \mathcal{E}(b)) - 1 \geq \frac{1}{2} c_s c(b) + o(1),
\]

for a suitable choice of \( C \). The rest of the proof of Lemma 6.11 works as before. We obtain, using (6.73),

\[
\mathbb{P}(\mathcal{E}(b, C) \cap \mathcal{E}_a) = \mathbb{P}(\mathcal{E}(b) \cap \mathcal{E}_1(C) \cap \mathcal{E}_a) \geq \frac{1}{2} c_s c(b)^2 + o(1).
\]

Hence we conclude, using (6.58) as before, that, for any \( B > 1 \),

\[
\mathbb{P}(\{\tilde{A}_n(t) > 0 : t \in [t_1, Bt_1]\} \cap \mathcal{E}_a) \geq \mathbb{P}(B) + o(1)
\]

for some (new) \( p(B) > 0 \), where \( t_1 \) is as in (6.17). Finally, the proof of Theorem 2.12(ii) above yields, cf. (6.70), \( \mathbb{P}(\{v(C_1) \geq Kn^{2/3}R_n^{-1/3} \} \cap \mathcal{E}_a) \geq p(B) + o(1) \), and thus \( \mathbb{P}(v(C_1) \geq Kn^{2/3}R_n^{-1/3} | \mathcal{E}_a) \geq p(B) + o(1) \), which completes the proof of Theorem 2.12(ii) for the simple random graph \( G_n \).

It remains only to prove Lemma 6.14. This could be done by the method used for similar results in [37, 38], see also [35], but we prefer an alternative, simpler, argument.

**Proof of Lemma 6.14** Consider the conditional analogue of (6.20); the proof of conditional (6.21) is identical.

Let \( a \in \mathbb{R} \) and let \( \mathcal{E}_a := \{ (\tilde{S}_n(t_1) - \mathbb{E}\tilde{S}_n(t_1)) / \sigma_n \leq a \} \); thus, by (6.20),

\[
\mathbb{P}(\mathcal{E}_a) \rightarrow \Phi(a).
\]

Let \( T' \) denote the first time that a connected component is completely explored after time \( t_1 \). Let \( B > 1 \). If \( T' > Bt_1 \), then the component \( C \) explored until \( T' \) has at least \( V_n(t_1) - V_n(t') \geq V_n(t_1) - V_n(Bt_1) \) vertices, and hence, using Lemmas 6.12 and 6.13,

\[
v(C_1) \geq c(\mathcal{E}_a) \geq V_n(t_1) - V_n(Bt_1) = n\mu_n(B - 1)t_1 + o_p(nt_1).
\]

It follows from Theorem 2.12(i) that, for any \( \delta > 0 \) and any fixed \( B \) such that \( \mu(B - 1) > K(\delta) \), we have \( \mathbb{P}(T' > Bt_1) < \delta + o(1) \). Consequently, if \( B_n \rightarrow \infty \), then \( \mathbb{P}(T' > B_n t_1) < 2\delta \) for any \( \delta > 0 \) and all large \( n \), and thus \( T' \leq B_n t_1 \) w.h.p. Note that (6.17) and (6.7) imply that

\[
t_1 R_n = n^{-1/3} R_n^{2/3} = o(1),
\]

and that, since 1 = \( O(R_n) \), we also have \( t_1 = o(1) \). We may thus fix a sequence \( B_n \rightarrow \infty \) such that \( B_n t_1 = o(1) \) and \( B_n t_1 R_n = o(1) \).

Let \( T'' \) be the first time that the number of sleeping half-edges \( S_n(t) \) drops below \( \ell_n/2 \). (Recall that \( S_n(0) = \ell_n = n\mu_n \).) At time \( B_n t_1 \), the expected number of times that \( \mathcal{C}3 \) has been performed is at most \( B_n t_1 \ell_n = o(\ell_n) \), and corresponding to a few of these times also \( \mathcal{C}1 \) was performed; it follows easily that the expected number of sleeping half-edges at \( B_n t_1 \) is \( \ell_n - o(\ell_n) \), and thus w.h.p. \( T'' > B_n t_1 \).
Let $B_{T'}$ denote the event that all the components explored by time $T'$ are simple.

The probability that vertex $i$ is awakened no later than time $(B_{n}t_{1}) \wedge T''$ by using C1 or C3 is $O(B_{n}t_{1}d_{i})$, and, in the event that it is awakened, the probability that two of its half-edges will form a loop is $O(d_{i}^{2}/\ell_{n})$ and the probability that it will be joined by a multiple edge to a vertex $j$ awakened later is $O(d_{i}d_{j}^{2}/\ell_{n}^{2})$. Consequently,

$$
\mathbb{P} \left( B_{T'}^{c} \cap \{ T' \leq B_{n}t_{1} \} \cap \{ T'' > B_{n}t_{1} \} \right) \\
\leq O(B_{n}t_{1}) \sum_{i \in [n]} d_{i} \left[ \frac{d_{i}^{2}}{\ell_{n}} + \frac{d_{i}^{2}}{\ell_{n}} \sum_{j \in \pi_{i}} \frac{d_{j}^{2}}{\ell_{n}} \right] = O(t_{1}B_{n}R_{n}) = o(1),
$$

(6.79) and thus $\mathbb{P}(B_{T'}^{c}) = o(1)$, that is, $B_{T'}$ holds w.h.p.

Then, we condition on the $\sigma$-algebra $F_{T'}$ of all randomness up to time $T'$, and note that $E_{a}$ and $B_{T'}$ are $F_{T'}$-measurable to obtain

$$
\mathbb{P}(E_{a} \cap E_{s}) = \mathbb{P}(E_{a} \cap E_{s} \cap B_{T'}) + o(1) = \mathbb{E}[\mathbb{1}_{E_{a} \cap B_{T'}} \mathbb{P}(E_{s} | F_{T'})] + o(1).
$$

(6.80)

The configuration model multigraph can be partitioned into the connected components found until time $T'$ and those that are found afterwards. The multigraph consisting of all the connected components found after time $T'$ is again (conditioned on $F_{T'}$) a configuration model, now with a random number $\tilde{n} = n(1 - o(1))$ vertices and degrees that are a (random) subset of size $\tilde{n}$ from $[n]$. We denote this degree sequence by $(\tilde{d}_{i})_{i \in [\tilde{n}]}$. In particular, conditional on $F_{T'}$,

$$
\mathbb{P}(E_{s} \mid F_{T'}) = \mathbb{P}(E_{s} \cap B_{T'} \mid F_{T'}) = \mathbb{P}(G(\tilde{n}, (\tilde{d}_{i})_{i \in [\tilde{n}]}) \text{ simple}).
$$

(6.81)

By the discussion above (6.79), the probability that the event $\{ T' \leq B_{n}t_{1} \leq T'' \}$ occurs and that vertex $i$ is part of one of the connected components found before time $T'$ is $O(B_{n}t_{1}d_{i})$. Hence,

$$
\mathbb{E} \left[ \left( \sum_{i \in [n]} d_{i}^{2} - \sum_{i \in [n]} \tilde{d}_{i}^{2} \right) \mathbb{1}_{\{ T' \leq B_{n}t_{1} \leq T'' \}} \right] \leq O(B_{n}t_{1}) \sum_{i \in [n]} d_{i}^{3} = O(nt_{1}B_{n}R_{n}) = o(n).
$$

(6.82)

Consequently, using Markov’s inequality and recalling that $T' \leq B_{n}t_{1} \leq T''$ w.h.p., we obtain

$$
\sum_{i \in [n]} \tilde{d}_{i}^{2} = \sum_{i \in [n]} d_{i}^{2} - o_{p}(n) = (1 + o_{p}(1)) \sum_{i \in [n]} d_{i}^{2}.
$$

(6.83)

Similarly, or as a consequence, $\sum_{i \in [\tilde{n}]} \tilde{d}_{i} = (1 + o_{p}(1)) \sum_{i \in [n]} d_{i}$.

Thus, with $\tilde{v}_{n}$ denoting $v_{n}$ in (2.5) for the (random) degree sequence $(\tilde{d}_{i})_{i \in [\tilde{n}]}$, and noting that $v_{n} = \sum_{i} d_{i}^{2} / \sum_{i} d_{i} - 1$ and $\tilde{v}_{n} = \sum_{i} \tilde{d}_{i}^{2} / \sum_{i} \tilde{d}_{i} - 1$ we obtain

$$
\tilde{v}_{n} = v_{n} + o_{p}(1) = 1 + o_{p}(1).
$$

(6.84)

Consequently, (6.71) yields

$$
\mathbb{P}(G(\tilde{n}, (\tilde{d}_{i})_{i \in [\tilde{n}]}) \text{ simple}) = e^{\tilde{v}_{n} - 1/2 - \tilde{v}_{n}^{2}/4 + o_{p}(1)} = e^{-3/4 + o_{p}(1)} = \mathbb{P}(E_{s}) + o_{p}(1).
$$

(6.85)
and, since \( B_T^r \) holds w.h.p., (6.81) yields
\[
\mathbb{P}(E_s \mid F_T^r) = \mathbb{P}(E_s) + o_p(1). \tag{6.86}
\]

Finally, (6.80) and (6.86) yield, together with (6.76),
\[
\mathbb{P}(E_s \cap E_a) = \mathbb{E}[\mathbb{1}_{E_s \cap B_T^r} \mathbb{P}(E_a)] + o(1) = \mathbb{P}(E_s \cap B_T^r) \mathbb{P}(E_a) + o(1) = \Phi(a) \mathbb{P}(E_s) + o(1), \tag{6.87}
\]
and thus \( \mathbb{P}(E_a \mid E_s) \to \Phi(a) \), which completes the proof of the lemma, and thus of the theorem. \( \square \)

7 | THE COMPLEXITY

Define the complexity of a component \( C \) by \( k(C) := e(C) - v(C) + 1 \); this is the number of independent cycles in \( C \). The estimates in Theorem 2.6 show only that \( k(C_1) = o_p(v(C_1)) \). (This is in contrast to the strongly supercritical case \( \mathbb{E} D(D - 2) > 0 \), when \( v(C_1) = c_v n (1 + o_p(1)) \) and \( e(C_1) = c_e n (1 + o_p(1)) \) for two positive constants \( c_v \) and \( c_e \), see for example [36, Theorem 2.3], and it is easily verified that \( c_e > c_v \) so \( k(C_1) \) also is linear in \( n \).) We can use our methods to obtain a much sharper result. As before, we write \( n = - \log(1 - \rho_n) \), where \( \rho_n \) is the survival probability of a branching process with offspring distribution \( \hat{D}_n = D_n^* - 1 \), with \( D_n^* \) the size-biased version of \( D_n \).

**Theorem 7.1** Suppose that (A1)-(A4) are satisfied, in particular \( \varepsilon_n = o(1) \). Suppose also that \( \varepsilon_n \gg n^{-1/3} (\mathbb{E} D_n^3)^{2/3} \). Then
\[
k(C_1) = n \chi(n(1 + o_p(1))), \tag{7.1}
\]
where
\[
\chi_n := \frac{1}{2} \mu_n (1 - (1 - \rho_n)^2) - \mathbb{E} \left( 1 - (1 - \rho_n)^{D_n} \right) \tag{7.2}
\]
\[
= \frac{1}{2} \mu_n (1 - e^{-2 \alpha_n}) - \mathbb{E} \left( 1 - e^{-\alpha_n D_n} \right) \tag{7.3}
\]
\[
= \mathbb{E} h(\alpha_n D_n) - \frac{1}{2} \mathbb{E} D_n^2 h(2 \alpha_n), \tag{7.4}
\]
with
\[
h(x) := \left( 1 + \frac{x}{2} \right) e^{-x} - 1 + \frac{x}{2} = \frac{1}{2} \sum_{n \geq 3} (-1)^{n-1} \frac{n-2}{n!} x^n. \tag{7.5}
\]

Moreover, \( n \chi_n \to \infty \), \( \chi_n = O(\alpha_n^2 \varepsilon_n) = O(\varepsilon_n^3) \) and
\[
\chi_n \asymp \alpha_n \gamma_n \asymp \mathbb{E} \left( (\alpha_n D_n) \wedge (\alpha_n D_n)^3 \right). \tag{7.6}
\]

**Remark 7.2** The expression (7.2) is what would be intuitively expected from the branching process approximation: if we multiply by \( n \), then the first term is the number of edges \( (\ell_n/2 = n \mu_n/2) \) times the approximate probability that one of the endpoints of an edge attaches to the largest component, and the second term is the approximate probability that a random vertex attaches to the largest component. Indeed, it follows from Theorem 2.6 that the two terms approximate \( e(C_1)/n \) and \( v(C_1)/n \) within a factor \( 1 + o_p(1) \). However, the two terms in (7.2) differ only by a factor \( 1 + o(1) \), so there is a significant cancellation and we need a different argument to show the result.
Remark 7.3  By (7.5) and simple calculus, \( h(0) = h'(0) = 0 \) and \( h''(x) = \frac{1}{2} e^{-x} \), so \( h(x) \) is positive and convex on \( (0, \infty) \). Moreover, \( h(x) \sim \frac{1}{12} x^3 \) as \( x \to 0 \) and \( h(x) \leq \frac{1}{12} x^3 \) for \( x \geq 0 \). Although the expressions in (7.2)-(7.3) are simpler, there is (as said in Remark 7.3) a lot of cancellation, and (7.4) better highlights the order of \( \chi_n \).

We postpone the proof of Theorem 7.1 and state first some consequences for the most important cases.

Theorem 7.4  Suppose that (A1)-(A4) are satisfied, and that \( D_n^3 \) is uniformly integrable. Suppose further that \( \epsilon_n n^{1/3} \to \infty \). Then

\[
k(C_1) = \frac{k \mu}{12} n \rho_n^3 (1 + o_p(1)) = \frac{2 \mu}{3k^3} n \epsilon_n^3 (1 + o_p(1)),
\]

where \( k \in (0, \infty) \) is given by (2.21).

This extends the result for the Erdős-Rényi random graph \( G(n, p) \). There, in the barely supercritical case \( k(C_1) \sim \frac{2}{3} n \epsilon_n^3 \) (see, with more details, [49] and, for \( \epsilon \leq n^{1/12} \), [34]), which corresponds to the case \( D \sim \text{Po}(1) \) (when \( \mu = k = 1 \), of Theorem 7.4 by conditioning on the vertex degrees as in Section 2.6. The order of the complexity in (7.7) interpolates nicely between the known cases of \( \epsilon_n = \epsilon > 0 \) independently of \( n \), where \( k(C_1) \) is of order \( n \), and the critical case \( \epsilon_n = O(n^{-1/3}) \), where \( k(C_1) \) converges in distribution [18].

Theorem 7.5  Suppose that (A1)-(A4) are satisfied, and that \( \mathbb{E} D^3 = \infty \). (Thus \( \mathbb{E} D_n^3 \to \infty \).) Suppose further that \( \epsilon_n \gg n^{-1/3}(\mathbb{E} D_n^3)^{2/3} \). Then

\[
k(C_1) = o_p(n \epsilon_n^3).
\]

Example 7.6  (Power-law degrees)  Consider again the power-law example in Example 2.15, with \( 2 < \gamma < 3 \). It follows from (7.6), (5.58) and (2.38) that \( \chi_n \sim \alpha_n^\gamma \approx \epsilon_n^{3/(\gamma-2)} \). Again, this interpolates nicely between the known cases of \( \epsilon_n = \epsilon > 0 \) independently of \( n \), where \( k(C_1) \) is of order \( n \), and the critical case \( \epsilon_n = O(n^{-1/3}) \), where \( k(C_1) \) converges in distribution. The latter is shown in [19] under stronger power-law assumptions on the degrees, including that \( d_n^{-1/\gamma} \to c_i \) with \( \sum_{i \geq 1} c_i^3 < \infty \), such as for \( c_i / i^{-1/\gamma} \) with \( \gamma \in (2, 3) \). (Recall Remark 2.16, where this is discussed in more detail.)

Example 7.7  Suppose that (A1)-(A4) are satisfied, \( \mathbb{E} D^3 = \infty \), and, furthermore, \( \rho_n \Delta_n = O(1) \). Then Lemma 5.17 applies and yields together with (7.6)

\[
\chi_n \sim \alpha_n \gamma_n \sim \epsilon_n^3 / (\mathbb{E} D_n^3)^2,
\]

showing that (7.8) in this case can be sharpened to \( k(C_1) \approx n \epsilon_n^3 / (\mathbb{E} D_n^3)^2 \) w.h.p.

Lemma 7.8  Suppose that (A1)-(A4) are satisfied and that \( \epsilon_n \gg n^{-1/3}(\mathbb{E} D_n^3)^{2/3} \). Then \( n \alpha_n \gamma_n \to \infty \).

Proof  We consider only \( n \) such that \( \epsilon_n > 0 \); this holds at least for all large \( n \).
First, if \( \alpha_n \Delta_n \leq 1 \), then Lemma 5.17 and the assumptions yield

\[
n \alpha_n \gamma_n \asymp n - \frac{\varepsilon_n}{(n^2 D_n^2)} \to \infty. \tag{7.10}
\]

On the other hand, if \( \alpha_n \Delta_n > 1 \), then by (B8), which was verified in the proof of Theorem 2.6, 
\( 1 < \alpha_n \Delta_n = o(\alpha_n \gamma_n) \), and thus \( n \alpha_n \gamma_n \to \infty \) in this case too.

\[\blacksquare\]

Proof of Theorem 7.1

Let \( N(t) \) be the number of times up to time \( t \) that a new cycle is created. Thus, if \( T \) is a time when \( C1 \) is performed, then \( N(T) \) is the sum of the complexities of the components explored up to \( T \).

During the exploration process, we create a new cycle each time \( C3 \) is performed and the half-edge that dies is an active half-edge, i.e., each time an active half-edge dies spontaneously. This happens with rate \( A_n(t) \). Consequently,

\[
M(t) := N(t) - \int_0^t A_n(u) \, du \tag{7.11}
\]

is a martingale, with \( M(0) = 0 \).

Let \( T_1 \) and \( T_2 \) be as in the proof of Theorem 5.4, so w.h.p. \( C_1 \) is explored between \( T_1 \) and \( T_2 \). Thus w.h.p. \( k(C_1) = N(T_2) - N(T_1) \). Recall that, since \( \beta_n \) is set to \( \alpha_n \), \( T_1 / \alpha_n \xrightarrow{\mathbb{P}} 0 \) and \( T_2 / \alpha_n \xrightarrow{\mathbb{P}} \tau = 1 \), and note that \( T_2 \) is a stopping time.

Recall that (B1)-(B8) were verified in the proof of Theorem 2.6. By (B5) and Lemma 5.7,

\[
\sup_{t \leq T_2 / \alpha_n} \left| \frac{1}{n \gamma_n} A_n(\alpha_n t) - \psi_n(t) \right| \xrightarrow{\mathbb{P}} 0. \tag{7.12}
\]

Consequently, using also that \( \psi_n(t) \) is uniformly bounded on \([0, 2] \) by Remark 5.11, and that \( T_2 / \alpha_n \xrightarrow{\mathbb{P}} 1 \) so that \( T_2 / \alpha_n \leq 2 \) w.h.p.,

\[
\int_0^{T_2} A_n(u) \, du = \alpha_n \int_0^{T_2 / \alpha_n} A_n(\alpha_n u) \, du = n \gamma_n \alpha_n \int_0^{T_2 / \alpha_n} \psi_n(u) \, du + o_p(n \gamma_n \alpha_n) \tag{7.13}
\]

Let

\[
\Psi_n := \int_0^1 \psi_n(t) \, dt, \tag{7.14}
\]

and note that by Remark 5.11 and (B4)(d), \( \Psi_n \asymp 1 \). Define also the stopping time \( T \) by

\[
\int_0^T A_n(u) \, du = n \alpha_n \gamma_n (\Psi_n + 1). \tag{7.15}
\]

By (7.13), \( T_2 \leq T \) w.h.p.

All jumps in the martingale \( M(t) \) are \(+1\), so the quadratic variation (see e.g. [42, Theorem 26.6]) is

\[
[M, M]_t = \sum_{u \leq t} (\Delta M(u))^2 = \sum_{u \leq t} \Delta M(u) = N(t). \tag{7.16}
\]
Hence, for the stopped martingale $M(t \land T)$, using (7.11) and the definition (7.15) of $T$, as well as [50, Corollary 3 to Theorem II.6.27, p. 73],
\[
\mathbb{E}(M(T_2 \land T)^2) = \mathbb{E}[M, M]_{T_2 \land T} = \mathbb{E}[N(T_2 \land T) = \mathbb{E} \int_0^{T_2 \land T} A_n(u) \, du + \mathbb{E} M(T_2 \land T) \\
\leq n\alpha_n\gamma_n(\Psi_n + 1) + 0 = O(n\alpha_n\gamma_n).
\]

Hence it follows that, using also Lemma 7.8,
\[
M(T_2 \land T) = O_p((n\alpha_n\gamma_n)^{1/2}) = o_p(n\alpha_n\gamma_n). \tag{7.17}
\]

By (7.11), (7.13), (7.17) and $T_2 \land T = T_2$ w.h.p.,
\[
N(T_2) = \int_0^{T_2} A_n(u) \, du + M(T_2) = n\alpha_n\gamma_n\Psi_n + o_p(n\alpha_n\gamma_n). \tag{7.18}
\]

Furthermore, for any fixed $\delta > 0$, $T_1 < \delta\alpha_n$ w.h.p. and thus $N(T_1 \land T) \leq N(T \land (\delta\alpha_n))$. Hence, again since $M$ is a martingale,
\[
\mathbb{E} N(T_1 \land T) \leq \mathbb{E} N(T \land (\delta\alpha_n)) = \mathbb{E} \int_0^{T \land (\delta\alpha_n)} A_n(u) \, du. \tag{7.19}
\]

Furthermore, by (7.12) and Remark 5.11,
\[
\int_0^{T \land (\delta\alpha_n)} A_n(u) \, du \leq \int_0^{\delta\alpha_n} A_n(u) \, du = \alpha_n \int_0^{\delta} A_n(\alpha_n t) \, dt \leq n\alpha_n\gamma_n\left(\int_0^{\delta} \Psi_n(t) \, dt + o_p(1)\right) \leq n\alpha_n\gamma_n(\delta + o_p(1)). \tag{7.20}
\]

It follows from (7.19) and (7.20), by dominated convergence justified by (7.15), that
\[
(n\alpha_n\gamma_n)^{-1} \mathbb{E} N(T_1 \land T) \leq \delta + o(1). \tag{7.21}
\]

Since $\delta \in (0, 1)$ is arbitrary, it follows that $\mathbb{E} N(T_1 \land T) = o(n\alpha_n\gamma_n)$, and thus w.h.p. $N(T_1) = N(T_1 \land T) = o_p(n\alpha_n\gamma_n)$. Consequently, recalling (7.18), w.h.p.
\[
k(C_1) = N(T_2) - N(T_1) = n\alpha_n\gamma_n(\Psi_n + o_p(1)) = n\alpha_n\gamma_n\Psi_n(1 + o_p(1)), \tag{7.22}
\]

which shows (7.1) with
\[
\chi_n = \alpha_n\gamma_n\Psi_n. \tag{7.23}
\]

Recalling $\Psi_n \asymp 1$, we have $\chi_n \asymp \alpha_n\gamma_n$ and thus $n\chi_n \rightarrow \infty$ by Lemma 7.8. Furthermore, (7.6) follows from (5.20). It follows from (7.6) and (4.3) that
\[
\chi_n \asymp \mathbb{E}\left((\alpha_nD_n) \land (\alpha_nD_n)^3\right) \leq \mathbb{E}\left((\alpha_nD_n)^2 \land (\alpha_nD_n)^3\right) \leq \alpha_n^2\varepsilon_n, \tag{7.24}
\]

that is, $\chi_n = O(\alpha_n^2\varepsilon_n)$; furthermore $\alpha_n \sim \rho_n = O(\varepsilon_n)$ by (3.2).
It remains to evaluate \( \chi_n \) in (7.23) and show that it agrees with (7.2)-(7.4). By (7.14), (5.21) and Fubini’s theorem,

\[
\chi_n = \alpha_n \gamma_n \Psi_n = \alpha_n \int_0^1 \left( \mu_n e^{-2\alpha_n t} - \mathbb{E}(D_n e^{-\alpha_n t D_n}) \right) dt
\]

(7.25)

which shows (7.3). By the definition (4.1) of \( \alpha_n \), this is the same as (7.2). Furthermore, the equality of (7.4) and (7.3) follows by a simple calculation using (4.8).

---

**Proof of Theorem 7.4**  
Under the assumptions in Theorem 7.4, \( \gamma_n \sim \alpha_n^2 \mathbb{E} D^3 \) by (5.53) and

\[
\Psi_n = \int_0^1 \psi_n(t) dt \to \frac{\kappa \mu}{12 \mathbb{E} D^3}
\]

(7.26)

as a consequence of (5.54). Hence (7.7) follows from (7.1), (7.23) and (4.9).

---

**Proof of Theorem 7.5**  
As in the proof of Theorem 2.9, (3.6) yields \( \alpha_n = o(\epsilon_n) \). Hence, (7.24) implies \( \chi_n = o(\epsilon_n^3) \), and (7.8) follows.

---

**ACKNOWLEDGEMENTS**

This work was commenced while the authors were visiting the Mittag-Leffler Institute in 2009 for the program “Discrete Probability.” Part of the work was done during the authors’ visit to the International Centre for Mathematical Sciences in Edinburgh to attend a workshop “Networks: stochastic models for populations and epidemics” in 2011. The paper was completed during the authors’ visit to the Isaac Newton Institute for Mathematical Sciences for the program “Theoretical Foundations for Statistical Network Analysis,” supported by EPSRC grant EP/K032208/1. RvdH is supported by the Netherlands Organisation for Scientific Research (NWO) through VICI grant 639.033.806 and the Gravitation NETWORKS grant 024.002.003. S.J. is supported by a grant from the Knut and Alice Wallenberg Foundation and a grant from the Simons foundation. M.L. was supported by an EPSRC Leadership Fellowship, grant reference EP/J004022/2, and then by ARC Future Fellowship FT170100409.

**REFERENCES**


