Towards graphs compression: The degree distribution of duplication-divergence graphs

Alan Frieze

Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA, USA
alan@random.math.cmu.edu

Krzysztof Turowski

Center for Science of Information, Department of Computer Science, Purdue University, West Lafayette, IN, USA
krzysztof.szymon.turowski@gmail.com

Wojciech Szpankowski

Center for Science of Information, Department of Computer Science, Purdue University, West Lafayette, IN, USA
spa@cs.purdue.edu

Abstract

We present a rigorous and precise analysis of the degree distribution in a dynamic graph model introduced by Pastor-Satorras et al. in which nodes are added according to a duplication-divergence mechanism, i.e. by iteratively copying a node and then randomly inserting and deleting some edges for a copied node. This graph model finds many applications in the real world from biology to social networks. It is discussed in numerous publications with only very few rigorous results, especially for the degree distribution.

In this paper we focus on two related problems: the expected value and large deviation for the degree of a given node over the evolution of the graph and the expected value and large deviation of the average degree in the graph. We present exact and asymptotic results showing that both quantities may decrease or increase over time depending on the model parameters. Our findings are a step towards a better understanding of aspects of the graph behavior such as degree distribution, symmetry—that eventually will lead to structural compression, an important open problem in this area.

2012 ACM Subject Classification Mathematics of computing → Random graphs; Theory of computation → Random network models

Keywords and phrases dynamic graphs, duplication-divergence graphs, degree distribution, large deviation

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23

Funding This work was supported by NSF Center for Science of Information (CSoI) Grant CCF-0939370, and in addition by NSF Grant CCF-1524312, and National Science Center, Poland, Grant 2018/31/B/ST6/01294.
This work was also supported by NSF Grant DMS1661063.

1 Introduction

On the one hand, it is widely accepted that we live in the age of data deluge. On a daily basis we observe the increasing availability of data collected and stored in various forms, as sequences, expressions, interactions or structures. A large part of this data is given in a complex form which conveys also a "shape" of the structure, such as network data. Examples are various biological networks, social networks or Web graphs.

On the other hand, compression is a well-known area of information theory which mostly deals with the compression of *sequences*. Yet, we note that already in 1953 Shannon argued as to the importance of extending the theory to data without a linear structure, such as lattices [17]. Recently, we saw some work directed towards more complex data structures such as trees [10, 16] and graphs [5, 3, 13]. Compression for such non-conventional types of data has become an important issue, since e.g. graph data are nowadays widely used in Big Data computing [11]. It is therefore an imperative to provide efficient storage and processing to speed up computations and lower memory and hardware costs.

The recent survey by Besta and Hoefler [4] collected over 450 papers concerned with the topic of lossless graph compression. There were several well-known heuristics proposed for the compression of real-world graphs, such as the algorithm by Adler and Mitzenmacher [2] devised for the Web graph. But the first rigorous analysis of an asymptotically optimal algorithm for Erdős-Renyi graphs was presented in [5], while recently it was extended to the preferential attachment model (also known as Barábasi-Albert) graphs [14]. However, many real-world networks such as protein-protein and social networks follow a different model of generation known as the *duplication-divergence* model in which new nodes are added to the network as copies of existing nodes together with some random divergence, resulting in differences among the original nodes and their copies. In this paper we focus on analyzing the degree distribution – a first step towards graph compression – in such a network, which we first define more precisely.

Consider the most popular duplication-divergence model as introduced by Pastor-Satorras et al. [18], referred to below as $\mathtt{DD}(t, p, r)$. It is defined as follows: starting from a given graph on t_0 vertices (labeled from 1 to t_0) we add subsequent vertices labeled t_0, $t_0 + 1$, ..., t as copies of some existing vertices in the graph and then we introduce divergence by adding and removing some edges connected to the new vertex independently at random. Finally, we remove the labels and return the structure, i.e. the unlabeled graph.

In order to pursue compression and other algorithms (e.g., finding the node arrivals) for duplication-divergence model we need to observe [5, 13] the close affinity between (structural) compression and symmetries of the graph. In turn, graph symmetries (motivated further below), are closely related to the degree distribution, which is the main topic of this paper. Indeed, as discussed in [13] a graph is asymmetric if two properties hold: (i) new nodes do not make the same choices among old nodes, and (ii) old nodes have *distinct* degrees. Thus the degree distribution plays a crucial role in many graph algorithms including graph compression and others (e.g., inferring node arrival in such dynamic networks [15]).

Before we summarize our main results on the degree distribution in $\mathtt{DD}(t, p, r)$ networks, let us explore further the connection between compression and graph symmetries. The linking concepts here are the *graph entropy* $H(G)$ (also known as the labeled graph entropy) and *structural graph entropy* $H(S(G))$ (also known as the *unlabeled graph* entropy). Both quantities depend deeply on the degree distribution. Let $\mathcal{G}_n$ be the set of all labeled graphs on n vertices (with vertices having labels 1, 2, ..., n) and $\mathcal{S}_n$ be the set of all unlabeled graphs on n vertices. Then, the graph entropy and the structural graph entropy are defined as

$$H(G) = \sum_{G \in \mathcal{G}_n} \Pr[G] \log \Pr[G],$$

$$H(S(G)) = \sum_{S(G) \in \mathcal{S}_n} \Pr[S(G)] \log \Pr[S(G)],$$

where $S(G)$ is the *structure* of graph G, that is, the graph G with labels removed.

It turns out that for many well-known random graph models, the structural graph entropy can be expressed by a following formula:

$$H(G) - H(S(G)) = \mathbb{E} \log |\mathrm{Aut}(G)| - \mathbb{E} \log |\Gamma(G)|$$

where $H(G)$ and $H(S(G))$ are, respectively, the entropy of the labelled and unlabelled graph generated by a given model, $\mathrm{Aut}(G)$ is the automorphism group of the graph G (representing graph symmetries) and $\Gamma(G)$ is the set of all re-labelings of G that give a graph which can be generated by the given graph model with positive probability [13].

In fact, many real-world networks, such as protein-protein and social networks, have been shown to contain lots of symmetries, as presented in Table 1. This is in stark contrast to the Erdős-Renyi and preferential attachment models, as both generate completely asymmetric graphs with high probability , that is $\log |\mathrm{Aut}(G)| = 0$ [5, 13], and therefore we do not consider these models as likely matches for these kinds of networks.

Network	Nodes	Edges	$\log \lvert\mathrm{Aut}(G)\rvert$
Baker's yeast protein-protein interactions	6,152	531,400	546
Fission yeast protein-protein interactions	4,177	58,084	675
Mouse protein-protein interactions	6,849	18,380	305
Human protein-protein interactions	17,295	296,637	3026
ArXiv high energy physics citations	7,464	116,268	13
Simple English Wikipedia hyperlinks	10,000	169,894	1019
CollegeMsg online messages	1,899	59,835	232

Table 1 Symmetries of the real-world networks [19, 22].

Consequently, in order to study and understand the behavior of real-world networks we need dynamic graph models that naturally generate internal graph symmetries. It turns out that the discussed duplication-divergence model is such a candidate. However, at the moment there do not exist any rigorous general results on symmetries for such graphs. Experimentally, when generating multiple graphs from this model with different parameters, we observe the pattern presented in Figure 1: there is a large set of parameters for which the generated graphs are highly symmetric, as exhibited by the size of their automorphisms group (expressed in a logarithmic scale), $\log |\mathrm{Aut}(G)|$. Moreover, as it was shown by Sreedharan et al. [19], the possible values of the parameters for real-world networks under the assumption that they were generated by this model lie in the blue-violet area, indicating a lot of symmetry.

In view of these, it is imperative that we understand symmetry and degree distribution in duplication-divergence networks. Overall, both questions are tightly related, as already discussed above. We note that in the previous work on various graph models, such as preferential attachment [13], the analysis of the degree distribution was a vital step in proving results on structural compression. For this, as discussed in [13], we need to study the average and large deviation of their degree sequence, which is the main topic of this conference paper.

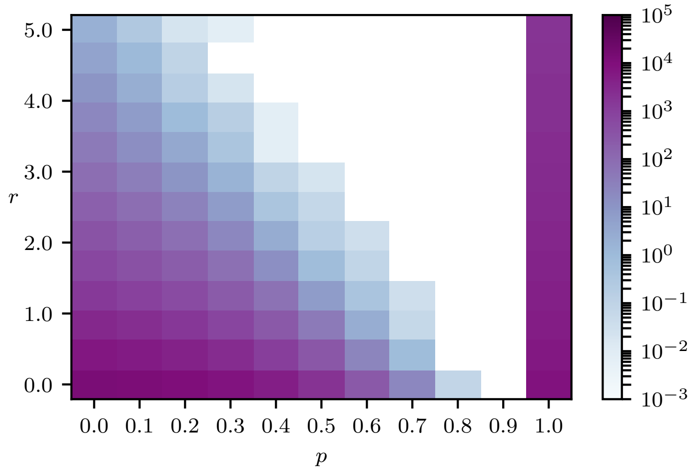

Figure 1 Symmetry of graphs ($\log|\mathrm{Aut}(G)|$) generated by Pastor-Satorras model.

Turowski et al. showed in [21] that for the special case of $p = 1$, $r = 0$ the expected logarithm of the number of automorphisms for graphs on t vertices is asymptotically $\Theta(t \log t)$, which indicates a lot of symmetry. Therefore, they were able to obtain asymptotically optimal compression algorithms for graphs generated by such models. However, their approach used certain properties of the model which cannot be applied for different parameter values.

For $r = 0$ and $p < 1$, it was recently proved by Hermann and Pfaffelhuber in [7] that depending on value of p either there exists a limiting distribution of degree frequencies with almost all vertices isolated or there is no limiting distribution as $t \to \infty$. Moreover, it is shown in [12] that the number of vertices of degree one is $\Omega(\ln t)$ but again the precise rate of growth of the number of vertices with degree $k > 0$ is as yet unknown. Recently, also for $r = 0$, Jordan [9] showed that the *non-trivial connected component* has a degree distribution which conforms to a power-law behavior, but only for $p < e^{-1}$. In this case the exponent is equal to γ which is the solution of $3 = \gamma + p^{\gamma-2}$.

In this paper we approach the problem of the degree distribution from a different perspective. We focus on presenting exact and precise asymptotic results for the expected degree and large deviations of a given vertex s at time t (denoted by $\deg_t(s)$) and the average degree in the graph (denoted by $D(G_t)$).

We discuss in Theorems 2–7 exact and precise asymptotics of these quantities when $t \to \infty$. We show that $\mathbb{E}[\deg_t(s)]$ and $\mathbb{E}[D(G_t)]$ exhibit phase transitions over the parameter space: as a function of p and r. In particular, we find that $\mathbb{E}[\deg_t(s)]$ grows respectively like $\left(\frac{t}{s}\right)^p$, $\sqrt{\frac{t}{s}} \log s$ or $\left(\frac{t}{s}\right)^p s^{2p-1}$, depending whether $p < \frac{1}{2}$, $p = \frac{1}{2}$ or $p > \frac{1}{2}$. Furthermore, $\mathbb{E}[D(G_t)]$ is either $\Theta(1)$, $\Theta(\log t)$ or $\Theta(t^{2p-1})$ for the same ranges of p. We also determine the exact constants for the leading terms that strictly depend on p, r, t_0 and the structure of the seed graph G_{t_0}. This confirms the empirical findings of [8] regarding the seed graph influence on the structure of G_t.

We also present some results concerning the the tail of the asymptotic distribution of the variables $D(G_t)$ and $\deg_t(s)$ for $s = O(1)$. It turns out that it is sufficient to only go a polylogarithmic factor under or over the mean to obtain a polynomial tail, that is to get an $O(t^{-A})$ tail probability.

These findings allow us to better understand why the $\mathtt{DD}(t, p, r)$ model differs quite substantially from other graph models such as the preferential attachment model [13, 23]. In particular, we observe that the expected degree behaves differently as $t \to \infty$ for different values of s and p. For example, if $p > \frac{1}{2}$, then for $s = O(1)$ (that is, for very old nodes) we observe that $\mathbb{E}[\deg_s(t)] = \Omega(t^p)$ while for $s = \Theta(t)$ (i.e., very young nodes) we have $\mathbb{E}[\deg_s(t)] = O(t^{2p-1})$. This behavior is very different than the degree distribution for, say, the preferential attachment model, for which the expected degree of a vertex s in a graph on t vertices is of order $\sqrt{t/s}$ for s up to an order of t^ε for some constant $\varepsilon > 0$ [13].

We now present our main results on degree distributions. All proofs are delegated to appendices.

2 Main results

In this section we present our main results with proofs and auxiliary lemmas presented in the respective appendices.

We use standard graph notation, e.g. from [6]: $V(G)$ denotes the set of vertices of graph G, $\mathcal{N}_G(u)$ – the set of neighbors of vertex u in G, $\deg_G(u) = |\mathcal{N}_G(u)|$ – the degree of u in G. For brevity we use the abbreviations for G_t, e.g. $\deg_t(u)$ instead of $\deg_{G_t}(u)$. All graphs are simple. Let us also introduce the *average degree* $D(G_t)$ of G as

$$D(G) = \frac{1}{|V(G)|} \sum_{v \in V(G)} \deg_G(u).$$

It is also known in the literature as the first moment of the degree distribution, and it is related to the number of edges.

Formally, we define the model $\texttt{DD}(t, p, r)$ as follows: let $0 \le p \le 1$ and $0 \le r \le t_0$ be the parameters of the model. Let also G_{t_0} be a graph on t_0 vertices, with $V(G_{t_0}) = \{1, \dots, t_0\}$. Now, for every $t = t_0, t_0 + 1, \dots$ we create G_{t+1} from G_t according to the following rules:

1. add a new vertex $t + 1$ to the graph,
2. pick vertex u from $V(G_t) = \{1, \dots, t\}$ uniformly at random – and denote u as *parent*$(t + 1)$,
3. for every vertex $i \in V(G_t)$:
   a. if $i \in \mathcal{N}_t(\textit{parent}(t + 1))$, then add an edge between i and $t + 1$ with probability p,
   b. if $i \notin \mathcal{N}_t(\textit{parent}(t + 1))$, then add an edge between i and $t + 1$ with probability $\frac{r}{t}$.

We focus now on the expected value of $\deg_t(s)$, that is, the degree of node s at time t. We start with a recurrence relation for $\mathbb{E}[\deg_t(s)]$. Observe that for any $t \ge s$ we know that vertex s may be connected to vertex $t + 1$ in one of the following two cases:

- either $s \in \mathcal{N}_t(\textit{parent}(t + 1))$ (which holds with probability $\frac{\deg_t(s)}{t}$) and we add an edge between s and $t + 1$ (with probability p),
- or $s \notin \mathcal{N}_t(\textit{parent}(t + 1))$ (with probability $\frac{t-\deg_t(s)}{t}$) and we an add edge between s and $t + 1$ (with probability $\frac{r}{t}$).

From the definition presented above we directly obtain the following recurrence for $\mathbb{E}[\deg_t(s)]$:

$$\begin{aligned}
\mathbb{E}[\deg_{t+1}(s) \mid G_t] &= \left(\frac{\deg_t(s)}{t}p + \frac{t - \deg_t(s)}{t}\frac{r}{t}\right)(\deg_t(s) + 1) \\
&\quad + \left(\frac{\deg_t(s)}{t}(1 - p) + \frac{t - \deg_t(s)}{t}\left(1 - \frac{r}{t}\right)\right)\deg_t(s) \\
&= \deg_t(s)\left(1 + \frac{p}{t} - \frac{r}{t^2}\right) + \frac{r}{t}.
\end{aligned}$$

After removing the conditioning on G_t, we find:

$$\mathbb{E}[\deg_{t+1}(s)] = \mathbb{E}[\deg_t(s)]\left(1 + \frac{p}{t} - \frac{r}{t^2}\right) + \frac{r}{t}. \tag{1}$$

This recurrence falls under a general recurrence of the form

$$\mathbb{E}[f(G_{t+1}) \mid G_t] = f(G_t g_1(t) + g_2(t) \tag{2}$$

where g_1 and g_2 are given functions. As we shall see these type of recurrences occur a few times in this paper, therefore we need appropriate tools to solve it. We derive a series of lemmas (Lemma 10–15), providing exact and asymptotic behavior of $\mathbb{E}[f(G_t)]$. They are based on well-known martingale theory and they use various asymptotic properties of Euler gamma function. For convenience, the associated lemmas with their proofs were moved to Appendix A.

First, we use Lemma 10 to obtain the equation for the exact behavior of the degree of a given node s at time t:

$$\mathbb{E}[\deg_t(s)] = \mathbb{E}[\deg_s(s)]\prod_{k=s}^{t-1}\left(1 + \frac{p}{k} - \frac{r}{k^2}\right) + \sum_{j=s}^{t-1}\frac{r}{j}\prod_{k=j+1}^{t-1}\left(1 + \frac{p}{k} - \frac{r}{k^2}\right). \tag{3}$$

However, we see that to solve this recurrence we need to know the expected value of $\deg_s(s)$ for all $s > t_0$, which we tackle next.

Turning our attention to this variable we find the following lemma connecting $\mathbb{E}[\deg_t(t)]$ and the average degree $\mathbb{E}[D(G_t)]$ (see proof in Appendix B):

▶ **Lemma 1.** *For any $t \geq t_0$ it holds that*

$$\mathbb{E}[\deg_{t+1}(t+1)] = \left(p - \frac{r}{t}\right)\mathbb{E}[D(G_t)] + r.$$

It is quite intuitive that the expected degree of a new vertex behaves as if we would choose a vertex with the average degree $\mathbb{E}[D(G_t)]$ as its parent, and then copy p fraction of its edges, adding also almost r more edges to all other vertices in the graph.

Thus to complete our analysis we need to find $\mathbb{E}[D(G_t)]$, that is, the average degree of G_t. Using a similar argument to the above, we find the following recurrence for the average degree of G_{t+1}:

$$\begin{aligned}
\mathbb{E}[D(G_{t+1}) \mid G_t] &= \frac{1}{t+1}\mathbb{E}\left[\sum_{i=1}^{t+1}\deg_{t+1}(i) \mid G_t\right] \\
&= \frac{1}{t+1}\mathbb{E}\left[\sum_{i=1}^{t}\deg_t(i) + 2\deg_{t+1}(t+1) \mid G_t\right] \\
&= \frac{1}{t+1}\left(\sum_{i=1}^{t}\deg_t(i) + 2\mathbb{E}\left[\deg_{t+1}(t+1) \mid G_t\right]\right) \\
&= \frac{1}{t+1}\left(tD(G_t) + 2\mathbb{E}[\deg_{t+1}(t+1) \mid G_t]\right) = D(G_t)\left(1 + \frac{2p-1}{t+1} - \frac{2r}{t(t+1)}\right) + \frac{2r}{t+1}.
\end{aligned}$$

Therefore, after removing the conditioning on G_t:

$$\mathbb{E}[D(G_{t+1})] = \mathbb{E}[D(G_t)]\left(1 + \frac{2p-1}{t+1} - \frac{2r}{t(t+1)}\right) + \frac{2r}{t+1}. \tag{4}$$

This is again recurrence of the form (2) that we can handle in a uniform manner as discussed above.

Finally, we obtain a recurrence which does not refer to any other variable defined over G_t or G_{t+1}. We can solve this recurrence by using Lemma 10 from the next section and derive Theorem 2. The proof is given in Appendix C.

▶ **Theorem 2.** *For $G_t \sim DD(t,p,r)$ and for all $t \geq t_0$ we have*

$$\mathbb{E}[D(G_t)] = \frac{\Gamma(t+c_3)\Gamma(t+c_4)}{\Gamma(t)\Gamma(t+1)} \left(D(G_{t_0})\frac{\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)} + 2r\sum_{j=t_0}^{t-1}\frac{\Gamma(j+1)^2}{\Gamma(j+c_3+1)\Gamma(j+c_4+1)}\right),$$

where $c_3 = p + \sqrt{p^2+2r}$, $c_4 = p - \sqrt{p^2+2r}$, and $\Gamma(z)$ is the Euler gamma function.

Furthermore, asymptotically as $t \to \infty$ we find

$$\mathbb{E}[D(G_t)] = \begin{cases} \frac{2r}{1-2p}(1+o(1)) & \text{if } p < \frac{1}{2} \text{ and } r > 0,\\ 2r\ln t\,(1+o(1)) & \text{if } p = \frac{1}{2} \text{ and } r > 0,\\ t^{2p-1}\frac{\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)}(1+o(1))\times & \\ \left(D(G_{t_0}) + \frac{2rt_0\,{}_3F_2\left[{t_0+1,t_0+1,1 \atop t_0+c_3+1,t_0+c_4+1};1\right]}{t_0^2+2pt_0-2r}\right) & \text{if } p > \frac{1}{2} \text{ or } r = 0,\end{cases}$$

where $D(G_{t_0})$ is the average degree of the initial graph G_{t_0} and

$${}_3F_2\left[{a_1,a_2,a_3 \atop b_1,b_2};z\right] = \sum_{l=0}^{\infty}\frac{(a_1)_l(a_2)_l(a_3)_l}{(b_1)_l(b_2)_l}\frac{z^l}{l!}$$

is the generalized hypergeometric function with $(a)_l = a(a+1)\ldots(a+l-1)$, $(a)_0 = 1$ the rising factorial (see [1] for details).

As we see, the asymptotic behavior of $\mathbb{E}[D(G_t)]$ has a threefold characteristic: when $p < \frac{1}{2}$ and $r > 0$, the majority of the edges are not created by copying them from parents, but actually by attaching them according to the value of r. For $p = \frac{1}{2}$ and $r > 0$ we note the curious situation of a phase transition (still with non-copied edges dominating), and only if either $p > \frac{1}{2}$ or $r = 0$ do the edges copied from the parents contribute asymptotically the major share of the edges.

Finally, we turn to estimations of the tails of the distribution of $D(G_t)$. It turns out that this variable is concentrated in the sense that with probability $1 - O(t^{-A})$ it is contained only within polylogarithmic ratio from the mean.

More specifically, the right tail of the distributions may be bounded as following:

▶ **Theorem 3.** *Asymptotically for $G_t \sim DD(t,p,r)$ it holds that*

$$\Pr[D(G_t) \geq A\,C\,\log^2(t)] = O(t^{-A}) \qquad \text{for } p < \frac{1}{2},$$
$$\Pr[D(G_t) \geq A\,C\,\log^3(t)] = O(t^{-A}) \qquad \text{for } p = \frac{1}{2},$$
$$\Pr[D(G_t) \geq A\,C\,t^{2p-1}\log^2(t)] = O(t^{-A}) \qquad \text{for } p > \frac{1}{2}.$$

for some fixed constant $C > 0$ and any $A > 0$.

Similarly, we have the behavior of the left tail:

▶ **Theorem 4.** *For $G_t \sim DD(t,p,r)$ with $p > \frac{1}{2}$ asymptotically it holds that*

$$\Pr\left[D(G_t) \le \frac{C}{A} t^{2p-1} \log^{-3-\varepsilon}(t)\right] = O(t^{-A}).$$

for some fixed constant $C > 0$ and any $\varepsilon, A > 0$.

Note that since $D(G_t) = O(\log t)$ for $p \le \frac{1}{2}$, the bounds of the above form are trivial and not interesting.

Now we return to the computation of the expected values of $\mathbb{E}[\deg_t(t)]$ and $\mathbb{E}[\deg_t(s)]$. By applying Theorem 2 to Lemma 1 we obtain the following corollary.

▶ **Corollary 5.** *For all $t > t_0$ it is true that*

$$\mathbb{E}[\deg_t(t)] = (pt - p - r)\frac{\Gamma(t+c_3-1)\Gamma(t+c_4-1)}{\Gamma(t)^2}$$
$$\left(D(G_{t_0})\frac{\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)} + 2r\sum_{j=t_0}^{t-2}\frac{\Gamma(j+1)^2}{\Gamma(j+c_3+1)\Gamma(j+c_4+1)}\right) + r,$$

where c_3, c_4 are as above.

Moreover, asymptotically as $t \to \infty$ it holds that

$$\mathbb{E}[\deg_t(t)] = \begin{cases} pt^{2p-1}\frac{\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)} D(G_{t_0})(1+o(1)) & \text{if } p \le \frac{1}{2},\ r = 0, \\ \frac{r}{1-2p}(1+o(1)) & \text{if } p < \frac{1}{2},\ r > 0, \\ 2rp\ln t\,(1+o(1)) & \text{if } p = \frac{1}{2},\ r > 0, \\ \frac{\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)} pt^{2p-1}(1+o(1)) & \text{if } p > \frac{1}{2}, \\ \quad \left(D(G_{t_0}) + \frac{2rt_0}{t_0^2+2pt_0-2r}\, {}_3F_2\left[{t_0+1, t_0+1, 1 \atop t_0+c_3+1, t_0+c_4+1}; 1\right]\right) & \end{cases}$$

with the same notation as in Theorem 2.

As was mentioned above, the asymptotic expected behavior is similar to the behavior of $\mathbb{E}[D(G_t)]$.

We are finally in a position to state the exact and asymptotic expressions for $\mathbb{E}[\deg_t(s)]$. This we need to split in two parts: first, for the initial vertices of G_{t_0} ($1 \le s \le t_0$) and all other vertices ($t_0 < s < t$). Note that the first of the theorems may be derived directly from Eqn. (3), (using only lemmas from Appendix A) and the second one requires Corollary 5. For the proofs of both theorems see Appendix C.

▶ **Theorem 6.** *For all $1 \le s \le t_0$ it is true that*

$$\mathbb{E}[\deg_t(s)] = \frac{\Gamma(t+c_1)\Gamma(t+c_2)}{\Gamma(t)^2}$$
$$\left[\deg_{t_0}(s)\frac{\Gamma(t_0)^2}{\Gamma(t_0+c_1)\Gamma(t_0+c_2)} + r\sum_{j=t_0}^{t-1}\frac{\Gamma(j)\Gamma(j+1)}{\Gamma(j+c_1+1)\Gamma(j+c_2+1)}\right],$$

where $c_1 = \frac{p+\sqrt{p^2+4r}}{2}$, $c_2 = \frac{p-\sqrt{p^2+4r}}{2}$, c_3 and c_4 as above.

Asymptotically as $t \to \infty$:

$$\mathbb{E}[\deg_t(s)] = \begin{cases} r \ln t\,(1+o(1)) & \text{if } p=0 \text{ and } r>0,\\ t^p \Big[\deg_{t_0}(s) \frac{\Gamma(t_0)^2}{\Gamma(t_0+c_1)\Gamma(t_0+c_2)} \\ \quad + \frac{r\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(t_0+c_1+1)\Gamma(t_0+c_2+1)} {}_3F_2\left[{t_0, t_0+1, 1 \atop t_0+c_1+1, t_0+c_2+1}; 1\right]\Big] \\ \quad (1+o(1)) & \text{if } p>0 \text{ or } r=0. \end{cases}$$

Here we observe only two regimes. In the first, for the case when $p = 0$, when edges are added mostly due to the parameter r, we have logarithmic growth of $\mathbb{E}[\deg_t(s)]$. In the second one, edges attached to s accumulate mostly by choosing vertices adjacent to s as parents of the new vertices, and therefore the expected degree of s grows proportionally to t^p.

▶ **Theorem 7.** *For all $t_0 < s < t$ it is true that*

$$\mathbb{E}[\deg_t(s)] = \frac{\Gamma(t+c_1)\Gamma(t+c_2)}{\Gamma(t)^2}$$
$$\Bigg[(ps-p-r)\,\frac{\Gamma(s+c_3-1)\Gamma(s+c_4-1)}{\Gamma(s+c_1)\Gamma(s+c_2)}$$
$$\left(D(G_{t_0}) \frac{\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)} + 2r \sum_{j=t_0}^{s-2} \frac{\Gamma(j+1)^2}{\Gamma(j+c_3+1)\Gamma(j+c_4+1)} \right)$$
$$+ \frac{r\Gamma(s)^2}{\Gamma(s+c_1)\Gamma(s+c_2)} + r \sum_{j=s}^{t-1} \frac{\Gamma(j)\Gamma(j+1)}{\Gamma(j+c_1+1)\Gamma(j+c_2+1)} \Bigg],$$

where c_1–c_4 are as above.
Asymptotically as $t \to \infty$:
(i) *for $s = O(1)$*

$$\mathbb{E}[\deg_t(s)] = t^p(1+o(1))$$
$$\Bigg[(ps-p-r)\,\frac{\Gamma(s+c_3-1)\Gamma(s+c_4-1)}{\Gamma(s+c_1)\Gamma(s+c_2)}$$
$$\left(D(G_{t_0}) \frac{\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)} + 2r \sum_{j=t_0}^{s-2} \frac{\Gamma(j+1)^2}{\Gamma(j+c_3+1)\Gamma(j+c_4+1)} \right)$$
$$+ \frac{r\Gamma(s)^2}{\Gamma(s+c_1)\Gamma(s+c_2)} \left(1 + {}_3F_2\left[{s, s+1, 1 \atop s+c_1+1, s+c_2+1}; 1\right] \frac{s}{s^2+ps-r} \right) \Bigg].$$

(ii) *for $s = \omega(1)$ and $s = o(t)$*

$$\mathbb{E}[\deg_t(s)] = \begin{cases} D(G_{t_0}) \frac{p\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)} \left(\frac{t}{s}\right)^p s^{2p-1}(1+o(1)) & \text{if } p \le \frac{1}{2},\ r=0,\\ r \log\left(\frac{t}{s}\right)(1+o(1)) & \text{if } p=0,\ r>0,\\ \frac{r(1-p)}{p(1-2p)} \left(\frac{t}{s}\right)^p (1+o(1)) & \text{if } 0<p<\frac{1}{2},\ r>0,\\ r\sqrt{\frac{t}{s}} \log s\,(1+o(1)) & \text{if } p=\frac{1}{2},\ r>0,\\ \left(D(G_{t_0}) + \frac{2rt_0}{t_0^2+2pt_0-2r}\, {}_3F_2\left[{t_0+1, t_0+1, 1 \atop t_0+c_3+1, t_0+c_4+1}; 1\right] \right) \\ \quad \frac{p\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)} \left(\frac{t}{s}\right)^p s^{2p-1}(1+o(1)) & \text{if } p>\frac{1}{2}. \end{cases}$$

(iii) *for* $s = ct - o(t)$, $0 < c \leq 1$,

$$\mathbb{E}[\deg_t(s)] = \begin{cases} D(G_{t_0})\frac{p\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)}t^{2p-1}c^{p-1}(1+o(1)) & \text{if } p \leq \frac{1}{2},\ r = 0, \\ r\,(1-\log c)\,(1+o(1)) & \text{if } p = 0,\ r > 0, \\ \left(\frac{r(1-p)}{p(1-2p)c^p} - \frac{r}{p}\right)(1+o(1)) & \text{if } 0 < p < \frac{1}{2},\ r > 0, \\ \frac{r}{\sqrt{c}}\log t\,(1+o(1)) & \text{if } p = \frac{1}{2},\ r > 0, \\ \left(D(G_{t_0}) + \frac{2rt_0}{t_0^2+2pt_0-2r}\,{}_3F_2\left[\begin{smallmatrix} t_0+1, t_0+1, 1 \\ t_0+c_3+1, t_0+c_4+1 \end{smallmatrix}; 1\right]\right) & \\ \qquad \frac{p\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)}t^{2p-1}c^{p-1}(1+o(1)) & \text{if } p > \frac{1}{2}. \end{cases}$$

The theorem above shows that there is a threefold behavior with respect to the range of s: s small (constant), s medium (growing, but slower than t), and s large (when s is directly proportional to t). In the first case we observe a behavior very similar to the one for $1 \leq s \leq t_0$. In the second case we have a dependency on both s and t depending on the values of p and r. When the majority of the edges are created due to the copying (for $r = 0$ or $p > \frac{1}{2}$), then $\mathbb{E}[\deg_t(s)] = \Theta\left(\left(\frac{t}{s}\right)^p s^{2p-1}\right)$. When the majority of the edges are created due to the random addition (for $r > 0$ and $p < \frac{1}{2}$), then $\mathbb{E}[\deg_t(s)] = \Theta\left(\left(\frac{t}{s}\right)^p\right)$. Finally, we observe a phase transition for $p = \frac{1}{2}$, $r = 0$ with $\mathbb{E}[\deg_t(s)] = \Theta\left(\left(\frac{t}{s}\right)^p \log s\right)$. In the last case, the rates of growth of $\mathbb{E}[\deg_t(s)]$ are exactly like for $\mathbb{E}[\deg_t(t)]$: $\Theta(1)$, $\Theta(\log t)$ or $\Theta(t^{2p-1})$ respectively for different ranges of p and r.

Note that given the results presented in [19] and [22] we expect the real-world networks to fit the range $p > \frac{1}{2}$ and $r > 0$.

Finally, we derive the theorems showing the concentration of the quantity $\deg_t(s)$, given G_s. It is possible to show the following result:

▶ **Theorem 8.** *Asymptotically for $G_t \sim DD(t, p, r)$ and $s = O(1)$ it holds that*

$$\Pr[\deg_t(s) \geq A\,C\,t^p \log^2(t)] = O(t^{-A})$$

for some fixed constant $C > 0$ and any $A > 0$.

We also prove a respective lower bound:

▶ **Theorem 9.** *For $G_t \sim DD(t, p, r)$ with $p > 0$ and $s = O(1)$ it holds asymptotically that*

$$\Pr\left[\deg_t(s) \leq \frac{C}{A} t^p \log^{-3-\varepsilon}(t)\right] = O(t^{-A})$$

for some fixed constant $C > 0$ and any $A > 0$.

Note that in the $p = 0$ case, missing from Theorem 9 it is clear that we have with high probability at least a positive constant fraction of vertices with degree 0, as $\deg_s(t) \sim Bin\left(t, \frac{r}{t}\right)$.

Finally, we strongly believe that since $\deg_t(t)$ is closely dependent on the degree distribution in G_{t-1}, it is very unlikely that for s close to t the analogous bounds with only logarithmic factor from the mean for $\deg_t(s)$ exist.

3 Discussion

In this paper we have focused on a rigorous and precise analysis of the average degree of a given node over the evolution of the network as well as the average degree. We present exact and asymptotic results showing the behavior of important graph variables such as $D(G_t)$, $\deg_t(t)$ and $\deg_t(s)$.

It is worth noting that it is the parameter p that drives the rate of growth of expected value for these parameters. The value of the parameter r and the structure of the starting graph G_{t_0} impact only the leading constants and lower order terms.

We note that there are several phase transitions of these quantities as a function of p and r. However, as demonstrated in [19], it is seems that all real-world networks fall within a range $\frac{1}{2} < p < 1$, $r > 0$ – and this case should probably be the main topic of further investigation.

The proposed methodology can be easily extended to obtain variance and higher moments of the above quantities. Future work may include investigations into both the large deviation of the degree distribution as well as proving properties of the degree distribution (i.e., the number of nodes of degree k) as a function of both degree and time t. This, in turn, would allow us to differentiate between the ranges of parameters for which we obtain an asymmetric graph with high probability and the range where non-negligible symmetry occurs. Estimation of the graph entropy and the structural entropy would give us a way towards our ultimate aim: good quality (and efficient) algorithms which would match the entropy for this graph model.

References

1. Milton Abramowitz and Irene Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Dover Publications, 1972.
2. Micah Adler and Michael Mitzenmacher. Towards compressing web graphs. In *Proceedings DCC 2001. Data Compression Conference*, pages 203–212. IEEE, 2001.
3. David Aldous and Nathan Ross. Entropy of some models of sparse random graphs with vertex-names. *Probability in the Engineering and Informational Sciences*, 28(2):145–168, 2014.
4. Maciej Besta and Torsten Hoefler. Survey and taxonomy of lossless graph compression and space-efficient graph representations. *arXiv preprint arXiv:1806.01799*, 2018.
5. Yongwook Choi and Wojciech Szpankowski. Compression of graphical structures: Fundamental limits, algorithms, and experiments. *IEEE Transactions on Information Theory*, 58(2):620–638, 2012.
6. Reinhard Diestel. *Graph Theory.* Springer, 2005.
7. Felix Hermann and Peter Pfaffelhuber. Large-scale behavior of the partial duplication random graph. *ALEA*, 13:687–710, 2016.
8. Fereydoun Hormozdiari, Petra Berenbrink, Nataša Pržulj, and Süleyman Cenk Sahinalp. Not all scale-free networks are born equal: the role of the seed graph in PPI network evolution. *PLoS Computational Biology*, 3(7):e118, 2007.
9. Jonathan Jordan. The connected component of the partial duplication graph. *ALEA – Latin American Journal of Probability and Mathematical Statistics*, 15:1431–1445, 2018.
10. John Kieffer, En-Hui Yang, and Wojciech Szpankowski. Structural complexity of random binary trees. In *2009 IEEE International Symposium on Information Theory*, pages 635–639, 2009.
11. Jure Leskovec and Rok Sosič. Snap: A general-purpose network analysis and graph-mining library. *ACM Transactions on Intelligent Systems and Technology*, 8(1):1, 2016.
12. Si Li, Kwok Pui Choi, and Taoyang Wu. Degree distribution of large networks generated by the partial duplication model. *Theoretical Computer Science*, 476:94–108, 2013.
13. Tomasz Łuczak, Abram Magner, and Wojciech Szpankowski. Asymmetry and structural information in preferential attachment graphs. *Random Structures and Algorithms (arXiv preprint arXiv:1607.04102)*, pages 1–24, 2019.

14 Tomasz Łuczak, Abram Magner, and Wojciech Szpankowski. Compression of Preferential Attachment Graphs. In *2019 IEEE International Symposium on Information Theory*, 2019.

15 Abram Magner, Jithin Sreedharan, Ananth Grama, and Wojciech Szpankowski. Inferring temporal information from a snapshot of a dynamic network. *Nature Scientific Reports*, 9:3057–3062, 2019.

16 Abram Magner, Krzysztof Turowski, and Wojciech Szpankowski. Lossless compression of binary trees with correlated vertex names. *IEEE Transactions on Information Theory*, 64(9):6070–6080, 2018.

17 Claude Shannon. The lattice theory of information. *Transactions of the IRE Professional Group on Information Theory*, 1(1):105–107, 1953.

18 Ricard Solé, Romualdo Pastor-Satorras, Eric Smith, and Thomas Kepler. A model of large-scale proteome evolution. *Advances in Complex Systems*, 5(01):43–54, 2002.

19 Jithin Sreedharan, Krzysztof Turowski, and Wojciech Szpankowski. Revisiting Parameter Estimation in Biological Networks: Influence of Symmetries, 2019.

20 Wojciech Szpankowski. *Average case analysis of algorithms on sequences*. John Wiley & Sons, 2011.

21 Krzysztof Turowski, Abram Magner, and Wojciech Szpankowski. Compression of Dynamic Graphs Generated by a Duplication Model. In *56th Annual Allerton Conference on Communication, Control, and Computing*, pages 1089–1096, 2018.

22 Krzysztof Turowski, Jithin Sreedharan, and Wojciech Szpankowski. Temporal Ordered Clustering in Dynamic Networks, 2019.

23 Remco Van Der Hofstad. *Random graphs and complex networks*. Cambridge University Press, 2016.

A Useful lemmas

Here we derive a series of lemmas useful for the analysis of the following type of recurrence

$$\mathbb{E}[f(G_{n+1}) \mid G_n] = f(G_n)g_1(n) + g_2(n) \tag{5}$$

for some nonnegative functions $g_1(n)$, $g_2(n)$ and a Markov process G_n. It should be again noted that our recurrences for $\mathbb{E}[\deg_t(s)]$ and $\mathbb{E}[D(G_t)]$ (e.g., see (1) and (4)) fall under this pattern.

First lemma is a generalization of a result obtained in [7], where only the case $g_1(n) = 1+\frac{a}{n}$, $a > 0$, was analyzed.

▶ **Lemma 10.** *Let $(G_n)_{n=n_0}^{\infty}$ be a Markov process for which $\mathbb{E}f(G_{n_0}) > 0$ and (5) holds with $g_1(n) > 0$, $g_2(n) \geq 0$ for all $n = n_0, n_0+1, \ldots$. Then*

(ii) *The process $(M_n)_{n=n_0}^{\infty}$ defined by $M_{n_0} = f(G_{n_0})$ and*

$$M_n = f(G_n) \prod_{k=n_0}^{n-1} \frac{1}{g_1(k)} - \sum_{j=n_0}^{n-1} g_2(j) \prod_{k=n_0}^{j} \frac{1}{g_1(k)}$$

is a martingale.

(ii) *For all $n \geq n_0$*

$$\mathbb{E}f(G_n) = f(G_{n_0}) \prod_{k=n_0}^{n-1} g_1(k) + \sum_{j=n_0}^{n-1} g_2(j) \prod_{k=j+1}^{n-1} g_1(k)$$
$$= \prod_{k=n_0}^{n-1} g_1(k) \left(f(G_{n_0}) + \sum_{j=n_0}^{n-1} g_2(j) \prod_{k=n_0}^{j} \frac{1}{g_1(k)} \right).$$

Proof. Observe that

$$\mathbb{E}[M_{n+1} \mid G_n] = \mathbb{E}[f(G_{n+1}) \mid G_n] \prod_{k=n_0}^{n} \frac{1}{g_1(k)} - \sum_{j=n_0}^{n} g_2(j) \prod_{k=n_0}^{j} \frac{1}{g_1(k)}$$
$$= f(G_n) \prod_{k=n_0}^{n-1} \frac{1}{g_1(k)} - \sum_{j=n_0}^{n-1} g_2(j) \prod_{k=n_0}^{j} \frac{1}{g_1(k)} = M_n$$

which proves (i). Furthermore, after some algebra and taking expectation with respect to G_n we arrive at

$$\mathbb{E}f(G_n) = \mathbb{E}[M_n] \prod_{k=n_0}^{n-1} g_1(k) + \sum_{j=n_0}^{n-1} g_2(j) \prod_{k=n_0}^{j} \frac{1}{g_1(k)} \prod_{k=n_0}^{n-1} g_1(k)$$
$$= f(G_{n_0}) \prod_{k=n_0}^{n-1} g_1(k) + \sum_{j=n_0}^{n-1} g_2(j) \prod_{k=j+1}^{n-1} g_1(k)$$

which completes the proof. ◀

We now observe that any solution of recurrences of type (5) contains sophisticated products and sum of products (e.g., see Eqn. (3)) with which we must deal to find asymptotics. The next lemma shows how to handle such products.

► **Lemma 11.** *Let $W_1(k)$, $W_2(k)$ be polynomials of degree d with respective roots a_i, b_i ($i = 1, \dots, d$), that is, $W_1(k) = \prod_{i=1}^{d}(k - a_i)$ and $W_2(k) = \prod_{j=1}^{d}(k - b_j)$. Then*

$$\prod_{k=n_0}^{n-1} \frac{W_1(k)}{W_2(k)} = \prod_{i=1}^{d} \frac{\Gamma(n-a_i)}{\Gamma(n-b_i)} \frac{\Gamma(n_0-b_i)}{\Gamma(n_0-a_i)}.$$

Proof. We have

$$\prod_{k=n_0}^{n-1} \frac{W_1(k)}{W_2(k)} = \prod_{k=n_0}^{n-1}\prod_{i=1}^{d} \frac{k-a_i}{k-b_i} = \prod_{i=1}^{d}\prod_{k=n_0}^{n-1} \frac{k-a_i}{k-b_i} = \prod_{i=1}^{d} \frac{\Gamma(n-a_i)}{\Gamma(n-b_i)} \frac{\Gamma(n_0-b_i)}{\Gamma(n_0-a_i)}$$

which completes the proof. ◄

The next lemma presents well-known asymptotic expansion of the gamma function but we include it here for the sake of completeness.

► **Lemma 12** (Abramowitz, Stegun [1])**.** *For any $a, b \in \mathbb{R}$ if $n \to \infty$, then*

$$\begin{aligned}\frac{\Gamma(n+a)}{\Gamma(n+b)} &= n^{a-b} \sum_{k=0}^{\infty} \binom{a-b}{k} B_k^{(a-b+1)}(a) \cdot n^{-k} \\ &= n^{a-b}\left(1 + \frac{(a-b)(a+b-1)}{2n} + O\left(\frac{1}{n^2}\right)\right),\end{aligned}$$

where $B_k^{(l)}(x)$ are the generalized Bernoulli polynomials.

Now we deal with sum of products as seen in (5). In particular, we are interested in the following sum of products

$$\sum_{j=n_0}^{n} \frac{\prod_{i=1}^{k}\Gamma(j+a_i)}{\prod_{i=1}^{k}\Gamma(j+b_i)}$$

with $a = \sum_{i=1}^{k} a_i$, $b = \sum_{i=1}^{k} b_i$. In the next three lemmas we consider three cases: $a+1 > b$, $a+1 = b$ and $a+1 < b$.

► **Lemma 13.** *Let $a_i, b_i \in \mathbb{R}$ ($k \in \mathbb{N}$) with $a = \sum_{i=1}^{k} a_i$, $b = \sum_{i=1}^{k} b_i$ such that $a+1 > b$. Then it holds asymptotically for $n \to \infty$ that*

$$\sum_{j=n_0}^{n} \frac{\prod_{i=1}^{k}\Gamma(j+a_i)}{\prod_{i=1}^{k}\Gamma(j+b_i)} = \frac{n^{a-b+1}}{a-b+1} + O\left(n^{\max\{a-b,0\}}\right)$$

Proof. We estimate the sum using Lemma 12 and the Euler-Maclaurin formula [20, p. 294]

$$\begin{aligned}\sum_{j=n_0}^{n} \frac{\prod_{i=1}^{k}\Gamma(j+a_i)}{\prod_{i=1}^{k}\Gamma(j+b_i)} &= \sum_{j=n_0}^{n} j^{a-b}\left(1 + O\left(\frac{1}{j}\right)\right) = \int_{n_0}^{n} j^{a-b}\left(1 + O\left(\frac{1}{j}\right)\right) \mathrm{d}j \\ &= \left[j^{a-b+1}\left(\frac{1}{a-b+1} + O\left(\frac{1}{j}\right)\right)\right]_{n_0}^{n} = n^{a-b+1}\left(\frac{1}{a-b+1} + O\left(\frac{1}{n}\right)\right) + O(1)\end{aligned}$$

which completes the proof. ◄

► **Lemma 14.** *Let $a_i, b_i \in \mathbb{R}$ ($k \in \mathbb{N}$) with $a = \sum_{i=1}^{k} a_i$, $b = \sum_{i=1}^{k} b_i$ such that $a+1 = b$. Then asymptotically*

$$\sum_{j=n_0}^{n} \frac{\prod_{i=1}^{k}\Gamma(j+a_i)}{\prod_{i=1}^{k}\Gamma(j+b_i)} = \ln n + O(1)$$

Proof. We proceed as before

$$\sum_{j=n_0}^{n} \frac{\prod_{i=1}^k \Gamma(j+a_i)}{\prod_{i=1}^k \Gamma(j+b_i)} = \sum_{j=n_0}^{n} \frac{1}{j}\left(1+O\left(\frac{1}{j}\right)\right) = \int_{n_0}^{n} \frac{1}{j}\left(1+O\left(\frac{1}{j}\right)\right) \mathrm{d}j = \ln n + O(1)$$

which completes the proof. ◀

▶ **Lemma 15.** *Let $a_i, b_i \in \mathbb{R}$ ($i = 1, \ldots, k$, $k \in \mathbb{N}$) with $a = \sum_{i=1}^k a_i$, $b = \sum_{i=1}^k b_i$ such that $a + 1 < b$. Then it holds for every $n \in \mathbb{N}_+$ that*

$$\sum_{j=n}^{\infty} \frac{\prod_{i=1}^k \Gamma(j+a_i)}{\prod_{i=1}^k \Gamma(j+b_i)} = \frac{\prod_{i=1}^k \Gamma(n+a_i)}{\prod_{i=1}^k \Gamma(n+b_i)} \, {}_{k+1}F_k\left[{n+a_1,\ldots,n+a_k,1 \atop n+b_1,\ldots,n+b_k}; 1\right]$$

where ${}_pF_q[{\mathbf{a} \atop \mathbf{b}}; z]$ is the generalized hypergeometric function. Moreover it is true that asymptotically

$$\sum_{j=n}^{\infty} \frac{\prod_{i=1}^k \Gamma(j+a_i)}{\prod_{i=1}^k \Gamma(j+b_i)} = n^{a-b+1}\left(\frac{1}{b-a-1} + O\left(\frac{1}{n}\right)\right).$$

Proof. The proof of the first formula follows directly from the definition of the generalized hypergeometric function. Second formula follows from Lemma 12, as we know that for $n \to \infty$:

$$\sum_{j=n}^{\infty} \frac{\prod_{i=1}^k \Gamma(j+a_i)}{\prod_{i=1}^k \Gamma(j+b_i)} = \sum_{j=n}^{\infty} j^{a-b}\left(1+O\left(\frac{1}{j}\right)\right) = \int_{n}^{\infty} j^{a-b}\left(1+O\left(\frac{1}{j}\right)\right) \mathrm{d}j$$
$$= \left[j^{a-b+1}\left(\frac{1}{b-a-1} + O\left(\frac{1}{j}\right)\right)\right]_n^{\infty} = n^{a-b+1}\left(\frac{1}{b-a-1} + O\left(\frac{1}{n}\right)\right)$$

as desired. ◀

B Proof of Lemma 1

Now we turn our attention to the proof of Lemma 1. We first observe that it follows from the definition of the model that the degree of the new vertex $t+1$ is the total number of edges from $t+1$ to $N_t(parent(t+1))$ (chosen independently with probability p) and to all other vertices (chosen independently with probability $\frac{r}{t}$). Note that it can be expressed as a sum of two independent binomial variables

$$\deg_{t+1}(t+1) \sim \mathrm{Bin}\left(\deg_t(parent(t+1)), p\right) + \mathrm{Bin}\left(t - \deg_t(parent(t+1)), \frac{r}{t}\right).$$

Hence

$$\mathbb{E}[\deg_{t+1}(t+1) \mid G_t] = \sum_{k=0}^{t} \Pr(\deg_t(parent(t+1)) = k) \sum_{a=0}^{k} \binom{k}{a} p^a (1-p)^{k-a}$$
$$\sum_{b=0}^{t-k} \binom{t-k}{b} \left(\frac{r}{t}\right)^b \left(1-\frac{r}{t}\right)^{t-k-b} (a+b)$$
$$= \sum_{k=0}^{t} \Pr(\deg_t(parent(t+1)) = k) \left(pk + \frac{r}{t}(t-k)\right)$$

$$= \left(p - \frac{r}{t}\right) \sum_{k=0}^{t} k \Pr(\deg_t(parent(t+1)) = k) + r.$$

Since parent sampling is uniform, we know that $\Pr(parent(t+1) = i) = \frac{1}{t}$ and therefore

$$D(G_t) = \sum_{i=1}^{t} \Pr(parent(t+1) = i) \deg_t(i) = \sum_{k=0}^{t} k \Pr(\deg_t(parent(t+1)) = k).$$

Combining the last two equations above with the law of total expectation we finally establish Lemma 1.

C Proofs of Theorem 2 and Theorems 6–7

We start with the proof of Theorem 2. First, we observe that by combining Eqn. (4) with Lemmas 10 and 11 we prove the first part of Theorem 1. In similar fashion, the second part of Theorem 2 follows directly from the first part, combined with Lemmas 13, 14 and 15 for the respective ranges of p.

Finally, we proceed to the proof of Theorems 6 and 7. First, we apply Lemma 10 with $g_1(t) = 1 + \frac{p}{t} - \frac{r}{t^2}$ and $g_2(t) = \frac{r}{t}$ to Eqn. (1) and we obtain aforementioned Eqn. (3). Now we combine this result with Lemma 11. First, we if we apply it for $1 \le s \le t_0$ we obtain directly the exact formula in Theorem 6.

Similarly, for Theorem 7, we get the almost identical formula. The only difference is that we do not stop the recurrence at G_{t_0}, but at G_s:

$$\mathbb{E}[\deg_t(s)] = \frac{\Gamma(t+c_1)\Gamma(t+c_2)}{\Gamma(t)^2}$$
$$\left(\mathbb{E}[\deg_s(s)] \frac{\Gamma(s)^2}{\Gamma(s+c_1)\Gamma(s+c_2)} + \sum_{j=s}^{t-1} \frac{r\Gamma(j)\Gamma(j+1)}{\Gamma(j+c_1+1)\Gamma(j+c_2+1)}\right)$$

where $c_1 = \frac{p+\sqrt{p^2+4r}}{2}$, $c_2 = \frac{p-\sqrt{p^2+4r}}{2}$.

Now it is sufficient to apply Corollary 5 to this equation to get the exact formula for $\mathbb{E}[\deg_t(s)]$.

The asymptotic formulas in Theorems 6 and 7 – as it was in the case of $\mathbb{E}[D(G_t)]$ above – are derived as straightforward consequences of Lemmas 13, 14 and 15.

D Proof of Theorem 3

In order to prove the theorem we proceed as following: first we provide an asymptotic bound on $\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1)) | G_t\right]$, then we apply it for a suitable choices of λ, which allow us to use Chernoff bound.

▶ **Lemma 16.** *For any $\lambda = O(\frac{1}{t})$ it holds that*

$$\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1)) | G_t\right] \le \exp\left(\lambda p D(G_t)(1 + O(\lambda t)) + \lambda r(1 + O(\lambda))\right).$$

Proof.

$$\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1)) | G_t\right]$$
$$= \frac{1}{t} \sum_{i=1}^{t} \mathbb{E}\left[\exp\left(\lambda Bin(\deg_t(i), p) + \lambda Bin\left(t - \deg_t(i), \frac{r}{t}\right)\right) | G_t\right]$$

$$\leq \frac{1}{t}\sum_{i=1}^{t}\left(1-p+pe^{\lambda}\right)^{\deg_t(i)}\left(1-\frac{r}{t}+\frac{r}{t}e^{\lambda}\right)^{t-\deg_t(i)}.$$

Since $e^x \leq 1+x+x^2$ for all $x \in [0,1]$, $(1+x)^y \leq 1+xy+(xy)^2$ for $0 \leq xy \leq 1$ and $1+x \leq e^x$ for any x:

$$\begin{aligned}
&\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1))|G_t\right] \\
&\quad \leq \frac{1}{t}\sum_{i=1}^{t}(1+p\lambda(1+O(\lambda))^{\deg_t(i)}\left(1+\frac{r\lambda}{t}(1+O(\lambda))\right)^{t-\deg_t(i)} \\
&\quad \leq \frac{1}{t}\sum_{i=1}^{t}(1+p\lambda\deg_t(i)(1+O(\lambda t))\left(1+r\lambda(1+O(\lambda))\right) \\
&\quad \leq \frac{1}{t}\sum_{i=1}^{t}(1+p\lambda\deg_t(i)(1+O(\lambda t)))\exp\left(r\lambda(1+O(\lambda))\right) \\
&\quad = (1+p\lambda D(G_t)(1+O(\lambda t)))\exp\left(r\lambda(1+O(\lambda))\right) \\
&\quad \leq \exp\left(\lambda p D(G_t)(1+O(\lambda t))+\lambda r(1+O(\lambda))\right).
\end{aligned}$$

◀

Now we are ready to finally prove the theorem.

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda_{t+1}D(G_{t+1})\right) \mid G_t\right] &= \mathbb{E}\left[\exp\left(\lambda_{t+1}\left(\frac{t}{t+1}D(G_t)+\frac{2}{t+1}\deg_{t+1}(t+1)\right)\right) \mid G_t\right] \\
&= \exp\left(\frac{\lambda_{t+1}t}{t+1}D(G_t)\right)\mathbb{E}\left[\exp\left(\frac{2\lambda_{t+1}}{t+1}\deg_{t+1}(t+1)\right) \mid G_t\right]
\end{aligned}$$

Now we may use Lemma 17 with $\lambda = \frac{2\lambda_{t+1}}{t+1}$ to get

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda_{t+1}D(G_{t+1})\right) \mid G_t\right] = \\
&\quad \leq \exp\left(\lambda_{t+1}D(G_t)\left(1-\frac{2p-1}{t+1}\right)(1+O(\lambda_{t+1}))+\frac{2r\lambda_{t+1}}{t+1}(1+o(t^{-1}))\right).
\end{aligned}$$

Let us define for $k = t_0, \ldots, t-1$

$$\lambda_k = \lambda_{k+1}\left(1+\left(\frac{2p-1}{t+1}\right)(1+O(\lambda_{k+1}))\right)$$

and let $\varepsilon_t \geq \lambda_k$ for all k.

Then clearly

$$\begin{aligned}
\lambda_{t_0} &\in \left[\lambda_t\prod_{k=t_0}^{t-1}\left(1+\frac{2p-1}{k+1}\right), \lambda_t\prod_{k=t_0}^{t-1}\left(1+\left(\frac{2p-1}{k+1}\right)(1+O(\varepsilon_t))\right)\right] \\
&\subseteq \left[\lambda_t\left(\frac{t}{t_0}\right)^{2p-1}(1+o(1)), \lambda_t\left(\frac{t}{t_0}\right)^{(2p-1)(1+O(\varepsilon_t))}(1+o(1))\right]
\end{aligned}$$

It follows that

$$\mathbb{E}\left[\exp\left(\lambda_t D(G_t)\right)\right] \leq \exp\left(\lambda_{t_0}D(G_{t_0})\right)\prod_{k=t_0}^{t-1}\exp\left(\frac{2r\lambda_{k+1}}{k+1}\left(1+o(k^{-1})\right)\right)$$

$$\leq \exp\left(\lambda_{t_0} D(G_{t_0})\right) \exp\left(2r\varepsilon_{t+1} \ln \frac{t}{t_0} + C_1\right) = \exp\left(\lambda_{t_0} D(G_{t_0})\right) \left(\frac{t}{t_0}\right)^{2r\varepsilon_{t+1} + C_1}$$

for a certain constant C_1.

Finally, let $\lambda_t = \varepsilon_t \left(\frac{t}{t_0}\right)^{-(2p-1)(1+O(\varepsilon_t))}$ so that $\lambda_{t_0} \leq \varepsilon_t$. Then from Chernoff bound it follows that

$$\begin{aligned}
\Pr[D(G_t) \geq \alpha \mathbb{E} D(G_t)] &= \Pr[\exp(D(G_t) - \alpha \mathbb{E} D(G_t)) \geq 1] \\
&\leq \exp\left(-\alpha \lambda_t \mathbb{E} D(G_t)\right) \mathbb{E}[\exp\left(\lambda_t D(G_t)\right)] \\
&\leq \exp\left(-\alpha \lambda_t \mathbb{E} D(G_t)\right) \exp\left(\lambda_{t_0} D(G_{t_0})\right) \left(\frac{t}{t_0}\right)^{2r\varepsilon_{t+1} + C_1}
\end{aligned}$$

Assume $\varepsilon_t = \frac{1}{\ln(t/t_0)}$. For $p > \frac{1}{2}$ we have $\mathbb{E} D(G_t) = C_2 \left(\frac{t}{t_0}\right)^{2p-1} (1 + o(1))$, and therefore

$$\begin{aligned}
&\Pr\left[D(G_t) \geq \alpha C_2 \left(\frac{t}{t_0}\right)^{2p-1} (1 + o(1))\right] \\
&\quad \leq \exp\left(-\alpha C_2 \varepsilon_t \left(\frac{t}{t_0}\right)^{-(2p-1)\varepsilon_t}\right) \exp\left(\varepsilon_t (t_0 - 1)\right)\right) \left(\frac{t}{t_0}\right)^{2r\varepsilon_{t+1} + C_1} \\
&\quad \leq \exp\left(-\alpha C_2 \frac{\exp\left(-2p+1\right)}{\ln(t/t_0)}\right) \exp\left(\frac{t_0 - 1}{\ln(t/t_0)}\right) \exp\left(2r + C_1\right)
\end{aligned}$$

The last two elements are bounded by a constant, so it is sufficient to pick $\alpha = \frac{A}{C_2} \exp(2p - 1) \ln^2(t)$ to complete the proof for the case $p > \frac{1}{2}$.

Now, for $p < \frac{1}{2}$ and $p = \frac{1}{2}$ it is sufficient to use $\mathbb{E} D(G_t) = C_2(1 + o(1))$ and $\mathbb{E} D(G_t) = C_2 \ln t (1 + o(1))$, respectively.

E Proof of Theorem 4

We start the proof by obtaining a simple lemma, analogous to Lemma 16:

▶ **Lemma 17.** *For any $\lambda = O(\frac{1}{t})$ it holds that*

$$\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1)) | G_t\right] \leq \exp\left(2\lambda p D(G_t)(1 + O(\lambda)) + 2\lambda r(1 + O(\lambda))\right).$$

Proof.

$$\begin{aligned}
&\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1)) | G_t\right] \\
&\quad = \frac{1}{t} \sum_{i=1}^{t} \mathbb{E}\left[\exp\left(\lambda Bin(\deg_t(i), p) + \lambda Bin\left(t - \deg_t(i), \frac{r}{t}\right)\right) | G_t\right] \\
&\quad \leq \frac{1}{t} \sum_{i=1}^{t} \left(1 - p + pe^{\lambda}\right)^{\deg_t(i)} \left(1 - \frac{r}{t} + \frac{r}{t} e^{\lambda}\right)^{t - \deg_t(i)}.
\end{aligned}$$

Since $e^x \leq 1 + x + x^2$ for all $x \in [0, 1]$, $(1+x)^y \leq 1 + 2xy$ for $0 \leq xy \leq 1$, and $1 + x \leq e^x$ for all x

$$\begin{aligned}
&\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1)) | G_t\right] \\
&\quad \leq \frac{1}{t} \sum_{i=1}^{t} \left(1 + p\lambda(1 + O(\lambda))\right)^{\deg_t(i)} \left(1 + \frac{r\lambda}{t}(1 + O(\lambda))\right)^{t - \deg_t(i)}
\end{aligned}$$

$$\leq \frac{1}{t}\sum_{i=1}^{t}(1+2p\lambda \deg_t(i)(1+O(\lambda)))\,(1+2r\lambda(1+O(\lambda))))$$
$$\leq \frac{1}{t}\sum_{i=1}^{t}(1+2p\lambda \deg_t(i)(1+O(\lambda)))\exp\left(2r(1+O(\lambda))\right)$$
$$= (1+2p\lambda D(G_t)(1+O(\lambda)))\exp\left(2r(1+O(\lambda))\right))$$
$$\leq \exp\left(2\lambda p D(G_t)(1+O(\lambda))+2\lambda r(1+O(\lambda))\right).$$

◀

Next, using the lemma above and Theorem 3 we limit the growth of $D(G_t)$ over certain intervals:

▶ **Lemma 18.** *Let $p > \frac{1}{2}$. For sufficiently large t and all $k < t$ it is true that*

$$\Pr[D(G_{(k+1)t}) - D(G_{kt}) \geq AC((k+1)^{2p-1} - k^{2p-1})t^{2p-1}\log^2(t)] = O(t^{-A})$$

for some fixed constant $C > 0$ and any $A > 1$.

Proof. First, let us define events $\mathcal{B}_i = [D(G_{i+1}) \geq (A+1)\,C_1\,i^{2p-1}\log^2(i)]$ with a constant C_1 such that by Theorem 3 it is true that $\Pr[\mathcal{B}_i] = O(i^{-A-1})$. Let us also denote $\mathcal{A}_k = \bigcup_{i=kt}^{(k+1)t-1}\mathcal{B}_i$ and observe that $\Pr\left[\mathcal{A}_k\right] = O(t^{-A})$.

Now, we note that from Lemma 16 for any $\lambda = o(1)$

$$\mathbb{E}\left[\exp\left(\lambda(D(G_{t+1}) - D(G_t))\right)\middle|G_t, \neg\mathcal{B}_t\right]$$
$$\leq \mathbb{E}\left[\exp\left(\frac{2\lambda}{t+1}\deg_{t+1}(t+1)\right)\middle|G_t, \neg\mathcal{B}_t\right]$$
$$\leq \left[\exp\left(\frac{2\lambda p}{t+1}D(G_t)(1+O(\lambda)) + \frac{2\lambda r}{t+1}(1+O(\lambda))\right)\middle|\neg\mathcal{B}_t\right]$$
$$\leq \exp\left(\lambda\,(A+1)\,C_2\,t^{2p-2}\log^2(t)(1+o(1))\right)$$

for a certain constant C_2.

Now we proceed as following:

$$\Pr[D(G_{(k+1)t}) - D(G_{kt}) \geq d|G_{kt}]$$
$$\leq \Pr[D(G_{(k+1)t}) - D(G_{kt}) \geq d|G_{kt}, \neg\mathcal{A}_k]\Pr[\neg\mathcal{A}] + \Pr[\mathcal{A}_k]$$
$$\leq \exp(-\lambda d)\mathbb{E}\left[\exp\left(\lambda(D(G_{(k+1)t}) - D(G_{kt}))\right)|G_{kt}, \neg\mathcal{A}_k\right] + O(t^{-A})$$
$$\leq \exp(-\lambda d)\prod_{i=kt}^{(k+1)t-1}\mathbb{E}\left[\exp\left(\lambda(D(G_{i+1}) - D(G_i))\right)\middle|G_i, \neg\mathcal{B}_i\right] + O(t^{-A})$$
$$\leq \exp(-\lambda d)\prod_{i=kt}^{(k+1)t-1}\exp\left(\lambda\,(A+1)\,C_2\,i^{2p-2}\log^2(i)(1+o(1))\right) + O(t^{-A})$$
$$\leq \exp(-\lambda d)\exp\left(\sum_{i=kt}^{(k+1)t-1}\lambda\,(A+1)\,C_3\,i^{2p-2}\log^2(t)(1+o(1))\right) + O(t^{-A})$$
$$\leq \exp(-\lambda d)\exp\left(\lambda\,(A+1)\,C_3((k+1)^{2p-1} - k^{2p-1})t^{2p-1}\log^2(t)\right) + O(t^{-A})$$

for a certain constant C_3.

Finally, it is sufficient to take $\lambda = \left(((k+1)^{2p-1} - k^{2p-1})\log^2(t)\right)^{-1}$ and $d = AC_4((k+1)^{2p-1} - k^{2p-1})t^{2p-1}\log^2(t)$ for sufficiently large C_4 to obtain the final result. ◀

Now we may return to the main theorem. Let $Y_k = D(G_{(k+1)t}) - D(G_{kt})$. We know that for $p > \frac{1}{2}$

$$\mathbb{E}Y_k = \mathbb{E}D(G_{(k+1)t}) - \mathbb{E}D(G_{kt}) = C_1\left((k+1)^{2p-1} - k^{2p-1}\right)t^{2p-1}(1+o(1))$$

for some constant C_1.

Let now define the following events:

$$\mathcal{A}_1 = \left[Y_k \le \frac{t^{2p-1}}{f(t)}\right]$$
$$\mathcal{A}_2 = \left[\frac{t^{2p-1}}{f(t)} < Y_k \le C_2((k+1)^{2p-1} - k^{2p-1})t^{2p-1}\log^2(t)\right]$$
$$\mathcal{A}_3 = \left[Y_k > C_2((k+1)^{2p-1} - k^{2p-1})t^{2p-1}\log^2(t)\right]$$

for a constant C_2 such that (from the lemma above) $\Pr[\mathcal{A}_3] = O(t^{-2})$. Here $f(t)$ is any (monotonic) function such that $f(t) \to \infty$ as $t \to \infty$.

We know that

$$\mathbb{E}Y_k = \mathbb{E}\left[Y_k|\mathcal{A}_1\right]\Pr\left[\mathcal{A}_1\right] + \mathbb{E}\left[Y_k|\mathcal{A}_2\right]\Pr\left[\mathcal{A}_2\right] + \mathbb{E}\left[Y_k|\mathcal{A}_3\right]\Pr\left[\mathcal{A}_3\right]$$
$$\mathbb{E}Y_k \ge C_1\left((k+1)^{2p-1} - k^{2p-1}\right)t^{2p-1}$$
$$\mathbb{E}\left[Y_k|\mathcal{A}_1\right] \le \frac{t^{2p-1}}{f(t)}$$
$$\mathbb{E}\left[Y_k|\mathcal{A}_2\right] \le C_2((k+1)^{2p-1} - k^{2p-1})t^{2p-1}\log^2(t)$$
$$\mathbb{E}\left[Y_k|\mathcal{A}_3\right] \le (k+1)t$$

and therefore for sufficiently large t it holds that

$$\Pr[\mathcal{A}_1] \le \frac{C_2\left((k+1)^{2p-1} - k^{2p-1}\right)\log^2(t) - C_1\left((k+1)^{2p-1} - k^{2p-1}\right)}{C_2\left((k+1)^{2p-1} - k^{2p-1}\right)\log^2(t) - \frac{1}{f(t)}}$$
$$\le 1 - \frac{C_1}{2C_2\log^2(t)}.$$

Let now $\tau = kt$.

$$\Pr\left[D(G_\tau) \le t^{2p-1}f^{-1}(t)\right] = \Pr\left[\bigcap_{i=1}^{k} Y_i \le \frac{t^{2p-1}}{f(t)}\right]$$
$$\le \prod_{i=1}^{k}\Pr\left[Y_i \le \frac{t^{2p-1}}{f(t)}\right] \le \prod_{i=1}^{k}\left(1 - \frac{C_1}{2C_2\log^2(t)}\right)$$

Therefore, if we assume $k = \frac{2AC_2}{C_1}\log^3(t)$, we get

$$\Pr\left[D(G_\tau) \le \frac{t^{2p-1}}{f(t)}\right] = \exp\left(-A\log(t)\right) = O(t^{-A})$$

and finally

$$\Pr\left[D(G_t) \le \frac{C_3}{A^{2p-1}}t^{2p-1}\log^{-3(2p-1)-\varepsilon}(t)\right] = O(t^{-A}).$$

for some constant C_3 and any $\varepsilon > 0$.

F Proof of Theorem 8

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda_{t+1}\deg_{t+1}(s)\right)\mid G_t\right]= \\
&\quad=\left(\frac{\deg_t(s)}{t}p+\frac{t-\deg_t(s)}{t}\frac{r}{t}\right)\exp\left(\lambda_{t+1}\left(\deg_t(s)+1\right)\right) \\
&\quad\quad+\left(\frac{\deg_t(s)}{t}(1-p)+\frac{t-\deg_t(s)}{t}\left(1-\frac{r}{t}\right)\right)\exp\left(\lambda_{t+1}\deg_t(s)\right) \\
&\quad=\exp\left(\lambda_{t+1}\deg_t(s)\right) \\
&\quad\quad\left(\frac{\deg_t(s)}{t}\left(1-p+p\exp\left(\lambda_{t+1}\right)\right)+\frac{t-\deg_t(s)}{t}\left(1-\frac{r}{t}+\frac{r}{t}\exp\left(\lambda_{t+1}\right)\right)\right) \\
&\quad\le\exp\left(\lambda_{t+1}\deg_t(s)\right)\left(1+\left(\frac{p\deg_t(s)}{t}+\frac{r\left(t-\deg_t(s)\right)}{t^2}\right)\left(\lambda_{t+1}+\lambda_{t+1}^2\right)\right) \\
&\quad\le\exp\left(\lambda_{t+1}\deg_t(s)+\left(\frac{p\deg_t(s)}{t}+\frac{r\left(t-\deg_t(s)\right)}{t^2}\right)\left(\lambda_{t+1}+\lambda_{t+1}^2\right)\right) \\
&\quad=\exp\left(\lambda_{t+1}\deg_t(s)\left(1+\left(\frac{p}{t}-\frac{r}{t^2}\right)\left(1+\lambda_{t+1}\right)\right)\right)\exp\left(\lambda_{t+1}\left(1+\lambda_{t+1}\right)\frac{r}{t}\right).
\end{aligned}$$

Let us assume that $\lambda_k\le\varepsilon_t=o(1)$ for all $s\le k\le t$. Then for all $k=s,s+1,\dots,t$ we have

$$\lambda_k=\lambda_{k+1}\left(1+\left(\frac{p}{k}-\frac{r}{k^2}\right)\left(1+\lambda_{k+1}\right)\right)\le\lambda_{k+1}\left(1+\left(\frac{p}{k}-\frac{r}{k^2}\right)\left(1+\varepsilon_t\right)\right)$$

which lead us to

$$\begin{aligned}
\lambda_s&\le\lambda_t\prod_{k=s}^{t-1}\left(1+\left(\frac{p}{k}-\frac{r}{k^2}\right)\left(1+\varepsilon_t\right)\right)\le\lambda_t\exp\left(\left(1+\varepsilon_t\right)\sum_{k=s}^{t-1}\left(\frac{p}{k}-\frac{r}{k^2}\right)\right) \\
&\le\lambda_t\exp\left(\left(1+\varepsilon_t\right)\int_s^t\left(\frac{p}{k}-\frac{r}{k^2}\,\mathrm{d}k\right)\right)=\lambda_t\exp\left(\left(1+\varepsilon_t\right)\left(p\ln\frac{t}{s}+r\left(\frac{1}{t}-\frac{1}{s}\right)\right)\right) \\
&\le\lambda_t\left(\frac{t}{s}\right)^{p(1+\varepsilon_t)}\exp\left(\frac{r}{t}\left(1+\varepsilon_t\right)\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda_t\deg_t(s)\right)|G_s\right]\le\exp\left(\lambda_s\deg_s(s)\right))\prod_{k=s}^{t-1}\exp\left(\lambda_{k+1}\left(1+\lambda_{k+1}\right)\frac{r}{k}\right) \\
&\quad\le\exp\left(\lambda_s\deg_s(s)\right))\exp\left(\varepsilon_t\left(1+\varepsilon_t\right)r\ln\frac{t}{s}\right)\le\exp\left(\lambda_s\deg_s(s)\right))\left(\frac{t}{s}\right)^{r\varepsilon_t(1+\varepsilon_t)}
\end{aligned}$$

Now, let $\lambda_t=\epsilon_t\left(\frac{t}{s}\right)^{-p(1+\varepsilon_t)}\exp\left(-\frac{r}{t}\left(1+\varepsilon_t\right)\right)$ so that $\lambda_s\le\epsilon_t$. Then, from Chernoff bound it follows that

$$\begin{aligned}
\Pr[\deg_t(s)\ge\alpha\mathbb{E}\deg_t(s)|G_s]&=\Pr[\exp(\deg_t(s)-\alpha\mathbb{E}\deg_t(s))\ge 1|G_s] \\
&\le\exp\left(-\alpha\lambda_t\mathbb{E}[\deg_t(s)|G_s]\right)\mathbb{E}[\exp\left(\lambda_t\deg_t(s)\right)|G_s] \\
&\le\exp\left(-\alpha\lambda_t\mathbb{E}[\deg_t(s)|G_s]\right)\exp\left(\lambda_s\deg_s(s)\right)\left(\frac{t}{s}\right)^{r\varepsilon_t(1+\varepsilon_t)}.
\end{aligned}$$

Let's assume $\varepsilon_t=\frac{1}{\ln t}$. Recall now from Theorems 6 and 7 that if $s=O(1)$, then it holds that $\mathbb{E}[\deg_t(s)|G_s]=C_1t^p$ and therefore

$$\Pr[\deg_t(s)\ge\alpha C_1t^p|G_s]\le\exp\left(-\alpha C_2\epsilon_t t^{-p\varepsilon_t}\right)\exp\left(\epsilon_t\deg_s(s)\right))\left(\frac{t}{s}\right)^{r\varepsilon_t(1+\varepsilon_t)}$$

$$\leq \exp\left(-\frac{\alpha C_3}{\ln t}\right) \exp\left(\frac{\deg_s(s)}{\ln t}\right) \exp(2r)$$

for certain constants C_2, C_3.

Therefore, it is sufficient to set $\alpha = \frac{A}{C_3} \ln^2 t$ to get the final result.

G Proof of Theorem 9

We proceed similarly as in the proof of Theorem 4:

▶ **Lemma 19.** *Let $p > 0$ and $s = O(1)$. For sufficiently large t and all $k < t$ it is true that*

$$\Pr[\deg_{(k+1)t}(s) - \deg_{kt}(s) \geq AC((k+1)^p - k^p)t^p \log^2(t)] = O(t^{-A})$$

for some fixed constant $C > 0$ and any $A > 1$.

Proof. Let us define events $\mathcal{B}_i = [\deg_{i+1}(s) \geq (A+1)\, C_1\, i^p \log^2(i)]$ with a constant C_1 such that by Theorem 8 it is true that $\Pr[\mathcal{B}_i] = O(i^{-A-1})$.

Now, for any $\lambda = o(1)$ it holds that

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda(\deg_{t+1}(s) - \deg_t(s))\right) \middle| G_t, \neg\mathcal{B}_t\right] \\
&= \left[\frac{\deg_t(s)}{t}(1 - p + p\exp(\lambda)) + \frac{t - \deg_t(s)}{t}\left(1 - \frac{r}{t} + \frac{r}{t}\exp(\lambda)\right) \middle| \neg\mathcal{B}_t\right] \\
&\leq \exp\left(\left(\frac{p\deg_t(s)}{t} + \frac{r(t - \deg_t(s))}{t^2}\right)(\lambda + \lambda^2)\right) \\
&\leq \exp\left(\lambda(A+1)\, C_1\, pt^{p-1} \log^2(t)\, (1 + o(1))\right).
\end{aligned}$$

Let us now denote $\mathcal{A}_k = \bigcup_{i=kt}^{(k+1)t-1} \mathcal{B}_i$ and observe that $\Pr[\mathcal{A}_k] = O(t^{-A})$. We proceed similarly to the proof of Theorem 4:

$$\begin{aligned}
&\Pr[\deg_{(k+1)t}(s) - \deg_{kt}(s) \geq d | G_{kt}] \\
&\leq \Pr[\deg_{(k+1)t}(s) - \deg_{kt}(s) \geq d | G_{kt}, \neg\mathcal{A}_k] \Pr[\neg\mathcal{A}] + \Pr[\mathcal{A}_k] \\
&\leq \exp(-\lambda d)\mathbb{E}\left[\exp\left(\lambda(\deg_{(k+1)t}(s) - \deg_{kt}(s))\right) | G_{kt}, \neg\mathcal{A}_k\right] + O(t^{-A}) \\
&\leq \exp(-\lambda d) \prod_{i=kt}^{(k+1)t-1} \mathbb{E}\left[\exp\left(\lambda(\deg_{i+1}(s) - \deg_i(s))\right) \middle| G_i, \neg\mathcal{B}_i\right] + O(t^{-A}) \\
&\leq \exp(-\lambda d) \prod_{i=kt}^{(k+1)t-1} \exp\left(\lambda\,(A+1)\, C_1\, i^{p-1} \log^2(i)(1 + o(1))\right) + O(t^{-A}) \\
&\leq \exp(-\lambda d) \exp\left(\sum_{i=kt}^{(k+1)t-1} \lambda\,(A+1)\, C_1\, i^{p-1} \log^2(t)(1 + o(1))\right) + O(t^{-A}) \\
&\leq \exp(-\lambda d) \exp\left(\lambda\,(A+1)\, C_2((k+1)^p - k^p)t^p \log^2(t)\right) + O(t^{-A})
\end{aligned}$$

for a certain constant C_2.

Therefore, it is sufficient to take $\lambda = \left(((k+1)^p - k^p)\log^2(t)\right)^{-1}$ and $d = AC_3((k+1)^p - k^p)t^p \log^2(t)$ for sufficiently large C_3 to obtain the final result. ◀

Now we return to the proof of the main theorem. Let $Z_k = \deg_{(k+1)t}(s) - \deg_{kt}(s)$. We know that for $p > 0$

$$\mathbb{E}Z_k = \mathbb{E}D(G_{(k+1)t}) - \mathbb{E}D(G_{kt}) = C_1\left((k+1)^p - k^p\right)t^p(1+o(1))$$

for some constant C_1.

Let now define the following events:

$$\mathcal{A}_1 = \left[Z_k \le \frac{t^p}{f(t)}\right]$$
$$\mathcal{A}_2 = \left[\frac{t^p}{f(t)} < Z_k \le C_2((k+1)^p - k^p)t^p\log^2(t)\right]$$
$$\mathcal{A}_3 = \left[Z_k > C_2((k+1)^p - k^p)t^p\log^2(t)\right]$$

for a constant C_2 such that (from the lemma above) $\Pr[\mathcal{A}_3] = O(t^{-2})$. Here $f(t)$ is any (monotonic) function such that $f(t) \to \infty$ as $t \to \infty$.

We know that

$$\mathbb{E}Z_k = \mathbb{E}\left[Z_k|\mathcal{A}_1\right]\Pr\left[\mathcal{A}_1\right] + \mathbb{E}\left[Z_k|\mathcal{A}_2\right]\Pr\left[\mathcal{A}_2\right] + \mathbb{E}\left[Z_k|\mathcal{A}_3\right]\Pr\left[\mathcal{A}_3\right]$$
$$\mathbb{E}Z_k \ge C_1\left((k+1)^p - k^p\right)t^{2p-1}$$
$$\mathbb{E}\left[Z_k|\mathcal{A}_1\right] \le \frac{t^{2p-1}}{f(t)}$$
$$\mathbb{E}\left[Z_k|\mathcal{A}_2\right] \le C_2((k+1)^p - k^p)t^p\log^2(t)$$
$$\mathbb{E}\left[Z_k|\mathcal{A}_3\right] \le (k+1)t$$

and therefore for sufficiently large t it holds that

$$\Pr[\mathcal{A}_1] \le \frac{C_2\left((k+1)^p - k^p\right)\log^2(t) - C_1\left((k+1)^p - k^p\right)}{C_2\left((k+1)^p - k^p\right)\log^2(t) - \frac{1}{f(t)}}$$
$$\le 1 - \frac{C_1}{2C_2\log^2(t)}.$$

Let now $\tau = kt$. Then,

$$\Pr\left[D(G_\tau) \le t^p f^{-1}(t)\right] = \Pr\left[\bigcap_{i=1}^{k} Y_i \le \frac{t^p}{f(t)}\right] \le \prod_{i=1}^{k}\left(1 - \frac{C_1}{2C_2\log^2(t)}\right).$$

Therefore, if we assume $k = \frac{2AC_2}{C_1}\log^3(t)$, we get

$$\Pr\left[D(G_\tau) \le \frac{t^p}{f(t)}\right] = \exp\left(-A\log(t)\right) = O(t^{-A})$$

and finally

$$\Pr\left[D(G_t) \le \frac{C_3}{A^p}t^p\log^{-3p-\varepsilon}(t)\right] = O(t^{-A}).$$

for some constant C_3 and any $\varepsilon > 0$.