

A random graph of moderate density*

Ágnes Backhausz[†] Tamás F. Móri[‡]

Abstract

We analyse a randomly growing graph model in which the average degree is asymptotically equal to a constant times the square root of the number of vertices, and the clustering coefficient is rather small. In every step, we choose two vertices uniformly at random, check whether they are connected or not, and we either add a new edge or delete one and add a new vertex of degree two to the graph. This dependence on the status of the connection chosen vertices makes the total number of vertices random after n steps. We prove asymptotic normality for this quantity and also for the degree of a fixed vertex (with normalization $n^{1/6}$). We also analyse the proportion of vertices with degree greater than a fixed multiple of the average degree, and the maximal degree.

Keywords: random graphs; intermediate density; degree distribution.

MSC2020 subject classifications: 05C80.

Submitted to ECP on February 15, 2021, final version accepted on December 26, 2021.

1 Introduction

Random graphs have been intensively studied in the last decades [Durrett (2007), van der Hofstad (2016)]. In the current work, we study a randomly growing graph model which has intermediate (moderate) edge density, which includes deletion as a step, and which more likely has a small clustering coefficient. Before going into the details, based on these two properties, we briefly summarize the relation of our model to the ones that are studied in the literature.

For a growing graph sequence, the edge density (i.e. the number of edges divided by the number of pairs of vertices) has a major impact on the behavior of the model. Most preferential attachment graph models are sparse with fixed average degree [Barabási and Albert (1999)], while Erdős–Rényi random graphs with constant edge probabilities, or the stochastic block models are often dense (although they also have sparse versions); their edge density is positive. From the point of view of applications, the edge density can also be different, for example, the brain as a network can be very different from a proteome-proteome interaction network, or the physical network of the internet can be very different from an online social network from this point of view as well. This is one of the reasons why it is important to study random graphs of various densities. From the theoretical point of view, graph limit theory (see e.g. the monograph

*The research was partially supported by the NKFIH "Élvonal" KKP 133921 grant (to Á. B.), and by the Hungarian National Research, Development and Innovation Office NKFIH [grant number K 125569] (to T.F.M.).

[†]ELTE Eötvös Loránd University, Budapest, Hungary. E-mail: agnes.backhausz@ttk.elte.hu

[‡]Alfréd Rényi Institute of Mathematics. E-mail: mori.tamas@renyi.hu

[Lovász (2012)] usually used distinct approaches in the bounded degree case and in the dense case. Furthermore, it took several years while variants of the theory became available for graph sequences of intermediate density [Backhausz and Szegedy (2018), Borgs, Chayes, Cohn and Zhao (2018), Frenkel (2018)]. Although we do not study convergence or limit objects in the current paper, this also suggests that the understanding of random graphs of moderate degree needs different approaches and tools from probability theory.

Deletion of vertices and edges is also an important feature in many real-world networks. However, many of the probabilistic arguments applied for random graphs use some kind of induction with respect to the degrees of vertices and make use of the fact that the degree of a vertex can never decrease. Hence models including deletion of edges require other tools, see e.g. the papers [Deijfen and Lindholm (2009), Flaxman, Frieze and Vera (2007), Thörnblad (2015)]. In our model, when we delete an edge, we immediately add a new vertex connected to its endpoints, which means that the degree of a vertex cannot decrease, similarly to the classical cases. This makes the analysis of the degree of a fixed vertex simpler. Still, due to the average degree tending to infinity, the methods used, for example, for preferential attachment graphs, would not work here. In our model, due to the basic step mentioned above, the clustering coefficient is not expected to be large (although there are other ways how triangles can appear). This is similar to the behavior of many preferential attachment graph models.

In Section 2, we define our random graph model with intermediate density and prove the asymptotic normality of the number of vertices. In Section 3, we prove the main result of the paper, where we determine the asymptotics of the degree of a fixed vertex, and also prove that the fluctuations around the limit tend to Gaussian distribution. In Section 4, we make preliminaries for the following two sections, where we find the almost sure limit of the proportion of vertices with degree larger than a constant times the average degree, and describe the behavior of the maximum degree. For the proofs we are using methods from discrete-time martingale theory.

2 The model

This is a discrete time model. Let us start from an empty graph of size 2. At every step, we choose a pair of vertices uniformly at random. If they are not connected, we connect them with an edge. If they are connected, we delete the edge, add a new vertex to the graph, and connect it to both selected vertices. In this way the graph contains exactly n edges after the n th step. Figure 1 shows a realization for $n = 200$.

Let us denote the number of vertices after n steps by N_n , and let \mathcal{F}_n denote the σ -field generated by the first n steps. We will use the notation $N(\mu; \sigma^2)$ for the Gaussian distribution with mean μ and variance σ^2 .

Theorem 2.1. *For the number of vertices in the model after n steps, we have $N_n \sim 3^{1/3}n^{2/3}$ a.s. as $n \rightarrow \infty$ (that is, $N_n/(3^{1/3}n^{2/3}) \rightarrow 1$ almost surely as $n \rightarrow \infty$). Moreover, the following convergence in distribution holds:*

$$n^{-1/3}(N_n - 3^{1/3}n^{2/3}) \xrightarrow{d} N(0; 3^{1/3}5^{-1}).$$

For the proof we shall need the following lemma, interesting on its own right.

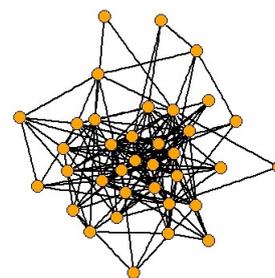


Figure 1: A realization of the random graph model of moderate density with $n = 200$ steps (edges)

Let $g(n)$, $n = 0, 1, \dots$, be an increasing positive sequence, and define $G(n) = \sum_{i=0}^{n-1} g(i)$. Assume $g(n) = O(G(n)^\delta)$ for some δ , $0 < \delta < 1$ (for example, every polynomial multiplied by a power of logarithms satisfies the condition). Roughly speaking, this means that g can only grow polynomially.

Lemma 2.2. Let $(\mathcal{F}_n)_{n \geq 0}$ be a filtration (an increasing sequence of σ -fields), and $A_n \in \mathcal{F}_n$, $n = 1, 2, \dots$. Let $S_n = \sum_{i=1}^n \mathbb{1}(A_i)$, where $\mathbb{1}(\cdot)$ stands for the indicator of the event in brackets. Finally, let $(\eta_n)_{n \geq 0}$ be an adapted sequence of nonnegative random variables such that $H_n = \sum_{i=0}^{n-1} \eta_i \rightarrow \infty$ with probability 1. Suppose

$$\mathbb{P}(A_n \mid \mathcal{F}_{n-1}) = \frac{\eta_{n-1}}{g(S_{n-1})}, \quad n = 1, 2, \dots, .$$

Then $G(S_n) \sim H_n$ a.s. as $n \rightarrow \infty$.

Proof. $\mathbb{E}(G(S_n) \mid \mathcal{F}_{n-1}) = G(S_{n-1}) + \eta_{n-1}$, thus $(M_n = G(S_n) - H_n; \mathcal{F}_n)_{n \geq 1}$ is a martingale. Consider the predictable increasing process B_n in the Doob decomposition of the submartingale M_n^2 . By supposition (namely, using that H_n is \mathcal{F}_{n-1} -measurable and the monotonicity of g), it follows that

$$\begin{aligned} B_n &= \sum_{i=1}^n \mathbb{E}((M_i - M_{i-1})^2 \mid \mathcal{F}_{i-1}) = \sum_{i=1}^n \text{Var}(G(S_i) - G(S_{i-1}) \mid \mathcal{F}_{i-1}) \\ &\leq \sum_{i=1}^n \mathbb{E}((G(S_i) - G(S_{i-1}))^2 \mid \mathcal{F}_{i-1}) = \sum_{i=1}^n g(S_{i-1})\eta_{i-1} \leq g(S_n)H_n \\ &\leq \kappa H_n G(S_n)^\delta \end{aligned}$$

with some positive constant κ . Consequently, by Proposition VII-2-4 of [Neveu (1975)] we have

$$M_n = O(B_n^c) = O(H_n^c G(S_n)^{\delta c}),$$

for every $c > 1/2$. This implies

$$\frac{G(S_n)}{H_n} = 1 + O\left(\left(\frac{G(S_n)}{H_n}\right)^{\delta c} H_n^{(1+\delta)c-1}\right). \tag{2.1}$$

By choosing c such that $1/2 < c < 1/(1 + \delta)$ we get that

$$\frac{G(S_n)}{H_n} = 1 + o\left(\left(\frac{G(S_n)}{H_n}\right)^{\delta c}\right),$$

holds with probability 1, and the proof is completed. □

Proof of Theorem 2.1. Let A_n denote the event that we select a connected pair of vertices at the n th step. Then $N_n = 2 + S_n$, where $S_n = \sum_{i=1}^n \mathbb{1}(A_i)$. Clearly,

$$\mathbb{P}(A_n \mid \mathcal{F}_{n-1}) = \frac{n-1}{\binom{N_{n-1}}{2}},$$

and we can apply Lemma 2.2 to S_n with $\eta_n = n$, $g(n) = \binom{n+2}{2}$. Then $H_n = \binom{n}{2}$, $G(n) = \binom{n+2}{3}$, and we obtain $\binom{N_n}{3} \sim \binom{n}{2}$; that is, $N_n \sim (3n^2)^{1/3}$.

We apply Corollary 3.1 of [Hall and Heyde (1980)] to find the limit distribution. Consider the martingale difference array $(X_{n,i}, \mathcal{F}_{n,i})$, $n \geq 1$, $1 \leq i \leq n$, where

$$X_{n,i} = n^{-5/3} \left[\binom{N_i}{3} - \binom{N_{i-1}}{3} - (i-1) \right], \quad \mathcal{F}_{n,i} = \mathcal{F}_i.$$

The row sums of conditional variances,

$$\begin{aligned} V_{n,n}^2 &= \sum_{i=1}^n \mathbb{E}(X_{n,i}^2 \mid \mathcal{F}_{n,i-1}) = n^{-10/3} \sum_{i=1}^n \left[\mathbb{E} \left(\left[\binom{N_i}{3} - \binom{N_{i-1}}{3} \right]^2 \mid \mathcal{F}_i \right) - (i-1)^2 \right] \\ &= n^{-10/3} \left[\sum_{i=1}^n \binom{N_{i-1}}{2} (i-1) - \frac{n(n-1)(2n-1)}{3} \right] \sim n^{-10/3} \cdot \sum_{i=1}^n \frac{(3i^2)^{2/3}}{2} (i-1) \end{aligned}$$

converge to $3^{5/3}/20$ as $n \rightarrow \infty$ (we have seen $N_n \sim (3n^2)^{1/3}$, and the second term goes to 0). Furthermore, since

$$\max_{1 \leq i \leq n} |X_{n,i}| \leq n^{-5/3} \binom{N_n}{2} = O(n^{-1/3}),$$

the conditional Lindeberg condition, namely,

$$\sum_{i=1}^n \mathbb{E}(X_{n,i}^2 \mathbf{1}(|X_{n,i}| \geq \varepsilon) \mid \mathcal{F}_{i-1}) \rightarrow 0$$

in probability as $n \rightarrow \infty$, for all $\varepsilon > 0$, is obviously satisfied. Hence,

$$n^{-5/3} \left[\binom{N_n}{3} - \binom{n}{2} \right] \xrightarrow{d} N(0; 3^{5/3}/20). \tag{2.2}$$

Recall that $N_n \sim (3n^2)^{1/3}$, hence

$$\binom{N_n}{3}^{1/3} = \frac{N_n}{6^{1/3}} + O(1), \quad \binom{n}{2}^{1/3} = \frac{n^{2/3}}{2^{1/3}} + O(1).$$

Therefore

$$\begin{aligned} N_n - 3^{1/3}n^{2/3} &\sim \binom{N_n}{3}^{1/3} \cdot 6^{1/3} - 3^{1/3} \cdot \binom{n}{2}^{1/3} \cdot 2^{1/3} \\ &= 6^{1/3} \cdot \frac{\binom{N_n}{3} - \binom{n}{2}}{\binom{N_n}{3}^{2/3} + \binom{N_n}{3}^{1/3} \binom{n}{2}^{1/3} + \binom{n}{2}^{2/3}} \sim 6^{1/3} \frac{\binom{N_n}{3} - \binom{n}{2}}{3 \cdot \binom{n}{2}^{2/3}}. \end{aligned}$$

Putting this together with equation (2.2), we obtain the limit distribution of N_n . □

Consider a fixed pair of vertices. They get connected and disconnected alternately. The conditional (with respect to the past) probability that the given pair is selected in the n th step is equal to $\binom{N_{n-1}}{2}^{-1} = O(n^{-4/3})$. Hence by the Lévy generalization of the Borel–Cantelli lemma (see Corollary VII-2-6 in [Neveu (1975)]), the number of changes is almost surely finite. Thus we get as a corollary that fixing an arbitrary (finite) subset of vertices, the induced subgraph will eventually freeze over: the state of the edges and non-edges between these vertices does not change after a certain (random) number of steps.

3 Degree process of a fixed vertex

In this section, we will analyse how fast the degree of a fixed vertex grows. The two vertices of the initial configuration are clearly interchangeable, therefore their degrees have the same asymptotic properties. Let one of them be labelled with zero; the other one remains unlabelled. All the other vertices will be labelled with 1, 2, ... in the order they are added to the graph. Let $d_n(k)$ denote the degree of vertex k at time n .

Theorem 3.1. For every $k = 0, 1, \dots$ we have $d_n(k) \sim 2 \cdot 3^{2/3}n^{1/3}$ a.s. as $n \rightarrow \infty$. Moreover, $n^{-1/6}(d_n(k) - 2 \cdot 3^{2/3}n^{1/3}) \xrightarrow{d} N(0; 2 \cdot 3^{2/3})$.

Proof. Let A_n denote the event that $N_{n-1} \geq k + 1$, and at the n th step we choose vertex k and another vertex that is not a neighbour of k (this is the case where $d_{n-1}(k)$ grows by 1). Then

$$d_n(k) = 2 \cdot \mathbf{1}(N_{n-1} \geq k + 1) + \sum_{i=1}^n \mathbf{1}(A_i). \tag{3.1}$$

Clearly,

$$\mathbb{P}(A_i | \mathcal{F}_{i-1}) = \mathbf{1}(N_{i-1} \geq k + 1) \frac{N_{i-1} - 1 - d_{i-1}(k)}{\binom{N_{i-1}}{2}} \leq \frac{N_{i-1} - 1}{\binom{N_{i-1}}{2}} = \frac{2}{N_{i-1}}, \tag{3.2}$$

hence

$$\sum_{i=1}^n \mathbb{P}(A_i | \mathcal{F}_{i-1}) = O(n^{1/3}).$$

By the Lévy generalization of the Borel–Cantelli lemma it also follows that $d_n(k) = O(n^{1/3})$. Consequently, since the indicator in equation (3.2) is 1 for large enough i (recall $N_n \sim (3n^2)^{1/3}$), and $d_{i-1}(k)/N_{i-1} \rightarrow 0$, we get

$$\mathbb{P}(A_n | \mathcal{F}_{n-1}) \sim 2 \cdot 3^{-1/3}n^{-2/3}, \tag{3.3}$$

implying

$$d_n(k) \sim \sum_{i=1}^n \mathbb{P}(A_i | \mathcal{F}_{i-1}) \sim 2 \cdot 3^{2/3}n^{1/3}.$$

The weak limit can be proved in the same way as in Theorem 2.1. This time let

$$X_{n,i} = n^{-1/6}(\mathbf{1}(A_i) - \mathbb{P}(A_i | \mathcal{F}_{i-1})), \quad \mathcal{F}_{n,i} = \mathcal{F}_i, \quad 1 \leq n, 1 \leq i \leq n.$$

Then

$$\begin{aligned} V_{n,n}^2 &= n^{-1/3} \sum_{i=1}^n \mathbb{P}(A_i | \mathcal{F}_{i-1})(1 - \mathbb{P}(A_i | \mathcal{F}_{i-1})) \\ &\sim n^{-1/3} \sum_{i=1}^n \mathbb{P}(A_i | \mathcal{F}_{i-1}) \stackrel{(3.3)}{\sim} n^{-1/3} \sum_{i=1}^n 2 \cdot 3^{-1/3}i^{-2/3}, \end{aligned}$$

which tends to $2 \cdot 3^{2/3}$ a.s. as $n \rightarrow \infty$. Since $\max_{1 \leq i \leq n} |X_{n,i}| \leq n^{-1/6}$, the Lindeberg condition trivially holds again, and the martingale central limit theorem can be applied to the row sums $\sum_{i=1}^n X_{n,i}$. In this way we obtain

$$n^{-1/6} \left(d_n(k) - \sum_{i=1}^n \mathbb{P}(A_i | \mathcal{F}_{i-1}) \right) \xrightarrow{d} N(0; 2 \cdot 3^{2/3}).$$

We still have to show

$$\sum_{i=1}^n \mathbb{P}(A_i | \mathcal{F}_{i-1}) = 2 \cdot 3^{2/3}n^{1/3} + o(n^{1/6}). \tag{3.4}$$

We already know that

$$\mathbb{P}(A_n | \mathcal{F}_{n-1}) = \frac{2}{N_n} + O(n^{-1}),$$

therefore what we need is a more precise asymptotic formula for N_n . This easily follows from (2.1) with $G(n)$ and H_n borrowed from the proof of Theorem 2.1. In this case $\delta = 2/3$. Set $c = 21/40$, say, then by (2.1),

$$\binom{N_n}{3} = \binom{n}{2} \left[1 + O\left(\binom{n}{2}^{-1/8}\right) \right],$$

which implies that $N_n = 3^{1/3}n^{2/3}(1 + o(n^{-1/6}))$. Thus,

$$\mathbb{P}(A_n | \mathcal{F}_{n-1}) = 2 \cdot 3^{-1/3}n^{-2/3} + o(n^{-5/6}),$$

and we finally obtain (3.4), as needed. □

4 Useful lemmas

Lemma 4.1. *Let $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n$ be an increasing sequence of σ -fields, and let $\xi_1, \xi_2, \dots, \xi_n$ random variables, such that $0 \leq \xi_i \leq 1$ and ξ_i is measurable w.r.t. \mathcal{F}_i , $i = 1, \dots, n$. Introduce*

$$S = \sum_{i=1}^n \xi_i, \quad T = \sum_{i=1}^n \mathbb{E}(\xi_i | \mathcal{F}_{i-1}).$$

Then for every positive c and B we have

$$\mathbb{P}(|S - B| \geq 2c) \leq 3\mathbb{P}(|T - B| > c) + 2 \exp\left\{-\frac{c^2}{2(B + \frac{4}{3}c)}\right\}.$$

Proof. Let us start from the following inequality (cf. Theorem 2.1 and Remark 2.1 in [Fan et al. (2012)]).

Let $(X_i; \mathcal{F}_i)_{i=0, \dots, n}$ be a supermartingale, where $X_0 = 0$, and $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \dots \subset \mathcal{F}_n$. Let $\langle X \rangle_n$ denote the quadratic characteristic of the supermartingale, i.e.,

$$\langle X \rangle_n = \sum_{i=1}^n \mathbb{E}([X_i - X_{i-1}]^2 | \mathcal{F}_{i-1}).$$

Suppose that the differences of the supermartingale are bounded by 1 from above. Then, for any $x \geq 0$ and $v \geq 0$,

$$\mathbb{P}(X_n \geq x \text{ and } \langle X \rangle_n \leq v^2) \leq \exp\left\{-\frac{x^2}{2(v^2 + \frac{1}{3}x)}\right\}. \tag{4.1}$$

Now, define

$$Q =: \sum_{i=1}^n \text{Var}(\xi_i | \mathcal{F}_{i-1}) \leq \sum_{i=1}^n \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1}) \leq T.$$

Then we have

$$\begin{aligned} \mathbb{P}(S \geq B + 2c) &\leq \mathbb{P}(S \geq T + c) + \mathbb{P}(T > B + c) \\ &\leq \mathbb{P}(S - T \geq c, Q \leq B + c) + \mathbb{P}(Q > B + c) + \mathbb{P}(T > B + c). \end{aligned}$$

Here $(X_k = \sum_{i=1}^k [\xi_i - \mathbb{E}(\xi_i | \mathcal{F}_{i-1})]; \mathcal{F}_k)_{k=1, \dots, n}$ is a zero mean martingale, its differences are bounded by 1, and $X_n = S - T$, $\langle X \rangle_n = Q$. Set $x = c$ and $v^2 = B + c$, then (4.1) implies that the first term on the right-hand side is estimated by $\exp\{-c^2/2(B + \frac{4}{3}c)\}$. Thus, using $Q \leq T$ as well, we obtain

$$\mathbb{P}(S \geq B + 2c) \leq \exp\left\{-\frac{c^2}{2(B + \frac{4}{3}c)}\right\} + 2\mathbb{P}(T > B + c). \tag{4.2}$$

On the other hand,

$$\begin{aligned} \mathbb{P}(S \leq B - 2c) &\leq \mathbb{P}(S \leq T - c) + \mathbb{P}(T < B - c) \\ &\leq \mathbb{P}(T - S \geq c, Q \leq B + c) + \mathbb{P}(Q > B + c) + \mathbb{P}(T < B - c). \end{aligned}$$

Here we can estimate the first term by $\exp\{-c^2/2(B + \frac{4}{3}c)\}$ again, considering the martingale $(-X_k; \mathcal{F}_k)$. Consequently,

$$\mathbb{P}(S \leq B - 2c) \leq \exp\left\{-\frac{c^2}{2(B + \frac{4}{3}c)}\right\} + \mathbb{P}(|T - B| > c).$$

The proof can be completed by combining this with (4.2). □

By substituting $c = \varepsilon B/2$ we obtain the following variant.

Corollary 4.2. *For every positive ε and B ,*

$$\mathbb{P}(|S - B| \geq \varepsilon B) \leq 3\mathbb{P}(|T - B| \geq \frac{1}{2}\varepsilon B) + 2\exp\left\{-\frac{\varepsilon^2 B}{8(1 + \frac{2}{3}\varepsilon)}\right\}.$$

Lemma 4.3. *There exist positive constants $\kappa_1, \kappa_2, \kappa_3, \kappa_4$ such that $\kappa_2 < 2/3$ and*

$$\mathbb{P}\left(|N_n - 3^{1/3}n^{2/3}| \geq \kappa_1 n^{\kappa_2}\right) \leq \exp(-\kappa_3 n^{\kappa_4})$$

holds for every $n \geq 1$.

Proof. Actually, we prove the desired inequality with $\kappa_2 = 10/19$ and $\kappa_4 = 1/19$. We will use the following result of [McDiarmid (1998)] (see Theorem 3.14).

Let $X_0 \equiv 0, X_1, \dots, X_n$ be a martingale with respect to the filtration $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n$. Let $(\alpha_i)_{i=1, \dots, n}$ and $(\beta_i)_{i=1, \dots, n}$ be predictable sequences of random variables such that $\alpha_i \leq X_i - X_{i-1} \leq \beta_i, i = 1, \dots, n$. Denote

$$R_n^2 = \sum_{i=1}^n (\beta_i - \alpha_i)^2. \tag{4.3}$$

Then, for every positive x and v ,

$$\mathbb{P}(X_n \geq x \text{ and } R_n^2 \leq v^2) \leq \exp(-2x^2/v^2), \tag{4.4}$$

and consequently,

$$\mathbb{P}(|X_n| \geq x) \leq 2\exp(-2x^2/v^2) + \mathbb{P}(R_n^2 > v^2). \tag{4.5}$$

The proof of the lemma will be performed in four similar steps. Throughout, c_1, c_2, \dots will denote suitable positive constants.

Clearly, it is sufficient to prove the claim when n is large enough, say $n > n_0$, because if κ_1 satisfies $n_0 + 1 < \kappa_1 n_0^{10/19}$, then the desired inequality trivially holds for $n \leq n_0$. Consider the martingale

$$\left(X_k = \binom{N_k}{3} - \binom{k}{2}, \mathcal{F}_k\right), \quad 1 \leq k \leq n.$$

Clearly, $X_1 \equiv 0$, and

$$\begin{aligned} -(k-1) \leq X_k - X_{k-1} &\leq \binom{N_k}{3} - \binom{N_{k-1}}{3} - (k-1) \\ &\leq \binom{N_{k-1} + 1}{3} - \binom{N_{k-1}}{3} - (k-1) = \binom{N_{k-1}}{2} - (k-1), \end{aligned}$$

hence

$$R_n^2 = \sum_{k=1}^n \binom{N_{k-1}}{2}^2 \leq nN_{n-1}^4.$$

Firstly, $N_{n-1} \leq n$, thus $R_n^2 \leq n^5$. Recall from the proof of Theorem 2.1 that $M_n = G(S_n) - H_n = \binom{N_n}{3} - \binom{n}{2}$. Since $\mathbb{P}(N_n \geq n^{16/19}) \leq \mathbb{P}(M_n \geq c_1 n^{48/19})$, by setting $x = c_1 n^{48/19}$, $v^2 = n^5$ and applying (4.4) we get

$$\mathbb{P}(N_n \geq n^{16/19}) \leq \exp(-c_2 n^{1/19}),$$

which, together with inequality (4.3), implies that

$$\mathbb{P}(R_n^2 > n^{83/19}) \leq \exp(-c_2 n^{1/19}).$$

Next, $\mathbb{P}(N_n \geq n^{14/19}) \leq \mathbb{P}(M_n \geq c_3 n^{42/19})$, so let $x = c_3 n^{42/19}$, $v^2 = n^{83/19}$; then (4.4) leads to

$$\mathbb{P}(N_n \geq n^{14/19}) \leq \exp(-c_4 n^{1/19}) + \exp(-c_2 n^{1/19}).$$

Therefore, using inequality (4.3) again,

$$\mathbb{P}(R_n^2 > n^{75/19}) \leq \exp(-c_5 n^{1/19}).$$

Now, let $x = n^2$ and $v^2 = n^{75/19}$ in (4.4). Then we have

$$\begin{aligned} \mathbb{P}(N_n \geq 3n^{2/3}) &\leq \mathbb{P}\left(\binom{N_n}{3} \geq \frac{3}{2}n^2\right) \leq \mathbb{P}(M_n \geq n^2) \\ &\leq \exp(-2n^{1/19}) + \exp(-c_5 n^{1/19}) \leq \exp(-c_6 n^{1/19}), \end{aligned} \quad (4.6)$$

and from this it follows that

$$\mathbb{P}(R_n^2 \geq c_7 n^{11/3}) \leq \exp(-c_6 n^{1/19}). \quad (4.7)$$

Finally, we have

$$\left| \frac{N_n}{6^{1/3}} - \frac{n^{2/3}}{2^{1/3}} \right| \leq \left| \frac{N_n^3}{6} - \frac{n^2}{2} \right| \left(\frac{n^2}{2} \right)^{-2/3}.$$

Here

$$0 \leq \frac{N_n^3}{6} - \binom{N_n}{3} \leq \frac{N_n^2}{2}, \quad \frac{n^2}{2} - \binom{n}{2} = \frac{n}{2},$$

therefore

$$\left| \frac{N_n}{6^{1/3}} - \frac{n^{2/3}}{2^{1/3}} \right| \leq \left(|M_n| + \frac{N_n^2 + n}{2} \right) 2n^{-4/3}.$$

Thus, by (4.6),

$$\mathbb{P}\left(\left| \frac{N_n}{6^{1/3}} - \frac{n^{2/3}}{2^{1/3}} \right| \geq 2n^{-4/3}|M_n| + 10\right) \leq \mathbb{P}(N_n \geq 3n^{2/3}) \leq \exp(-c_6 n^{1/19}),$$

and hence

$$\begin{aligned} \mathbb{P}(|N_n - 3^{1/3}n^{2/3}| \geq n^{10/19}) &\leq \mathbb{P}(6^{1/3}(2n^{-4/3}|M_n| + 10) \geq n^{10/19}) + \exp(-c_6 n^{1/19}) \\ &\leq \mathbb{P}(|M_n| \geq c_8 n^{10/19+4/3}) + \exp(-c_6 n^{1/19}). \end{aligned}$$

By (4.5), the first probability on the right-hand side can be estimated with $x = c_8 n^{10/19+4/3}$ and, in virtue of (4.7), $v^2 = c_7 n^{11/3}$. It follows that

$$\mathbb{P}(|M_n| \geq c_8 n^{10/19+4/3}) \leq 2 \exp(-c_9 n^{1/19}) + \exp(-c_6 n^{1/19}).$$

Altogether we get

$$\mathbb{P}(|N_n - 3^{1/3}n^{2/3}| \geq n^{10/19}) \leq \exp(-c_{10} n^{1/19})$$

if n is large enough, as needed. □

5 Asymptotic degree distribution

In many recursive random graph models, including several types of scale free graphs, the proportion of vertices of a fixed degree converges almost surely to a deterministic limit, which makes up a discrete probability distribution, often a power law. In those models the number of edges grows linearly with the number of vertices, thus the graphs are sparse. By contrast, our graph appears of moderate density, as the number of edges is $n = \Theta(N_n^{3/2})$. Not a surprise that an asymptotic degree distribution does not exist in the sense above. Instead of counting vertices with a fixed degree we compare actual degrees to the current average degree, which grows to infinity with n .

Let \bar{d}_n denote the average degree after n steps. That is, \bar{d}_n is equal to the sum of degrees divided by the number of vertices, where the numerator of this fraction is just the double of the number of edges. Thus $\bar{d}_n = 2n/N_n \sim 2 \cdot 3^{-1/3}n^{1/3}$. Notice that the degree of every fixed vertex is asymptotically the triple of \bar{d}_n . This is so because the newer a vertex is, the smaller its degree is and the younger vertices pull down the average.

Theorem 5.1. *Let $0 < t < 3$. Then the asymptotic proportion of vertices with degree greater than t times the average degree is a.s. convergent, namely,*

$$\lim_{n \rightarrow \infty} N_n^{-1} |\{k \leq N_n : d_n(k) > t\bar{d}_n\}| = (1 - t/3)^2.$$

This result is similar to Theorem 2.1 of [Backhausz and Móri (2012)]. There the asymptotic degree distribution was considered in a fixed level of the uniform recursive tree. The uniform recursive tree is defined recursively by starting from a single node called the root, and adding a new vertex at every step, which is to be connected to an existing vertex selected uniformly at random [Pittel (1994)]. Let $L_n(m)$ denote the set of vertices that are of distance m from the root after n steps; this is called level m . It can be proved that $|L_n(m)| \sim (\log n)^m/m!$ almost surely as $n \rightarrow \infty$. Then the average degree in level m is

$$\bar{d}_{n,m} = \frac{|L_n(m+1)|}{|L_n(m)|} + 1 \sim \frac{\log n}{m+1} \quad \text{a.s.,}$$

and for every t , $0 < t < m + 1$, we have

$$\lim_{n \rightarrow \infty} |L_n(m)|^{-1} |\{k \in L_n(m) : d_n(k) > t\bar{d}_{n,m}\}| = (1 - t/(m+1))^m$$

with probability 1. The method of the proof can be adapted to our present problem, but the fact that here the dynamics of the degree process $d_n(k)$ depends on the current value of the degree makes the analysis much harder.

Proof of Theorem 5.1. Throughout the proof we will say that a random event occurs with extra high probability (shortened to WEHP) if the probability of the complementary event has stretched exponential decay as $n \rightarrow \infty$: it can be majorized by $c_1 \exp(-c_2 n^{c_3})$ with suitable positive constants c_1, c_2, c_3 . For instance, our Lemma 4.3 can be rephrased as $|N_n - 3^{1/3}n^{2/3}| < \kappa_1 n^{\kappa_2}$ WEHP.

Let $t \in (0, 3)$ be fixed, and let ε be a sufficiently small positive number (it will tend to 0 in the end). Let n be sufficiently large (it will tend to ∞), and $\varepsilon n^{2/3} < k < (1 - 2\varepsilon)3^{1/3}n^{2/3}$. Then by Lemma 4.3 we have $k \leq N_n$ WEHP. Introduce

$$B = \int_{3^{-1/2}k^{3/2}}^n \frac{2}{3^{1/3}t^{2/3}} dt = 2 \cdot 3^{2/3}n^{1/3} - 2(3k)^{1/2}. \quad (5.1)$$

This quantity cannot be too small by the choice of k , namely, since $(3k)^{1/2} < (1 - \varepsilon)3^{2/3}n^{1/3}$, we have $B > 2\varepsilon \cdot 3^{2/3}n^{1/3}$.

Let us estimate $d_n(k)$ more precisely than in the asymptotic formula in Theorem 3.1. Let us start from (3.1). We are going to apply Corollary 4.2 to the sum $S = \sum_{i=1}^n \mathbb{1}(A_i)$ with B introduced in (5.1). To this end, we have to estimate the deviation of $T = \sum_{i=1}^n \mathbb{P}(A_i | \mathcal{F}_{i-1})$ from B .

On the one hand, by (3.2) and Lemma 4.3 it follows that

$$T \leq \sum_{i < n} \frac{2}{N_i} \mathbb{1}(N_i \geq k) \leq \left(1 + \frac{\varepsilon}{3}\right) \sum_{i < n} \frac{2}{3^{1/3} i^{2/3}} \mathbb{1}\left(3^{1/3} i^{2/3} \geq \left(1 - \frac{\varepsilon}{3}\right)k\right) \tag{5.2}$$

$$< \left(1 + \frac{\varepsilon}{2}\right)B \quad \text{WEHP.}$$

Note that here we are applying termwise estimations to a sum with an increasing number of terms. For the sake of the reader's convenience let us go into details. As $N_i \leq i + 1$, the index of the first nonzero term in the sum is of order $n^{2/3}$ at most (in fact, it is of order $n^{1/2}$, but we will not need that). Hence the probability that $\left(1 - \frac{\varepsilon}{3}\right)3^{1/3}i^{2/3} < N_i < \left(1 - \frac{\varepsilon}{3}\right)3^{1/3}i^{2/3}$ is not satisfied is bounded by $\exp(-\kappa_3 i^{\kappa_4})$. For the whole sum, from the union bound we get the upper bound $O\left(n \exp(-\kappa_3(\varepsilon n^{2/3})^{\kappa_4})\right)$ for the probability that inequality (5.2) is not satisfied, and this still decreases at a stretched exponential rate.

Therefore we also have

$$d_n(k) \leq 2 + S \leq 2 + (1 + \varepsilon)B < \frac{\varepsilon}{3}(N_n - 1) \quad \text{WEHP}$$

by (4.2) and Lemma 4.3.

On the other hand, T can be estimated from below as follows.

$$\frac{N_{i-1} - 1 - d_{i-1}(k)}{\binom{N_{i-1}}{2}} > \left(1 - \frac{\varepsilon}{3}\right) \frac{2}{N_i} \quad \text{WEHP,}$$

therefore

$$T > \left(1 - \frac{\varepsilon}{3}\right) \sum_{i < n, N_i \geq k} \frac{2}{N_i} \geq \left(1 - \frac{\varepsilon}{2}\right)B \quad \text{WEHP.}$$

Now, Corollary 4.2 implies that $(1 - \varepsilon)B < S < (1 + \varepsilon)B$ WEHP, thus

$$(1 - \varepsilon)2\left[3^{2/3}n^{1/3} - (3k)^{1/2}\right] < d_n(k) < (1 + \varepsilon)2\left[3^{2/3}n^{1/3} - (3k)^{1/2}\right].$$

Moreover, since

$$t\bar{d}_n < (1 + \varepsilon)2t \cdot 3^{-1/3}n^{1/3} < (t + 2\varepsilon)2 \cdot 3^{-1/3}n^{1/3},$$

and similarly,

$$t\bar{d}_n > (t - 2\varepsilon)2 \cdot 3^{-1/3}n^{1/3}$$

WEHP by Lemma 4.3, we finally get

$$d_n(k) - t\bar{d}_n < (1 + \varepsilon)2 \cdot 3^{2/3}n^{1/3} - (t - 2\varepsilon)2 \cdot 3^{-1/3}n^{1/3} - (1 + \varepsilon)2(3k)^{1/2}$$

$$< \left(1 - \frac{t}{3} + 2\varepsilon\right)2 \cdot 3^{2/3}n^{1/3} - (1 + \varepsilon)2(3k)^{1/2}$$

$$< (1 + \varepsilon)2 \cdot 3^{1/2} \left[\left(1 - \frac{t}{3} + 2\varepsilon\right)N_n^{1/2} - k^{1/2} \right] \quad \text{WEHP,}$$

and similarly,

$$d_n(k) - t\bar{d}_n > (1 - \varepsilon)2 \cdot 3^{1/2} \left[\left(1 - \frac{t}{3} - 2\varepsilon\right)N_n^{1/2} - k^{1/2} \right].$$

Consequently, if $d_n(k) > t\bar{d}_n$, then $k < (1 - t/3 + 2\varepsilon)^2 N_n$ WEHP, and $\varepsilon N_n < k < (1 - t/3 - 2\varepsilon)^2 N_n$ implies $d_n(k) > t\bar{d}_n$ WEHP. Thus,

$$\left(1 - \frac{t}{3} - 2\varepsilon\right)^2 - \varepsilon < \frac{|\{k \leq N_n : d_n(k) > t\bar{d}_n\}|}{N_n} < \left(1 - \frac{t}{3} + 2\varepsilon\right)^2$$

WEHP. This inequality is valid for every sufficiently large n by the Borel–Cantelli lemma, implying the same lower and upper bounds for the \liminf and \limsup of the middle expression almost surely. Now we can complete the proof by letting ε decrease to 0. \square

6 Maximal degree

Let $\mu_n = \max\{d_n(k) : 0 \leq k \leq N_n\}$ be the maximal degree in the graph after n steps. (In fact, this is not the maximal degree yet, as the degree of the unlabelled initial vertex is missing, but we shall see that asymptotically it makes no difference.) In many increasing sparse graph models the maximal degree increases at the same rate as the degree of each individual vertex, see e.g. [Móri (2005)]. It is no different now; though, unlike in the case of scale-free models, younger vertices do not have smaller degrees in the limit.

Theorem 6.1. *Let $\mu_n = \max\{d_n(k) : 0 \leq k \leq N_n\}$ be the maximal degree in the random graph of moderate density after n steps. Then we have*

$$\lim_{n \rightarrow \infty} n^{-1/3} \mu_n = 2 \cdot 3^{2/3} \quad \text{a.s.}$$

Proof. It is sufficient to show that $\mu_n \leq (1 + \varepsilon)2 \cdot 3^{2/3}$ WEHP, for all positive ε . Then the Borel–Cantelli lemma gives $\limsup_{n \rightarrow \infty} n^{-1/3} \mu_n \leq (1 + \varepsilon)2 \cdot 3^{2/3}$ a.s. On the other hand,

$$\liminf_{n \rightarrow \infty} n^{-1/3} \mu_n \geq \liminf_{n \rightarrow \infty} n^{-1/3} d_n(0) = 2 \cdot 3^{2/3}$$

almost surely.

Similarly to the proof of Theorem 5.1 we consider a vertex k and estimate the probability $\mathbb{P}(S \geq (1 + \varepsilon)2 \cdot 3^{2/3}n^{1/3})$ where $S = d_n(k) - 2$ is a sum of indicators, see (3.1). Since this time we only need an upper bound for S , we do not have to keep k apart from 0 and N_n . Again,

$$\begin{aligned} T &\leq 2 \sum_{i=0}^{n-1} N_i^{-1} < \varepsilon n^{1/3} + 2 \sum_{\varepsilon n^{1/3} \leq i < n} N_i^{-1} \leq \varepsilon n^{1/3} + \left(1 + \frac{\varepsilon}{4}\right) \int_0^n \frac{2}{3^{1/3}t^{2/3}} dt \\ &= \varepsilon n^{1/3} + \left(1 + \frac{\varepsilon}{4}\right)2 \cdot 3^{2/3}n^{1/3} < \left(1 + \frac{\varepsilon}{2}\right)2 \cdot 3^{2/3}n^{1/3} \end{aligned}$$

WEHP, not depending on k . By (4.2), we therefore have

$$S < (1 + \varepsilon)2 \cdot 3^{2/3}n^{1/3} \quad \text{WEHP.}$$

Consequently, $\mu_n \leq (1 + \varepsilon)2 \cdot 3^{2/3}n^{1/3}$ WEHP, as needed. \square

The difference from the sparse scale-free models can be caught by the fact that $n^{-1/6}(\mu_n - 2 \cdot 3^{2/3}n^{1/3})$ tends to infinity (at least in probability) instead of having a non-degenerate limit distribution. Indeed, similarly to the proof of Theorem 3.1, by using the multivariate version of the martingale central limit theorem, one can show that the individual degree processes become independent in the limit, thus for an arbitrary fixed positive integer k we have

$$n^{-1/6}(\mu_n - 2 \cdot 3^{2/3}n^{1/3}) \geq n^{-1/6} \max_{0 \leq i < k} (d_n(i) - 2 \cdot 3^{2/3}n^{1/3}),$$

where the right-hand side converges in distribution to the maximum of k independent $N(0; 2 \cdot 3^{2/3})$ random variables.

References

- [Backhausz and Móri (2012)] Backhausz Á, Móri TF (2012), Degree distribution in the lower levels of the uniform recursive tree. *Ann. Univ. Sci. Budapest Sect. Comput.* 36: 53–62. MR2914830
- [Backhausz and Szegedy (2018)] Backhausz Á, Szegedy B, Action convergence of operators and graphs. *Canadian Journal of Mathematics* (2020), 1-50.
<https://doi.org/10.4153/S0008414X2000070X>
- [Barabási and Albert (1999)] Barabási A-L, Albert R (1999), Emergence of scaling in random networks. *Science* 286(5439): 509–512. MR2091634
- [Borgs, Chayes, Cohn and Zhao (2018)] Borgs C, Chayes JT, Cohn H, Zhao Y (2018), An L^p theory of sparse graph convergence II: LD convergence, quotients, and right convergence. *Ann. Probab.* 46(1): 337–396. MR3758733
- [Deijfen and Lindholm (2009)] Deijfen M, Lindholm M (2009), Growing networks with preferential deletion and addition of edges. *Physica A: Statistical Mechanics and its Applications*, 388(19): 4297–4303.
<https://doi.org/10.1016/j.physa.2009.06.032>
- [Durrett (2007)] Durrett R (2007), *Random graph dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge Univ. Press, Cambridge. MR2271734
- [Fan et al. (2012)] Fan X, Grama I, Liu Q (2012), Hoeffding’s inequality for supermartingales. *Stochastic Process. Appl.* 122(10): 3545–3559. MR2956116
- [Flaxman, Frieze and Vera (2007)] Flaxman AD, Frieze AM, Vera J (2007), Adversarial deletion in a scale-free random graph process. *Combin. Probab. Comput.* 16: 251–270. MR2298813
- [Frenkel (2018)] Frenkel PE (2018), Convergence of graphs with intermediate density. *Trans. Amer. Math. Soc.* 370(5): 3363–3404. MR3766852
- [Hall and Heyde (1980)] Hall P, Heyde CC (1980) *Martingale Limit Theory and Its Application*. Academic Press: New York. MR0624435
- [van der Hofstad (2016)] van der Hofstad R, *Random graphs and complex networks* (2017). Cambridge series in statistical and probabilistic mathematics, 43. MR3617364
- [Lovász (2012)] Lovász L (2012), *Large networks and graph limits*, American Mathematical Society Colloquium Publications, 60, American Mathematical Society, Providence, RI. MR3012035
- [McDiarmid (1998)] McDiarmid C (1998), Concentration. In: Habib M et al. (eds.) *Probabilistic Methods for Algorithmic Discrete Mathematics*, Springer Verlag: Berlin, Heidelberg, 195–248. MR1678578
- [Móri (2005)] Móri TF (2005), The maximum degree of the Barabási–Albert random tree. *Comb Probab Computing* 14: 339–348. MR2138118
- [Neveu (1975)] Neveu J (1975), *Discrete-Parameter Martingales*, North-Holland: Amsterdam. MR0402915
- [Pittel (1994)] Pittel, B (1994), Note on the heights of random recursive trees and random m -ary search trees. *Random Structures Algorithms* 5(2): 337–347. MR1262983
- [Thörnblad (2015)] Thörnblad E (2015), Asymptotic degree distribution of a duplication-deletion random graph model. *Internet Math.* 11(3): 289–305. MR3344243