Local Degree Distribution in Scale Free Random Graphs

Article *in* ELECTRONIC JOURNAL OF PROBABILITY · July 2010

DOI: 10.1214/EJP.v16-916 · Source: arXiv

2 authors, including:

Tamás F Móri
Alfréd Rényi Institute of Mathematics

LOCAL DEGREE DISTRIBUTION IN SCALE FREE RANDOM GRAPHS

ÁGNES BACKHAUSZ AND TAMÁS F. MÓRI

Department of Probability Theory and Statistics, Faculty of Science
Eötvös Loránd University
Pázmány P. s. 1/C, H-1117 Budapest, Hungary
E-mail address: agnes@cs.elte.hu, moritamas@ludens.elte.hu

Abstract. In several scale free graph models the asymptotic degree distribution and the characteristic exponent change when only a smaller set of vertices is considered. Looking at the common properties of these models, we present sufficient conditions for the almost sure existence of asymptotic degree distribution constrained on the set of selected vertices, and identify the chararteristic exponent belonging to it.

Date: 14 July 2010.
2000 *Mathematics Subject Classification.* 60G42, 05C80.
Key words and phrases. Scale free, random graphs, recursive trees, martingales, regular variation.
Research supported by the Hungarian National Foundation for Scientific Research, Grant No. K67961.

1. Introduction

Since the end of the nineties several complex real world networks and their random graph models have been investigated [4, 5, 6]. Many of them possess the scale free property: the tail of the degree distribution decreases polynomially fast, that is, if c_d denotes the proportion of vertices of degree d, then $c_d \approx C \cdot d^{-\gamma}$ holds for large values of d [1]. γ is called the characteristic exponent.

If the whole network is completely known, the empirical estimator of the characteristic exponent may have nice properties. However, real world networks usually are too large and complex, hence our knowledge of the graph is partial. For several models of evolving random graphs the degree distribution and the characteristic exponent change when attention is restricted to a set of selected vertices that are close to the initial configuration [10, 11, 13].

Starting from these phenomena, in this paper the degree distribution constrained on a set of selected vertices will be investigated, assuming that the graph model possesses the scale free property with characteristic exponent $\gamma > 1$, and the number of selected vertices grows regularly with exponent $0 < \alpha \le 1$. Sufficient conditions for the almost sure existence of the local asymptotic degree distribution will be given. It will be shown that under these conditions the characteristic exponent of the constrained degree distribution is $\alpha\,(\gamma - 1) + 1$.

The proofs are based on the methods of martingale theory. Applications of the general results to different graph models (e.g. to the Albert–Barabási random tree [1]) will be shown.

In Section 2 we present the family of random graph models to be examined and formulate the sufficient conditions. In Sections 3 and 4 we mention some results about martingales and slowly varying sequences to be applied in the proofs. Section 5 contains the proof of the main results, and in Section 6 we give some examples and applications.

2. Main results

In this section we present sufficient conditions for the almost sure existence of asymptotic degree distribution constrained on the set of selected vertices, and we describe that distribution.

Let $(G_n = (V_n, E_n))_{n\in\mathbb{N}}$ be a sequence of evolving simple random graphs. Some vertices are distinghuised; let $S_n \subseteq V_n$ denote the set of selected vertices.

We start from a finite, simple graph $G_0 = (V_0, E_0)$, this is the initial configuration with $V_0 = \{u_1, u_2, \dots, u_l\}$. $S_0 \subseteq V_0$ is arbitrarily chosen. For $n \ge 1$, at the nth step

- one new vertex, v_n, is added to the graph: $V_n = V_0 \cup \{v_1, \dots, v_n\}$;
- the new vertex gets some random edges, thus $E_{n-1} \subseteq E_n$, and every edge from $E_n \setminus E_{n-1}$ is connected to v_n;

- the new vertex can be added to the set of selected vertices, $v_n \in S_n$ is a random choice.

The σ-field of events generated by the first n steps is denoted by $\mathcal{F}_n$.

For $v \in V_n$ let the degree of v in G_n be denoted by $\deg_n(v)$. Furthermore, for $n \geq 1$, and $d \geq 0$ define

$$X[n,d] = |\{v \in V_n : \deg_n(v) = d\}|;$$
$$Y[n,d] = |\{v \in V_n : \deg_n(v) = d,\ (v, v_{n+1}) \in E_{n+1}\}|;$$
$$I[n,d] = \begin{cases} 1 & \text{if } \deg_n(v_n) = d, \\ 0 & \text{otherwise.} \end{cases}$$

In some models it is possible that the new vertex does not get any edges at some steps. In other models the degree of the new vertex is fixed, for example, the degree of the new vertex is always 1 in random tree models. If the new vertex gets at least m edges at each step for some $m > 0$, then $X[n,d]$ is at most $|V_0|$ for all n and $d < m$. Thus we denote the minimal initial degree of the new vertex by m, and we consider $X[n,d]$ only for $d \geq m$. Of course, $m = 0$ is also possible.

2.1. **Conditions on the graph model.** We say that a discrete probability distribution (a_n) is *exponentially decreasing* if $a_n \leq C \cdot q^n$ holds for all $n \geq 1$ for some $C > 0$ and $0 < q < 1$. A sequence (a_n) is *slowly varying* if $a_{[sn]}/a_n \to 1$ as $n \to \infty$ for all $s > 0$.

Throughout this paper, for two sequences $(a_n), (b_n)$ of nonnegative numbers, $a_n \sim b_n$ means that $b_n > 0$ except finitely many terms, and $a_n/b_n \to 1$ as $n \to \infty$.

Now we can formulate the conditions on the graph model.

Condition 1. *$X[n,d] \sim c_d \cdot n$ holds as $n \to \infty$ for every $d \geq m$ with probability 1, where (c_d) is a probability distribution and c_d is positive for all $d \geq m$.*

This means that asymptotic degree distribution exists in this graph model. Note that $X[n,d] \to \infty$ as $n \to \infty$ almost surely.

Condition 2. *$c_d \sim K \cdot d^{-\gamma}$ holds as $d \to \infty$ for some positive numbers K and γ.*

This is the so called scale free property with characteristic exponent γ. That is, the asymptotic degree distribution decays polynomially with exponent γ. This implies that c_d is positive for every d large enough, but we will need it for all $d \geq m$, this is included in Condition 1.

Condition 3. *For every $n \geq 0$, if $w_1, w_2 \in V_n$ and $\deg_n(w_1) = \deg_n(w_2)$, then*

$$\mathbb{P}\left((w_1, v_{n+1}) \in E_{n+1} \middle| \mathcal{F}_n\right) = \mathbb{P}\left((w_2, v_{n+1}) \in E_{n+1} \middle| \mathcal{F}_n\right).$$

In other words, at each step, conditionally on the past, old vertices of the same degree get connected to the new vertex with the same probability.

Condition 4. $\sum_{i=1}^{n+1} I\left[i,d\right] = p_d \cdot n + o\left(n\right)$ *holds as* $n \to \infty$ *for every* $d \geq m$ *with probability* 1*, where* (p_d) *is an exponentially decreasing probability distribution.*

Loosely speaking, the degree of the new vertex has an exponentially decreasing asymptotic distribution. This trivially holds if the degree of the new vertex is fixed.

Condition 5. *For every* $d \geq m$ *there exists a random variable* $Z_d \geq 0$ *with exponentially decreasing distribution such that*

$$\mathbb{P}\left(Y\left[i,d\right] \geq l \mid \mathcal{F}_i\right) \leq \mathbb{P}\left(Z_d \geq l\right), \quad i \geq 1,\ l \geq 1.$$

In many particular cases the following stronger condition is also met.

There exists a random variable $Z \geq 0$ *with exponentially decreasing distribution such that*

$$\mathbb{P}\left(\deg_n(v_n) \geq l \mid \mathcal{F}_{n-1}\right) \leq \mathbb{P}\left(Z \geq l\right), \quad n \geq 1,\ l \geq m. \tag{1}$$

This is a sort of upper bound for the initial degree of the new vertex.

Condition 6. *For every* $d \geq m$ *we have*

$$k_d = \sum_{j=m}^{d} \left(p_j - c_j\right) > 0.$$

We will see later that the nonnegativity of k_d follows from the previous conditions; however, the positivity of k_d cannot be omitted, as an example will show.

2.2. **Conditions on the set of selected vertices.** Recall that $S_n \subseteq V_n$ is the set of selected vertices in G_n. We emphasize that $\deg_n\left(v\right)$ always denotes the degree of vertex v in G_n, not in S_n.

We will need the following notations. The σ-field generated by the first n steps and adding the edges of v_{n+1} at the $(n+1)$st step is denoted by $\mathcal{F}_n^+$. Furthermore, for $n \geq 1$ and $d \geq m$ let

$$X^*\left[n,d\right] = \left|\left\{v \in S_n :\ \deg_n\left(v\right) = d\right\}\right|;$$

$$Y^*\left[n,d\right] = \left|\left\{v \in S_n :\ \deg_n\left(v\right) = d, \left(v, v_{n+1}\right) \in E_{n+1}\right\}\right|;$$

$$I^*\left[n,d\right] = \begin{cases} 1 & \text{if } v_n \in S_n \text{ and } \deg_n\left(v_n\right) = d, \\ 0 & \text{otherwise;} \end{cases}$$

$$I^*\left(n\right) = \sum_{d=m}^{n} I^*\left[n,d\right] = \begin{cases} 1 & \text{if } v_n \in S_n, \\ 0 & \text{otherwise.} \end{cases}$$

The conditions on the set of selected vertices are the following.

Condition 7. *$S_n \subseteq S_{n+1}$ for all $n \geq 0$.*

Vertices cannot be deleted from the set of selected vertices.

Condition 8. *$I^*(n+1)$ is $\mathcal{F}_n^+$-measurable for all $n \geq 0$.*

At each step we have to decide whether the new vertex is to be selected immediately after choosing its neighbours. Selecting the neighbours of a fixed vertex is an example.

Condition 9. *There exists a sequence of positive random variables (ζ_n) that are slowly varying as $n \to \infty$, and $|S_n| = \sum_{i=1}^n I^*(i) \sim \zeta_n \cdot n^\alpha$ for some $\alpha > 0$, with probability 1.*

This means that the size of the set of selected vertices is regularly growing with exponent $\alpha > 0$.

Condition 10. *For every $d \geq m$*

$$\sum_{i=1}^n E\left(I^*[i,d] \,|\, \mathcal{F}_{i-1}\right) = (q_d + o(1)) \sum_{i=1}^n E\left(I^*(i) \,|\, \mathcal{F}_{i-1}\right)$$

holds a.s. as $n \to \infty$, with some exponentially decreasing probability distribution $(q_d)_{d \geq m}$.

The last condition holds if the degree of the new vertex v_n is fixed, or its degree and $I^*(n)$ are independent conditionally on $\mathcal{F}_{n-1}$. In that case sequence $q_d = p_d$ satisfies the condition. It is also possible that the asymptotic degree distribution of the new selected vertices is different from (p_d) if only it decays exponentially fast.

2.3. **Description of the local degree distribution.** Now we formulate the main results.

Theorem 1. *Suppose that Conditions 1–10 hold for a random graph model (G_n, S_n), then the limits*

$$\lim_{n\to\infty} \frac{X^*[n,d]}{|S_n|} = x_d$$

exist for all $d \geq m$ with probability 1.

The constants x_d satisfy the following recursive equations.

$$x_m = \frac{\alpha\, q_m}{\alpha + \frac{p_m - c_m}{c_m}}, \quad x_d = \frac{x_{d-1} \cdot \frac{k_{d-1}}{c_{d-1}} + \alpha \cdot q_d}{\alpha + \frac{k_d}{c_d}} \quad (d \geq m+1).$$

Sequence (x_d) is a probability distribution, that is, it sums up to 1. Moreover, $x_d \sim L \cdot d^{-\gamma^*}$ as $d \to \infty$ with $L > 0$ and

$$\gamma^* = \alpha(\gamma - 1) + 1.$$

Remark 1. *From the proof it is clear that with Condition 2 dropped the limits x_d still exist and the recursive equations remain valid. The role of the scale free property of the graph is just to guarantee that the asymptotic degree distribution constrained on the set of selected vertices is also polynomially decaying.*

3. Martingales

We will extensively use the following propositions that are based on well-known facts of martingale theory.

Proposition 1. *Let $(M_n, \mathcal{G}_n)$ be a square integrable martingale with $M_1 = 0$, $\mathcal{G}_0 = \{\emptyset, \Omega\}$. Introduce*

$$A_n = \sum_{i=2}^{n} E\left((M_i - M_{i-1})^2 \middle| \, \mathcal{G}_{i-1} \right),$$

that is, the predictable increasing process in the Doob decomposition of M_n^2. Then $M_n = o\left(A_n^{1/2} \log A_n \right)$ holds almost surely on the event $\{A_\infty = \infty\}$, and M_n converges to a finite limit, as $n \to \infty$, almost surely on the event $\{A_\infty < \infty\}$.

This is a corollary of Propositions VII-2-3 and VII-2-4 of [14].

Proposition 2. *Let $(M_n, \mathcal{G}_n)$ be a square integrable nonnegative submartingale, and*

$$A_n = EM_1 + \sum_{i=2}^{n} \left(E\left(M_i | \, \mathcal{G}_{i-1} \right) - M_{i-1} \right), \quad B_n = \sum_{i=2}^{n} \operatorname{Var}\left(M_i | \, \mathcal{G}_{i-1} \right).$$

If $B_n^{1/2} \log B_n = O\left(A_n\right)$, then $M_n \sim A_n$ on the event $\{A_n \to \infty\}$.

This is easy to prove applying Proposition 1 to the martingale part of the Doob decomposition of M_n.

Proposition 3. *Let $Y_1, Y_2, \ldots$ be nonnegative, uniformly bounded random variables, and $\mathcal{G}_n = \sigma\left(Y_1, \ldots, Y_n\right)$. Then the symmetric difference of the events $\left\{ \sum_{n=1}^{\infty} Y_n < \infty \right\}$ and $\left\{ \sum_{n=1}^{\infty} E\left(Y_n | \, \mathcal{G}_{n-1} \right) < \infty \right\}$ has probability 0. Moreover,*

$$\frac{\sum_{n=1}^{\infty} Y_n}{\sum_{n=1}^{\infty} E\left(Y_n | \, \mathcal{G}_{n-1} \right)} \to 1 \quad (n \to \infty)$$

holds almost everywhere on the event $\left\{ \sum_{n=1}^{\infty} Y_n = \infty \right\}$.

This proposition follows from the Lévy generalization of the Borel–Cantelli lemma that can be found in [14] (Corollary VII-2-6).

4. Slowly varying sequences

In the proofs we will use the basic results of the theory of regularly varying sequences, see e.g. [2, 3, 7].

We say that a sequence of positive numbers (β_n) is regularly varying with exponent μ if the following holds:

$$\beta_n \sim \gamma_n n^{\mu} \quad (n \to \infty)$$

where (γ_n) is slowly varying.

(β_n) is regularly varying with exponent μ if and only if $\beta_{[sn]}/\beta_n \to s^{\mu}$ as $n \to \infty$ for all $s > 0$, see Bingham [2].

Proposition 4. *Let (α_n), (β_n) be nonnegative sequences such that (α_n) is slowly varying as $n \to \infty$, and $n^{-\lambda}\beta_n \to 1$ as $n \to \infty$ for some $\lambda > -1$. Then the following holds.*

$$\sum_{i=1}^{n} \alpha_i \beta_i \sim \alpha_n \sum_{i=1}^{n} \beta_i \quad (n \to \infty)\,.$$

This is a consequence of the results of Bojanić and Seneta [2, 3].

Proposition 5. *Let (α_n), (β_n) be nonnegative sequences such that (α_n) is regularly varying with exponent δ.*

a) Suppose $\sum_{i=1}^{n} \beta_i = B_n$ is regularly varying with exponent $\mu > 0$, and $\mu + \delta > 0$. Then

$$\sum_{i=1}^{n} \alpha_i \beta_i \sim \frac{\mu}{\delta + \mu}\, \alpha_n B_n \quad (n \to \infty)\,.$$

b) Suppose $\sum_{i=1}^{n} \beta_i = o(B_n)$, where (B_n) is regularly varying with exponent $\mu > 0$, and $\mu + \delta > 0$. Then

$$\sum_{i=1}^{n} \alpha_i \beta_i = o\left(\alpha_n B_n\right) \quad (n \to \infty)\,.$$

Proof. *a*) Suppose first that $\delta = 0$, that is, (α_n) is slowly varying. By Bojanić and Seneta [3], for a nonnegative slowly varying sequence (α_n) there always exists another nonnegative sequence (α'_n) such that $\alpha_n \sim \alpha'_n$ as $n \to \infty$, and

$$\lim_{n\to\infty} n\left(1 - \frac{\alpha'_{n-1}}{\alpha'_n}\right) = 0. \tag{2}$$

This implies that $\alpha_{n+1}/\alpha_n \to 1$ as $n \to \infty$.

All sequences are nonnegative, hence we have

$$\sum_{i=1}^{n}\alpha_i'\beta_i=\alpha_n'\sum_{j=1}^{n}\beta_j+\sum_{i=1}^{n-1}\left(\alpha_i'-\alpha_{i+1}'\right)\sum_{j=1}^{i}\beta_j \tag{3}$$
$$=\alpha_n'\sum_{j=1}^{n}\beta_j-\sum_{i=1}^{n-1}i\left(1-\frac{\alpha_i'}{\alpha_{i+1}'}\right)\left(\frac{\alpha_{i+1}'}{i}\sum_{j=1}^{i}\beta_j\right)$$

as $n\to\infty$.

Sequence (α_n') is slowly varying. By supposition, $\sum_{j=1}^{n}\beta_j=\gamma_n n^{\mu}$ as $n\to\infty$ with some slowly varying sequence (γ_n), hence $i^{-1}\sum_{j=1}^{i}\beta_j=\gamma_i\, i^{\mu-1}$ as $i\to\infty$. Since $\lambda=\mu-1>0$, by applying Proposition 4 we obtain that

$$\sum_{i=1}^{n}\frac{\alpha_{i+1}'}{i}\sum_{j=1}^{i}\beta_j\sim\sum_{i=1}^{n}\alpha_{i+1}'\gamma_i\, i^{\mu-1}\sim\alpha_{n+1}'\gamma_n\sum_{i=1}^{n}i^{\mu-1}\sim\frac{1}{\mu}\,\alpha_n\gamma_n n^{\mu}$$

as $n\to\infty$. Combining this with (2) we get that the second term on the right-hand side of (3) is $o\left(\alpha_n\gamma_n n^{\mu}\right)$ as $n\to\infty$.

The first term is asymptotically equal to $\alpha_n\gamma_n n^{\mu}$ as $n\to\infty$. Thus we get that

$$\sum_{i=1}^{n}\alpha_i\beta_i\sim\sum_{i=1}^{n}\alpha_i'\beta_i\sim\alpha_n\gamma_n n^{\mu}\quad(n\to\infty)\,.$$

Next, let δ differ from 0. Let $\alpha_n=\kappa_n n^{\delta}$, and $B_n=\gamma_n n^{\mu}$ with slowly varying sequences (κ_n) and (γ_n). We have

$$\sum_{i=1}^{n}i^{\delta}\beta_i=\sum_{i=1}^{n}i^{\delta}\left(B_i-B_{i-1}\right)=n^{\delta}B_n+\sum_{i=1}^{n-1}\left(i^{\delta}-(i+1)^{\delta}\right)B_i$$
$$=n^{\delta}B_n-\delta\sum_{i=1}^{n-1}i^{\delta-1}\left(1+o\left(1\right)\right)B_i$$
$$=\gamma_n n^{\delta+\mu}-(1+o(1))\,\delta\sum_{i=1}^{n-1}\gamma_i i^{\delta+\mu-1}\quad(n\to\infty)\,.$$

(γ_n) is slowly varying, and $\lambda=\delta+\mu-1>-1$, thus Proposition 4 applies, and we obtain that

$$\sum_{i=1}^{n}i^{\delta}\beta_i\sim\gamma_n n^{\delta+\mu}-\delta\gamma_n\sum_{i=1}^{n-1}i^{\delta+\mu-1}$$
$$\sim\gamma_n n^{\delta+\mu}-\delta\gamma_n\frac{n^{\delta+\mu}}{\delta+\mu}=\frac{\mu}{\delta+\mu}\,\gamma_n n^{\delta+\mu}\quad(n\to\infty)\,.$$

Let us apply the already proved particular case to (κ_n) and $\left(n^\delta \beta_n\right)$. Then we get that

$$\sum_{i=1}^n \alpha_i\beta_i = \sum_{i=1}^n \kappa_i i^\delta \beta_i \sim \kappa_n \sum_{i=1}^n i^\delta \beta_i \sim \frac{\mu}{\delta+\mu}\,\kappa_n\gamma_n n^{\delta+\mu} = \frac{\mu}{\delta+\mu}\,\alpha_n B_n.$$

b) We can suppose that B_n is increasing, since for every regularly varying sequence with positive exponent one can find another, increasing one, which is equivalent to it. Introduce $\beta_n' = \beta_n + B_n - B_{n-1}$, with $B_0 = 0$. Then $\beta_n' \geq 0$, and

$$\sum_{i=1}^n \beta_i' = \sum_{i=1}^n \beta_i + B_n \sim B_n,$$

hence it is regularly varying with exponent μ. By part *a*) we have

$$\sum_{i=1}^n \alpha_i\beta_i' \sim \frac{\mu}{\delta+\mu}\,\alpha_n B_n,$$

and also

$$\sum_{i=1}^n \alpha_i\left(B_i - B_{i-1}\right) \sim \frac{\mu}{\delta+\mu}\,\alpha_n B_n.$$

After subtraction we obtain that $\sum_{i=1}^n \alpha_i\beta_i = o(\alpha_n B_n)$. $\square$

Proposition 6. *Let $a_1, a_2 \dots$ and $b_1, b_2, \dots$ be nonnegative numbers satisfying*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n a_i = K < \infty, \qquad \lim_{n\to\infty} nb_n = 1.$$

Then

a) $\exp\left(\sum_{i=1}^n a_i b_i\right)$ *is regularly varying with exponent K as $n \to \infty$;*
b) $\exp\left(\sum_{i=1}^n a_i^2 b_i^2 s_i\right)$ *is slowly varying as $n \to \infty$ for every bounded sequence of real numbers (s_n).*

Proof. *a*) Let $m = [tn],\ t > 1$. We have

$$\begin{aligned}\sum_{i=n+1}^m a_i b_i &= \sum_{i=n+1}^m \frac{a_i q_i}{i} + \sum_{i=n+1}^m \frac{1}{i}\left[i(K+r_i) - (i-1)(K+r_{i-1})\right] \\ &= \sum_{i=n+1}^m \frac{a_i q_i}{i} + K\sum_{i=n+1}^m \frac{1}{i} + \sum_{i=n+1}^m (r_i - r_{i-1}) + \sum_{i=n+1}^m \frac{r_{i-1}}{i}.\end{aligned}$$

The first sum on the right-hand side tends to 0, since

$$\left|\sum_{i=n+1}^m \frac{a_i q_i}{i}\right| \leq \sum_{i=n+1}^m \frac{a_i\left|q_i\right|}{i} \leq \frac{t}{m}\sum_{i=n+1}^m a_i\left|q_i\right| = o\left(1\right). \tag{4}$$

The second sum is $K\log t + o\left(1\right)$, the third one is $r_m - r_{n-1} = o\left(1\right)$, and the last one also converges to 0.

b) Let $m = [tn],\ t > 1$. Now we have

$$\sum_{i=n+1}^{m} a_i^2 b_i^2 s_i = \sum_{i=n+1}^{m} \frac{a_i}{i} \cdot \frac{(1+q_i)^2 a_i s_i}{i}.$$

By supposition

$$\frac{a_i}{i} = \frac{K + ir_i - (i-1)\, r_{i-1}}{i} = \frac{K}{i} + r_i - r_{i-1} + \frac{r_{i-1}}{i} \to 0 \quad (i \to \infty),$$

hence

$$q_i' = \frac{(1+q_i)^2 a_i s_i}{i} \to 0 \quad (i \to \infty).$$

We can complete the proof similarly to (4).

$$\left| \sum_{i=n+1}^{m} a_i^2 b_i^2 s_i \right| = \left| \sum_{i=n+1}^{m} \frac{a_i q_i'}{i} \right| \le \frac{t}{m} \sum_{i=n+1}^{m} a_i \left| q_i' \right| = o(1). \quad \square$$

5. Proofs

For sake of convenience, instead of $X^*[n,d]$, we consider the number $Z^*[n,d]$ of selected vertices with degree greater than or equal to d. That is, for $n \ge 1$ and $d \ge m$ let

$$Z^*[n,d] = \left| \{ v \in S_n :\ \deg_n(v) \ge d \} \right| = \sum_{j=d}^{n} X^*[n,j]. \tag{5}$$

We also need the following notations.

$$J^*[n,d] = \sum_{j=d}^{n} I^*[n,j], \quad J[n,d] = \sum_{j=d}^{n} I[n,j].$$

First we show that Theorem 1 is implied by the following proposition. For all $d \ge m$ we have $Z^*[n,d] \sim z_d |S_n|$ a.s. as $n \to \infty$ with some positive constants z_d. In addition,

$$z_m = 1, \quad z_d = \frac{z_{d-1} \frac{k_{d-1}}{c_{d-1}} + \alpha \sum_{j=d}^{\infty} q_j}{\alpha + \frac{k_{d-1}}{c_{d-1}}} \quad (d \ge m+1). \tag{6}$$

It is clear that

$$X^*[n,d] = Z^*[n,d] - Z^*[n,d+1] \quad (n \ge 1,\ d \ge m),$$

hence

$$X^*[n,d] = (z_d - z_{d+1})\, |S_n| + o(|S_n|)$$

a.s. as $n \to \infty$. Thus the limits

$$\lim_{n \to \infty} \frac{X^*[n,d]}{|S_n|} = x_d$$

exist for all $d \ge m$ almost surely, and $x_d = z_d - z_{d+1}$ for all $d \ge m$.

It is easy to derive the recursive equations for $x_d = z_d - z_{d+1}$ from $z_m = 1$ and equation (6). The denominators are positive, because

Conditions 1, 6, and 9 guarantee that c_d is nonnegative, α is positive, and k_d is positive.

It is also easy to check that sequence (x_d) is a probability distribution. We have

$$x_m\alpha + x_m\frac{k_m}{c_m} = \alpha q_m,$$

and

$$x_d\alpha + x_d\,\frac{k_d}{c_d} = x_{d-1}\,\frac{k_{d-1}}{c_{d-1}} + \alpha q_d \quad (d \geq m+1)\,.$$

Summing up the equations above we get that

$$\sum_{d=m}^{\infty} x_d = \sum_{d=m}^{\infty} q_d = 1,$$

since, by Conditions 9 and 10, $\alpha > 0$ and the sequence (q_d) is a probability distribution.

The next step is solving the recursion for (x_d). Set

$$t_d = \frac{k_d}{c_d}, \quad a_d = \prod_{i=m}^{d-1}\frac{t_i+\alpha}{t_i} \quad (d \geq m)\,.$$

It is easy to check that the recursive equations of Theorem 1 are satisfied by the sequence

$$x_d = \frac{1}{t_d+\alpha}\,\sum_{i=0}^{d} q_i\alpha\,\prod_{j=i}^{d-1}\frac{t_j}{t_j+\alpha} \quad (d \geq m)\,.$$

By Condition 2, $c_d \sim K\cdot d^{-\gamma}$ holds as $d\to\infty$, and by Condition 4 the sequence (p_j) is exponentially decreasing. Hence it follows, as $d\to\infty$, that

$$\begin{aligned}
k_d &= -\sum_{j=m}^{d}(c_j - p_j) \sim -K\cdot\frac{d^{-\gamma+1}}{-\gamma+1};\\
t_d &= \frac{k_d}{c_d} \sim \frac{-K\cdot\frac{d^{-\gamma+1}}{-\gamma+1}}{K\cdot d^{-\gamma}} = \frac{d}{\gamma-1};\\
a_d &= \prod_{i=0}^{d-1}\left(1+\frac{\alpha}{t_i}\right) \sim \prod_{i=0}^{d-1}\left(1+\frac{\alpha\,(\gamma-1)}{i}\right) \sim K'\cdot d^{\alpha(\gamma-1)}
\end{aligned}$$

for some $K' > 0$. By Condition 10 the sequence (q_d) is exponentially decreasing, thus the series in the expression

$$x_d = \frac{1}{a_d\,(t_d+\alpha)}\sum_{i=m}^{d} a_i q_i\alpha$$

converges. Using the asymptotics of (a_d) and (t_d) we get that

$$x_d = \frac{1}{a_d\left(t_d+\alpha\right)}\sum_{i=0}^{d} a_i q_i \alpha \sim L\cdot d^{-\alpha(\gamma-1)-1}$$

for some $L>0$.

Consequently, the degree distribution constrained on the set of selected vertices decays polynomially, and the new characteristic exponent is determined by α and γ, namely, $\gamma^* = \alpha\left(\gamma-1\right)+1$, as stated.

Therefore Theorem 1 is indeed a consequence of (6).

Let us continue with the proof of (6). We proceed by induction on d.

The case $d=m$ is obvious, because the initial degree is never less than m, and the degree of a vertex cannot decrease, thus every vertex in $S_n \setminus S_0$ has at least m edges.

Suppose that

$$Z^*\left[n,d-1\right] \sim z_{d-1}\left|S_n\right| \quad (n\to\infty)$$

holds for some $z_{d-1}>$ and $d\geq m+1$ almost surely.

First we determine the expected number of vertices of degree $\geq d$ in S_{n+1}, given $\mathcal{F}_n$, for $n\geq 1$. Every vertex in S_n counts if its degree is at least d in G_n, or if its degree is equal to $d-1$ in G_n and it gets a new edge from v_{n+1}. The new vertex v_{n+1} counts if it falls into S_{n+1} and its degree is $\geq d$ in G_{n+1}. Thus the following equality holds for every $n\geq 1$.

$$Z^*\left[n+1,d\right] = Z^*\left[n,d\right] + Y^*\left[n,d-1\right] + J^*\left[n+1,d\right]. \tag{7}$$

Taking conditional expectations with respect to $\mathcal{F}_n$ we obtain that

$$\begin{aligned} &E\left(Z^*\left[n+1,d\right]\middle|\,\mathcal{F}_n\right) \\ &\qquad = Z^*\left[n,d\right] + E\left(Y^*\left[n,d-1\right]\middle|\,\mathcal{F}_n\right) + E\left(J^*\left[n+1,d\right]\middle|\,\mathcal{F}_n\right). \end{aligned} \tag{8}$$

By Condition 3, vertices of the same degree are connected to v_{n+1} with the same conditional probability. This implies that

$$E\left(\frac{Y^*\left[n,d\right]}{X^*\left[n,d\right]}\middle|\,\mathcal{F}_n\right) = E\left(\frac{Y\left[n,d\right]}{X\left[n,d\right]}\middle|\,\mathcal{F}_n\right), \quad (n\geq 1). \tag{9}$$

$X\left[n,d\right]$ may be equal to zero, then $Y\left[n,d\right]=0$ as well. We will consider all quotients of the form $0/0$ as 1.

The middle term on the right-hand side of (8) can be transformed by the help of (9).

$$\begin{aligned} &E\left(Z^*\left[n+1,d\right]\middle|\,\mathcal{F}_n\right) \\ &= Z^*\left[n,d\right] + X^*\left[n,d-1\right]\frac{E\left(Y\left[n,d-1\right]\middle|\,\mathcal{F}_n\right)}{X\left[n,d-1\right]} + E\left(J^*\left[n+1,d\right]\middle|\,\mathcal{F}_n\right). \end{aligned} \tag{10}$$

By (5), $X^*[n, d-1] = Z^*[n, d-1] - Z^*[n, d]$, hence from equation (10) we obtain that

$$E(Z^*[n+1, d] | \mathcal{F}_n) = Z^*[n, d] \left(1 - \frac{E(Y[n, d-1] | \mathcal{F}_n)}{X[n, d-1]}\right) + Z^*[n, d-1] \frac{E(Y[n, d-1] | \mathcal{F}_n)}{X[n, d-1]} + E(J^*[n+1, d] | \mathcal{F}_n) \tag{11}$$

for all $n \geq 1$.

For $i \geq 1$ define

$$b[i, d] = \begin{cases} 1 & \text{if } X[i, d] = 0; \\ \left(1 - \frac{E(Y[i,d]|\mathcal{F}_i)}{X[i,d]}\right)^{-1} & \text{if } X[i, d] > 0. \end{cases}$$

Set $c[1, d] = 1$ and for $n \geq 2$ define

$$c[n, d] = \prod_{i=1}^{n-1} b[i, d]. \tag{12}$$

Then for n large enough we have

$$\frac{c[n, d]}{c[n+1, d]} = \left(1 - \frac{E(Y[n, d] | \mathcal{F}_n)}{X[n, d]}\right).$$

For several particular models it is quite easy to compute the conditional expectations $E(Y[i, d-1] | \mathcal{F}_i)$, and hence, to determine the asymptotics of $c[n, d]$. In the present general case the conditional expectation is not specified. However, as the following sequence of lemmas shows, the asymptotics of the partial sums can be described, and one can calculate the asymptotics of $c[n, d]$. The proof of the lemmas will be postponed to the second part of this section. We emphasize that in the lemmas the induction hypothesis is assumed all along.

Consider the partial sums

$$S[n, d] = \sum_{i=1}^{n} E(Y[i, d] | \mathcal{F}_i) \quad (n \geq 1).$$

Lemma 1. *For all $d \geq m$ we have*

$$S[n, d] = \sum_{i=1}^{n} E(Y[i, d] | \mathcal{F}_i) = k_d \cdot n + o(n) \quad (n \to \infty) \tag{13}$$

with probability 1.

Remark 2. *It is clear from the definition that $S[n, d]$ is nonnegative, hence Lemma 1 immediately implies $k_d \geq 0$ for all $d \geq m$ (cf. Condition 6).*

Lemma 2.

$$c[n, d] \sim a[n, d] \cdot n^{k_d/c_d} \quad (n \to \infty),$$

a.s. for all $d \geq m$, where $a[n,d]$ is positive and slowly varying as $n \to \infty$.

By equation (11), the process

$$V[n,d] = c[n,d-1]\, Z^*[n,d] \quad (n \geq 1) \tag{14}$$

is a submartingale. Let $A[n,d]$ denote the increasing process in the Doob decomposition of $V[n,d]$; it is given by

$$\begin{aligned} A[n,d] = \sum_{i=1}^{n} c[i+1,d-1]\, Z^*[i,d-1] \frac{E\left(Y[i,d-1] \middle| \mathcal{F}_i\right)}{X[i,d-1]} \\ + \sum_{i=1}^{n} c[i+1,d-1]\, E\left(J^*[i+1,d] \middle| \mathcal{F}_i\right). \end{aligned} \tag{15}$$

First we describe the asymptotics of $A[n,d]$.

Lemma 3. *Suppose that $Z^*[n,d-1] \sim z_{d-1}|S_n|$ holds a.s. for some $d \geq m+1$, as $n \to \infty$, then*

$$A[n,d] \sim \frac{z_{d-1}\frac{k_{d-1}}{c_{d-1}} + \alpha\sum_{j=d}^{\infty} q_j}{\alpha + \frac{k_{d-1}}{c_{d-1}}}\, a[n,d]\, \zeta_n n^{\alpha + k_{d-1}/c_{d-1}} \quad a.s.$$

Next, we compute an upper bound for the conditional variances. Define

$$B[n,d] = \sum_{i=2}^{n} \operatorname{Var}\left(V[i,d] \middle| \mathcal{F}_{i-1}\right) \quad (n \geq 2).$$

Lemma 4. *Suppose that $Z^*[n,d-1] \sim z_{d-1}|S_n|$ holds a.s. for some $d \geq m+1$, as $n \to \infty$, then $B[n,d]^{1/2} \log B[n,d] = O(A[n,d])$.*

Therefore Proposition 2 implies that $V[n,d] \sim A[n,d]$ almost surely as $n \to \infty$. Finally, by Lemma 2 and Lemma 3 we obtain the asymptotics

$$Z^*[n,d] \sim \frac{z_{d-1}\frac{k_{d-1}}{c_{d-1}} + \alpha\sum_{j=d}^{\infty} q_j}{\alpha + \frac{k_{d-1}}{c_{d-1}}}\, \zeta_n n^{\alpha} \quad (n \to \infty).$$

Consequently, we have

$$Z^*[n,d] \sim z_d \zeta_n n^{\alpha} \quad (n \to \infty)$$

with

$$z_d = \frac{z_{d-1}\frac{k_{d-1}}{c_{d-1}} + \alpha\sum_{j=d}^{\infty} q_j}{\alpha + \frac{k_{d-1}}{c_{d-1}}}. \tag{16}$$

The size of S_n is asymptotically equal to $\zeta_n n^{\alpha}$ by Condition 9. Thus the proof of (6) can be completed by using Lemmas 1–4. □

Now we continue with the proofs of Lemmas 1–4.

Proof of Lemma 1. Similarly to equation (7), but considering all vertices, we see that

$$X[i+1,j] = X[i,j] - Y[i,j] + Y[i,j-1] + I[i+1,j]$$

for every $i \geq 0$ and $j \geq m$. Adding up for $i = 1, \ldots, n$ we obtain that

$$X[n+1,j] - X[1,j] = -\sum_{i=1}^{n} Y[i,j] + \sum_{i=1}^{n} Y[i,j-1] + \sum_{i=2}^{n+1} I[i,j] \tag{17}$$

for every $j \geq m$ and $n \geq 1$. By Conditions 1 and 4, from (17) it follows that

$$\sum_{i=1}^{n} Y[i,j-1] - \sum_{i=1}^{n} Y[i,j] = (c_j - p_j) \cdot n + o(n)$$

holds almost surely, as $n \to \infty$, for every $j \geq m$. Adding this up for $j = m, \ldots, d$ we get

$$\sum_{i=1}^{n} Y[i,d] = -\sum_{j=m}^{d} (c_j - p_j) \cdot n + o(n) = k_d \cdot n + o(n) \tag{18}$$

a.s., as $n \to \infty$. Therefore it is sufficient to prove that

$$\frac{1}{n} \sum_{i=1}^{n} \left(Y[i,d] - E\left(Y[i,d] \middle| \mathcal{F}_i \right) \right) \to 0 \quad (n \to \infty).$$

Fix $d \geq m$, and for $n \geq 1$ let $M_n = \sum_{i=1}^{n} \left(Y[i,d] - E\left(Y[i,d] \middle| \mathcal{F}_i \right) \right)$, $\mathcal{G}_n = \mathcal{F}_{n+1}$. It is clear that $(M_n, \mathcal{G}_n)$ is a martingale. Using Condition 5 we will derive an upper bound for the corresponding increasing process A_n introduced in Proposition 1.

$$\begin{aligned} A_n = \sum_{i=1}^{n} \operatorname{Var}\left(Y[i,d] \middle| \mathcal{F}_i \right) &\leq \sum_{i=1}^{n} E\left(Y[i,d]^2 \middle| \mathcal{F}_i \right) \leq \\ \leq \sum_{i=1}^{n} C_i E\left(Y[i,d] \middle| \mathcal{F}_i \right) &+ \sum_{i=1}^{n} E\left(Y[i,d]^2 I\left(Y[i,d] > C_i \right) \middle| \mathcal{F}_i \right) \\ &\leq \sum_{i=1}^{n} C_i E\left(Y[i,d] \middle| \mathcal{F}_i \right) + \sum_{i=1}^{n} E\left(Z_d^2 I\left(Z_d > C_i \right) \right) \end{aligned} \tag{19}$$

for any $C_i > 0$. Fix $\varepsilon > 0$ such that $\kappa = E\left(e^{\varepsilon Z_d} \right)$ is finite, and for $i \geq 3$ choose $C_i = \frac{2}{\varepsilon} \log i$. The function $z \mapsto z^2 e^{-\varepsilon z}$ is decreasing for $z > \frac{2}{\varepsilon}$, hence $z^2 e^{-\varepsilon z} \leq C_i^2 e^{-\varepsilon C_i}$ for $z > C_i$. This implies

$$\begin{aligned} E\left(Z_d^2 I\left(Z_d > C_i \right) \right) &\leq C_i^2 e^{-\varepsilon C_i} E\left(e^{\varepsilon Z_d} I\left(Z_d > C_i \right) \right) \\ &\leq (2/\varepsilon)^2 (\log i)^2 i^{-2} \kappa. \end{aligned} \tag{20}$$

The infinite sum of these terms converges, thus the second sum on the right-hand side of (19) is bounded for fixed d.

On the other hand, $Y[i,d] \le |V_i| \le i+l$ follows from the definition, therefore

$$\sum_{i=3}^{n} C_i E\left(Y[i,d] \mid \mathcal{F}_i\right) \le \sum_{i=3}^{n} \frac{2}{\varepsilon} \log i \cdot (i+l) = O\left(n^2 \log n\right).$$

Thus $A_n = O\left(n^2 \log n\right)$. This bound can be further improved as follows. Applying Proposition 1 to the martingale (M_n) we get that $M_n = O\left(n^{1+\eta}\right)$ a.s. for all $\eta > 0$. Equation (18) implies that $\sum_{i=1}^{n} Y[i,d] = O(n)$, therefore

$$\begin{aligned}\sum_{i=1}^{n} C_i E\left(Y[i,d] \mid \mathcal{F}_i\right) &\le C_n \sum_{i=1}^{n} E\left(Y[i,d] \mid \mathcal{F}_i\right) \\ &= C_n \left(\sum_{i=1}^{n} Y[i,d] - M_n\right) = O\left(n^{1+\eta} \log n\right).\end{aligned}$$

We obtain that $A_n = O\left(n^{1+\eta} \log n\right)$. Hence by Proposition 1 we have $M_n = o\left(n^{\frac{1}{2}+\eta} \log n\right)$ a.e. on the event $\{A_\infty = \infty\}$, for all $\eta > 0$. Therefore $M_n = o(n)$ holds almost surely, and this completes the proof of Lemma 1. $\square$

Proof of Lemma 2. Fix an arbitrary $d \ge m$. Lemma 1 and the induction hypothesis imply that

$$\frac{E\left(Y[n,d] \mid \mathcal{F}_n\right)}{X[n,d]} = \frac{S[n,d] - S[n-1,d]}{X[n,d]} \to 0 \quad (n \to \infty).$$

Thus in (12) we can apply the approximation $1 - x = e^{-x + O\left(x^2\right)}$ $(x \to 0)$. Set

$$a_i = \frac{S[i,d] - S[i-1,d]}{c_d}, \quad b_i = \frac{c_d}{X[i,d]}$$

if $X[i,d] > 0$, and $a_i = b_i = 0$ otherwise. Then a_i and b_i are nonnegative. In addition,

$$\frac{1}{n} \sum_{i=1}^{n} a_i = \frac{1}{c_d n} \sum_{i=1}^{n} \left(S[i,d] - S[i-1,d]\right) = \frac{S[n,d]}{c_d n} \to \frac{k_d}{c_d}$$

as $n \to \infty$, by Lemma 1. According to the induction hypothesis, $X[n,d] \sim c_d \cdot n$, which implies that $n b_n \to 1$ as $n \to \infty$. Therefore Proposition 6 applies to the sequences (a_n) and (b_n) with $K = k_d / c_d$. Thus, due to part a),

$$\exp\left(\sum_{i=1}^{n} a_i b_i\right) = \exp\left(\sum_{i=1}^{n} \frac{S[i,d] - S[i-1,d]}{X[i,d]}\right)$$

is regularly varying with exponent K.

The remainder terms produce a slowly varying function, because by part b) of Proposition 6 we get that

$$\exp\left(\sum_{i=1}^{n} a_i^2 b_i^2 s_i\right) = \exp\left(\sum_{i=1}^{n}\left(\frac{S\,[i,d]-S\,[i-1,d]}{X\,[i,d]}\right)^2 s_i\right)$$

is slowly varying supposed the sequence (s_i) is bounded.

From these the asymptotics of $c\,[n,d]$ readily follows. □

Proof of Lemma 3. By (15), $A[n,d] = A_1 + A_2$, where

$$A_1 = \sum_{i=1}^{n} c\,[i+1,d-1]\, Z^*\,[i,d-1]\,\frac{E\left(Y\,[i,d-1]|\,\mathcal{F}_i\right)}{X\,[i,d-1]}\,,$$

$$A_2 = \sum_{i=1}^{n} c\,[i+1,d-1]\, E\left(J^*\,[i+1,d]|\,\mathcal{F}_i\right).$$

Here we already know the asymptotics of $S\,[n,d-1]$ and $c\,[n,d-1]$ from Lemmas 1 and 2. In addition, $Z^*\,[n,d-1] \sim z_{d-1}\,|S_n| \sim z_{d-1}\zeta_n n^\alpha$ a.s., due to the induction hypothesis and Condition 9. It is clear from the definition that $Y\,[n,d-1]$ is nonnegative, thus we have

$$\begin{aligned}A_1 &\sim \sum_{i=1}^{n} a\,[i+1,d-1]\; i^{k_{d-1}/c_{d-1}}\, z_{d-1}\,\zeta_i\, i^\alpha\,\frac{1}{c_{d-1}i}\, E\left(Y\,[i,d-1]|\,\mathcal{F}_i\right)\\ &= \frac{z_{d-1}}{c_{d-1}}\sum_{i=1}^{n} a\,[i+1,d-1]\;\zeta_i\, i^{k_{d-1}/c_{d-1}+\alpha-1}\, E\left(Y\,[i,d-1]|\,\mathcal{F}_i\right).\end{aligned}$$

Let us apply part a) of Proposition 5 in the following setting.

$$\alpha_n = a\,[n+1,d-1]\,\zeta_n, \quad \beta_n = E\left(Y\,[n,d-1]|\,\mathcal{F}_n\right)$$

for $n \ge 1$;

$$\delta = \frac{k_{d-1}}{c_{d-1}} + \alpha - 1.$$

Condition 9 and Lemma 2 guarantee that α_n is slowly varying. Furthermore, $\sum_{i=1}^{n}\beta_i = S\,[n,d-1] \sim k_{d-1}n$ as $n \to \infty$ with $k_{d-1} > 0$, hence $\gamma_n = k_{d-1}$ and $\mu = 1$ satisfy the conditions. Finally, $\mu + \delta = k_{d-1}/c_{d-1} + \alpha > 0$, because c_{d-1}, k_{d-1} and α are positive due to Conditions 1, 6, and 9.

Applying Proposition 5 we obtain that

$$A_1 \sim \frac{z_{d-1}k_{d-1}}{c_{d-1}}\cdot\frac{1}{\alpha+\frac{k_{d-1}}{c_{d-1}}}\cdot a\,[n,d-1]\;\zeta_n\, n^{k_{d-1}/c_{d-1}+\alpha}$$

almost surely as $n \to \infty$, where z_{d-1}, k_{d-1} and c_{d-1} are positive.

Now we examine the second term in $A[n,d]$. Since

$$J^*[i+1,d] = I^*(i+1) - \sum_{j=m}^{d-1} I^*[i+1,j],$$

we have

$$E\left(J^*\left[i+1,d\right]\middle|\,\mathcal{F}_i\right)=E\left(I^*(i+1)\middle|\,\mathcal{F}_i\right)-\sum_{j=m}^{d-1}E\left(I^*[i+1,j]\middle|\,\mathcal{F}_i\right).$$

Hence by Lemma 2

$$A_2\sim\sum_{i=1}^{n}a\left[i+1,d-1\right]i^{k_{d-1}/c_{d-1}}\times$$
$$\times\left(E\left(I^*\left(i+1\right)\middle|\,\mathcal{F}_i\right)-\sum_{j=m}^{d-1}E\left(I^*\left[i+1,j\right]\middle|\,\mathcal{F}_i\right)\right).$$

Set $\alpha_n=a\left[n+1,d-1\right]$, $\delta=k_{d-1}/c_{d-1}$, and $\beta_n=E\left(I^*\left(n+1\right)\middle|\,\mathcal{F}_n\right)$. By Proposition 3 and Condition 9 we have

$$\sum_{i=1}^{n}\beta_i\sim\sum_{i=1}^{n}I^*\left(i+1\right)=\left|S_{n+1}\right|-I^*\left(1\right)\sim\zeta_n n^{\alpha}\quad\left(n\to\infty\right).\tag{21}$$

Thus we can apply part a) of Proposition 5 with $\mu=\alpha>0$. Assumption $\delta+\mu>0$ is satisfied. Therefore we get that

$$\sum_{i=1}^{n}c\left[i+1,d-1\right]I^*\left(i+1\right)\sim\frac{\alpha}{\alpha+\frac{k_{d-1}}{c_{d-1}}}\cdot a\left[n+1,d-1\right]\,\zeta_n\,n^{\alpha+k_{d-1}/c_{d-1}}$$

almost surely as $n\to\infty$.

On the other hand, for a fixed $j\le d-1$ we have

$$\sum_{i=1}^{n}c\left[i+1,d-1\right]E\left(I^*\left[i+1,j\right]\middle|\,\mathcal{F}_i\right)$$
$$\sim\sum_{i=1}^{n}a\left[i+1,d-1\right]i^{k_{d-1}/c_{d-1}}E\left(I^*\left[i+1,j\right]\middle|\,\mathcal{F}_i\right)$$

by Lemma 2. In this case α_n remains the same as before, and we set $\beta_n=E\left(I^*\left[n+1,j\right]\middle|\,\mathcal{F}_n\right)$. Using Condition 10 and equation (21) we obtain that

$$\sum_{i=1}^{n}\beta_i=\sum_{i=1}^{n}E\left(I^*\left[i+1,j\right]\middle|\,\mathcal{F}_i\right)$$
$$=\left(q_j+o\left(1\right)\right)\sum_{i=1}^{n}E\left(I^*\left(i+1\right)\middle|\,\mathcal{F}_i\right)=\left(q_j+o\left(1\right)\right)\zeta_n n^{\alpha}$$

almost surely as $n\to\infty$. Thus we can apply part a) or part b) of Proposition 5 with $\mu=\alpha$, according that q_j vanishes or it is positive.

Then we get that

$$\sum_{i=1}^{n} c\left[i+1,d-1\right] E\left(I^*\left[i+1,j\right]\middle|\,\mathcal{F}_i\right)$$
$$= \frac{\alpha q_j + o\left(1\right)}{\alpha + \frac{k_{d-1}}{c_{d-1}}} \cdot a\left[n,d-1\right]\zeta_n n^{\alpha + k_{d-1}/c_{d-1}}$$

almost surely as $n\to\infty$. Hence we conclude that

$$A_2 \sim \left(1 - \sum_{j=m}^{d-1} q_j + o\left(1\right)\right) \frac{\alpha}{\alpha + \frac{k_{d-1}}{c_{d-1}}} \cdot a\left[n,d-1\right]\,\zeta_n n^{\alpha + k_{d-1}/c_{d-1}} \tag{22}$$

almost surely as $n\to\infty$. Since (q_d) is a probability distribution by Condition 10, it follows that $\left(1-\sum_{j=m}^{d-1} q_j\right) = \sum_{j=d}^{\infty} q_j$. This completes the proof. $\square$

Proof of Lemma 4. From equation (14) it follows that

$$B\left[n,d\right] = \sum_{i=2}^{n} \mathrm{Var}\left(V\left[i,d\right]\middle|\,\mathcal{F}_{i-1}\right) = \sum_{i=2}^{n} c\left[i,d-1\right]^2 \mathrm{Var}\left(Z^*\left[i,d\right]\middle|\,\mathcal{F}_{i-1}\right).$$

By equation (7) we have

$$\begin{aligned}\mathrm{Var}\left(Z^*\left[i,d\right]\middle|\,\mathcal{F}_{i-1}\right) &\le E\left(\left(Z^*\left[i,d\right]-Z^*\left[i-1,d\right]\right)^2\middle|\,\mathcal{F}_{i-1}\right)\\ &= E\left(\left(Y^*\left[i-1,d-1\right]+J^*\left[i,d\right]\right)^2\middle|\,\mathcal{F}_{i-1}\right)\\ &\le 2E\left(Y^*\left[i-1,d-1\right]^2\middle|\,\mathcal{F}_{i-1}\right) + 2E\left(J^*\left[i,d\right]\middle|\,\mathcal{F}_{i-1}\right).\end{aligned}$$

Hence

$$\begin{aligned}B\left[n,d\right] &\le 2\sum_{i=2}^{n} c[i,d-1]^2\, E\left(Y^*\left[i-1,d-1\right]^2\middle|\,\mathcal{F}_{i-1}\right)\\ &\quad + 2\sum_{i=2}^{n} c[i,d-1]^2\, E\left(J^*[i,d]\middle|\,\mathcal{F}_{i-1}\right) = 2B_1 + 2B_2.\end{aligned} \tag{23}$$

We will estimate B_1 and B_2 separately.

Similarly to the proof of Lemma 1, fix a positive $\varepsilon>0$ such that $\kappa = E\left(e^{\varepsilon Z_{d-1}}\right) < \infty$, and set $C_i = \frac{2}{\varepsilon}\log i$. Using Condition 5 and inequality $Y^*\left[i,d\right] \le Y\left[i,d\right]$ one can see that

$$\begin{aligned}E\left(Y^*\left[i-1,d-1\right]^2\middle|\,\mathcal{F}_{i-1}\right) &\le C_i E\left(Y^*\left[i-1,d-1\right]\middle|\,\mathcal{F}_{i-1}\right)\\ &\quad + E\left(Z_{d-1}^2 I\left(Z_{d-1} > C_i\right)\right)\end{aligned}$$

holds. For estimating the first term on the right-hand side we make use of equation (9).

$$E\left(Y^*\left[i-1,d-1\right]\middle|\,\mathcal{F}_{i-1}\right) = \frac{E\left(Y\left[i-1,d-1\right]\middle|\,\mathcal{F}_{i-1}\right)}{X\left[i-1,d-1\right]}\,X^*\left[i-1,d-1\right] \le \frac{E\left(Y\left[i-1,d-1\right]\middle|\,\mathcal{F}_{i-1}\right)}{X\left[i-1,d-1\right]}\left|S_{i-1}\right|.$$

To the second term we can apply (20); it is $O\left((\log i)^2 i^{-2}\right)$.

From all these we obtain that

$$B_1 \le \sum_{i=2}^{n} c[i,d-1]^2 C_i\,\frac{E\left(Y[i-1,d-1]\middle|\,\mathcal{F}_{i-1}\right)}{X[i-1,d-1]}\left|S_{i-1}\right| + O\left(\sum_{i=2}^{n} c[i,d-1]^2(\log i)^2 i^{-2}\right)$$

Note that the second sum is convergent here. In the first sum $c[i,d-1]$ can be estimated by Lemma 2, $|S_{i-1}|$ by Condition 9, and $X[i-1,d-1]$ by Condition 1. In this way we obtain that

$$c[i,d-1]^2\,C_i\,\frac{|S_{i-1}|}{X[i-1,d-1]}$$

is regularly varying with exponent $\delta = 2k_{d-1}/c_{d-1}+\alpha-1$. On the other hand, by Lemma 1 the sum of $E\left(Y[i-1,d-1]\middle|\,\mathcal{F}_{i-1}\right)$ is regularly varying with exponent 1. Therefore part a) of Proposition 5 implies that

$$B_1 = O\left(a[n,d-1]^2(\log n)^2\zeta_n\,n^{\alpha+2k_{d-1}/c_{d-1}}\right).$$

For B_2 let us apply part a) of Proposition 5 with $\alpha_i = c[i,d-1]$ and $\beta_i = c[i,d-1]E\left(J^*[i,d]\middle|\,\mathcal{F}_{i-1}\right)$. The regular variation of $\sum\beta_i$ has already been proven in (22). Thus,

$$B_2 = O\left(a[n,d-1]^2\,\zeta_n\,n^{\alpha+2k_{d-1}/c_{d-1}}\right).$$

Returning to (23) we conclude that

$$B\left[n,d\right] = O\left(n^{\alpha+2k_{d-1}/c_{d-1}+\eta}\right)$$

for all $\eta > 0$. Consequently,

$$B\left[n,d\right]^{1/2}\log B\left[n,d\right] = O\left(n^{\alpha/2+k_{d-1}/c_{d-1}+\eta}\right)\quad(n\to\infty).$$

Now the proof can be completed by comparing this with Lemma 3. □

Remark 3. *Since $S\left[n,d\right]$ is clearly nonnegative, Lemma 1 implies that $k_d \ge 0$ for all $d \ge m$. This means that*

$$\sum_{j=0}^{d} c_j \ge \sum_{j=0}^{d} p_j \quad (d\ge m).$$

Loosely speaking, the degree of a typical vertex is asymptotically larger than or equal to the degree of the new vertex. This is in accordance with the fact that the degree of a fixed vertex cannot decrease.

Similarly to Lemma 1, one can prove that

$$\sum_{j=0}^{d} x_j \geq \sum_{j=0}^{d} q_j \quad (d \geq m)\,,$$

which means the same for the selected vertices.

6. Graph models

In this section we briefly review some scale free random graph models and sets of selected vertices to which the results of the previous section can be applied.

6.1. **Generalized plane oriented recursive tree.** We start from one edge, and at each step one new vertex and one new edge are added to the graph. At the nth step the probability that a given vertex of degree d is connected to v_n is $(d+\beta)/T_{n-1}$, where $\beta > -1$ is the parameter of the model, and $T_{n-1} = (2+\beta)(n+1)+\beta$. These kind of random trees are widely examined, see for example [5, 15]. $\beta = 0$ gives the Albert–Barabási tree [1].

We fix an integer $j \geq 1$. At the nth step v_n is added to the set of selected vertices if it is at distance j from u_1 in G_n. Thus S_n is the jth level of the tree G_n.

It is well known [9] that Condition 1 is satisfied with

$$c_d = \frac{(2+\beta)\,\Gamma(d+\beta)\,\Gamma(3+2\beta)}{\Gamma(1+\beta)\,\Gamma(d+3+2\beta)} \quad (d \geq 1)\,.$$

Consequently,

$$c_d \sim \frac{(2+\beta)\,\Gamma(3+2\beta)}{\Gamma(1+\beta)} \cdot d^{-(3+\beta)} \quad (d \to \infty)$$

and $\gamma = 3+\beta$ satisfies Condition 2. It is clear from the definition that Condition 3 holds, and since the degree of the new vertex is always 1, we have $m = 1$, and conditions 4, 5, and 10 are trivially satisfied. Using that $p_d = 0$ for $d \neq 1$ and $p_1 = 1$, Condition 6 is also easy to check.

The distance of v_n and u_1 does not change after generating the edges from v_n at the nth step. This guarantees Conditions 7 and 8. The results of [10] show that Condition 9 is satisfied with $\alpha = 1/(2+\beta)$. It is proven that

$$|S_n| \sim n\zeta \frac{\mu(n)^{j-1}}{(j-1)!}\, e^{-\mu(n)} \asymp n^{\frac{1}{2+\beta}} (\log n)^{j-1} \quad (n \to \infty)\,,$$

where ζ is a positive random variable, and $\mu(n) = \frac{1+\beta}{2+\beta}\log n$ [10, Theorem 2.1].

Thus Theorem 1 applies: the asymptotic degree distribution constrained on a fixed level of the tree does exist. The new characteristic exponent is the following (cf. [10, Theorem 3.1]).

$$\gamma^* = \alpha(\gamma - 1) + 1 = \frac{1}{2+\beta}(3+\beta-1) + 1 = 2.$$

6.2. **Independent edges.** We start from one edge. At the nth step, independently of each other, every old vertex is connected to the new one with probability $\lambda d/T_{n-1}$, where d is the degree of the old vertex in G_{n-1}, $0 < \lambda < 2$ is a fixed parameter, and T_{n-1} denotes the sum of degrees in G_{n-1}. The restriction on λ guarantees that the probability given above belongs to $[0, 1]$. It is clear that $m = 0$.

We fix one vertex, v, and S_n consists of its neighbours in G_n.

In [8, Theorem 3.1.] it is proven that the asymptotic degree distribution is given by

$$c_0 = p_0, \quad c_d = \frac{2}{d(d+1)(d+2)} \sum_{k=1}^{d} k(k+1)\, p_k,$$

where

$$p_k = \frac{\lambda^k}{k!}\, e^{-\lambda}.$$

Clearly, $c_d \sim 2\lambda(2+\lambda)\, d^{-3} \quad (d \to \infty)$. Thus the first two conditions are satisfied, and $\gamma = 3$. Condition 3 holds, because the probability that a given vertex gets a new edge depends only on its actual degree. It is also clear that Conditions 7 and 8 hold. Condition 9 is a corollary of [11, Theorem 2.1], and we have $\alpha = 1/2$.

In this case the initial degree of the new vertex is not fixed. It is proven in [8] that

$$\sum_{k=0}^{\infty} \left| E\left(I[n+1, d] \,|\, \mathcal{F}_n \right) - p_d \right| \to 0$$

almost surely as $n \to \infty$. This, and the fact that (p_d) is a Poisson distribution with parameter λ imply Condition 4.

Note that the conditional distribution of $Y[n, d]$ is binomial of order $X[n, d]$ and parameter $\lambda d/T_n < 1$. One can check Condition 5 with Z_d having a suitable Poisson distribution.

Condition 10 can be verified basing on the fact that the degree distribution of a new selected vertex is similar to the distribution of a new vertex because of the independent random choices, and the following results. Theorem 2.1 in [12] states that $T_n = 2\lambda n + o(n^{1-\varepsilon})$ almost surely if $\varepsilon > 0$ is sufficiently small. Moreover, Theorem 2.2 there implies that the maximum degree after n steps is $O(\sqrt{n})$ almost surely.

Our Theorem 1 can be applied, so the almost sure asymptotic degree distribution constrained on the neighbours of a fixed vertex exists. The new characteristic exponent is given by

$$\gamma^* = \alpha\,(\gamma - 1) + 1 = \frac{1}{2}\,(3 - 1) + 1 = 2$$

(cf. [11, Theorem 3.1]).

Let us modify this example in such a way that vertices of degree 1 never get new edges. Let

$$T_{n-1} = \sum_{d=2}^{n} X[n-1, d]\, d,$$

and choose S_n to contain all vertices of degree 1. Then we can see that the all conditions hold except Condition 6, but $x_d = 0$ for $d > 1$. This shows that positivity of k_d cannot be relaxed in order to obtain a polynomially decreasing degree distribution.

6.3. **Random multitrees.** For $M \geq 2$ an M-multicherry is a hypergraph on $M + 1$ vertices. One of them, called center, is distinguished, it is connected to all other vertices with ordinary edges (2-hyperedges), and the remaining M vertices form an M-hyperedge, called the base.

We start from the complete graph of M vertices; the vertices form a base. Then at each step we add a new vertex and an M-multicherry with the new vertex in its center. We select the base of the new multicherry from the existing bases uniformly. Finally, we add M new bases by replacing a vertex in the selected base with the new center in all possible ways.

The degree of the new vertex is always M, thus $m = M$.

Let S_n be the set of vertices that are at distance j from the initial configuration.

It is shown in [13] that Conditions 1, 2, and 9 are satisfied with $\gamma = 2 + \frac{1}{M-1}$ and $\alpha = \frac{M-1}{M}$. The other conditions are easy to check, using that distances in the multitree do not change.

Therefore Theorem 1 applies, and

$$\gamma^* = \alpha\,(\gamma - 1) + 1 = \frac{M-1}{M}\left(2 + \frac{1}{M-1} - 1\right) + 1 = 2.$$

Another option for the set of selected vertices is the following. Fix an integer $1 \leq k < M$ and k different vertices. Let S_n be the set of vertices that are connected to all of them. Since the model is the same, we only have to check the conditions on the set of selected vertices. Now Conditions 7, 8, and 10 clearly hold. Condition 9 can be proven by slight modifications of the proofs of [13]. In this case $\gamma = 2 + \frac{1}{M-1}$, $\alpha = 1 - \frac{k}{M}$, and

$$\gamma^* = 2 - \frac{k-1}{M-1} > 1.$$

7. Conclusions

We presented sufficient conditions for the existence of the asymptotic degree distribution constrained on the set of selected vertices. Scale free property and regular variation of the size of the set of selected vertices were essential. The new characteristic exponent depended only on γ and α.

We reviewed several models satisfying these conditions and identified their characteristic exponents applying our main result. In these models $\gamma^* \le 2$ and $\gamma^* \le \gamma$, thus the characteristic exponent decreased. One reason for that is the following. The selected vertices are closer to the initial configuration in some sense. There are more "old" vertices among them and their degree is larger than that of the "typical" ones.

References

[1] A-L. Barabási, R. Albert, *Emergence of scaling in random networks*, Science **286** (1999), 509–512.

[2] N. H. Bingham, *Regular variation,* Cambridge University Press, 1987.

[3] R. Bojanić, E. Seneta, *Slowly varying functions and asymptotic relations*, J. Math. Anal. Appl. **34** (1971), 302–315.

[4] B. Bollobás, *Random graphs*, Cambridge University Press, 2001.

[5] M. Drmota, *Random trees*, Springer-Verlag/Wien, 2009.

[6] R. Durrett, *Random graph dynamics*, Cambridge University Press, 2006.

[7] J. Galambos, E. Seneta, *Regularly varying sequences*, Proc. Amer. Math. Soc. **41** (1973), 110–116.

[8] Zs. Katona, T. F. Móri, *A new class of scale free random graphs*, Statist. Probab. Lett. **76** (2006), 1587–1593.

[9] T. F. Móri, *On random trees*, Studia Sci. Math. Hungar. **39** (2002), 143–155.

[10] T. F. Móri, *A surprising property of the Barabási–Albert random tree*, Studia Sci. Math. Hungar. **43** (2006), 265–273.

[11] T. F. Móri, *Degree distribution nearby the origin of a preferential attachment graph*, Electron. Comm. Probab. **12** (2007), 276–282.

[12] T. F. Móri, *On a 2-parameter class of scale free random graphs*, Acta Math. Hungar. **114** (2007), 37–48.

[13] T. F. Móri, *Random multitrees*, Studia Sci. Math. Hungar. **47** (2010), 59–80.

[14] J. Neveu, *Discrete-parameter martingales*, North-Holland, Amsterdam, 1975.

[15] B. Pittel, *Note on the heights of random recursive trees and random m-ary search trees*, Random Struct. Algorithms **5** (1994), 337–348.

Department of Probability Theory and Statistics, Faculty of Science, Eötvös Loránd University, Pázmány P. s. 1/C, H-1117 Budapest, Hungary

E-mail address: agnes@cs.elte.hu

Department of Probability Theory and Statistics, Faculty of Science, Eötvös Loránd University, Pázmány P. s. 1/C, H-1117 Budapest, Hungary

E-mail address: moritamas@ludens.elte.hu