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.szimon.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 degree of a given node over the evolution of the graph and the expected 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
Towards graphs compression: The degree distribution of duplication-divergence graphs

lattices [16]. Recently, we saw some work directed towards more complex data structures such as trees [10, 15] 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 Hoeffer [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-Rényi graphs was presented in [5], while recently it was extended to the preferential attachment model (also known as Barabasi-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. [17], referred to below as 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 \( t_0, t_0 + 1, \ldots, t \) 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 [?]).

Before we summarize our main results on the degree distribution in 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 \( G_n \) be the set of all labeled graphs on \( n \) vertices (with vertices having labels 1, 2, \ldots, \( n \)) and \( S_n \) be the set of all unlabeled graphs on \( n \) vertices. Then, the graph entropy and the structural graph entropy are defined as

\[
\begin{align*}
H(G) &= \sum_{G \in G_n} \Pr(G) \log \Pr(G), \\
H(S(G)) &= \sum_{S(G) \in S_n} \Pr(S(G)) \log \Pr(S(G)),
\end{align*}
\]

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:

\[
H(G) - H(S(G)) = \mathbb{E} \log |\text{Aut}(G)| - \mathbb{E} \log |\Gamma(G)|
\]
where $H(G)$ and $H(S(G))$ are, respectively, the entropy of the labelled and unlabelled graph generated by a given model, $\text{Aut}(G)$ 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 $\log |\text{Aut}(G)| = 0$ [5, 13], and therefore we do not consider these models as likely matches for these kinds of networks.

| Network                                      | Nodes  | Edges   | $\log |\text{Aut}(G)|$ |
|----------------------------------------------|--------|---------|------------------------|
| 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 [18, 21].

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. However, if we generate experimentally multiple random 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), $\log |\text{Aut}(G)|$. Moreover, as it was shown by Sreedharan et al. [18], 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 in duplication-divergence networks. Overall, the question of symmetry is tightly related to the behavior of the degree distribution, 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.

Turowski et al. showed in [20] 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 $\gamma$ 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 of a given vertex $s$ at time $t$ (denoted by $\text{deg}_t(s)$) and the average degree in the graph
(denoted by $D(G_t)$).

We present in Theorems 2–7 exact and precise asymptotics of these quantities when$t \to \infty$. We show that $\mathbb{E}[\text{deg}_t(s)]$ 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 $\mathbb{E}[\text{deg}_t(s)]$ grows respectively
like $(\frac{1}{2})^p \sqrt{\frac{2}{3} \log s}$ or $(\frac{1}{2})^p s^{p-1}$, 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 $\text{deg}_t(s)$ for $s = O(1)$. It turns out that it is sufficient to only go a
depolylogarithmic 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 $\mathbb{D}(t, p, r)$ model differs quite
substantially from other graph models such as the preferential attachment model [13, 22]. 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 $\mathbb{E}[\text{deg}_t(s)] = \Omega(t^p)$ while for $s = \Theta(t)$ (i.e., very young nodes) we have
$\mathbb{E}[\text{deg}_t(s)] = O(t^{2p-1})$. 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 $t^\varepsilon$ for some constant $\varepsilon > 0$ [13].

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

![Figure 1](image_url) Symmetry of graphs (log $|\text{Aut}(G)|$) generated by Pastor-Satorras model.
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 \( D(G) \) 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 \( \mathcal{D}(t, p, r) \) as follows: let \( 0 \leq p \leq 1 \) and \( 0 \leq r \leq t_0 \) be the parameters of the model. Let also \( G_{t_0} \) be a graph on \( t_0 \) vertices, with \( V(G_{t_0}) = \{1, \ldots, t_0\} \).

Now, for every \( t = t_0, t_0 + 1, \ldots \) 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 \( V(G_t) = \{1, \ldots, t\} \) uniformly at random and denote \( u \) as parent\((t + 1),
3. for every vertex \( i \in V(G_t)\):
   a. if \( i \in \mathcal{N}((\text{parent}(t + 1))) \) (which holds with probability \( \frac{\deg_t(s)}{t} \)) and we add an edge between \( i \) and \( t + 1 \) with probability \( p \),
   b. if \( i \notin \mathcal{N}((\text{parent}(t + 1))) \), 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 \( t \geq s \) we know that vertex \( s \) may be connected to vertex \( t + 1 \) in one of the following two cases:

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

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

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

After removing the conditioning on \( G_t \), we find:

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

This recurrence falls under a general recurrence of the form

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

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 9–14), providing exact and asymptotic behavior of \( \mathbb{E}[f(G_{n+1})] \). 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 9 to obtain the equation for the exact behavior of the degree of a given node \( s \) at time \( t \):

\[
E[\deg(s)] = E[\deg(s)] \prod_{k=s}^{t-1} \left( 1 + \frac{p}{k} - \frac{r}{k^2} \right) + \sum_{j=s}^{t-1} r \prod_{k=j+1}^{t-1} \left( 1 + \frac{p}{k} - \frac{r}{k^2} \right).
\]

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

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

\[
\textbf{Lemma 1.} \text{ For any } t \geq t_0 \text{ it holds that }
E[\deg(t+1)] = \left( p - \frac{r}{t} \right) E[D(G_t)] + r.
\]

It is quite intuitive that the expected degree of a new vertex behaves as if we would choose a vertex with the average degree \( E[D(G_t)] \) 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 \( E[D(G_t)] \), 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 \( G_{t+1} \):

\[
E[D(G_{t+1}) | G_t] = \frac{1}{t+1} \mathbb{E} \left[ \sum_{i=1}^{t+1} \deg(i+1) | G_t \right]
\]

\[
= \frac{1}{t+1} \mathbb{E} \left[ \sum_{i=1}^{t} \deg(i) + 2\deg(t+1) | G_t \right]
\]

\[
= \frac{1}{t+1} \left( tD(G_t) + 2E[\deg(t+1) | G_t] \right)
\]

\[
= \frac{1}{t+1} \left( tD(G_t) + 2E[\deg(t+1) | G_t] \right) = D(G_t) \left( 1 + \frac{2p-1}{t+1} - \frac{2r}{t(t+1)} \right) + \frac{2r}{t+1}.
\]

Therefore, after removing the conditioning on \( G_t \):

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

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 9 from the next section and derive Theorem 2. The proof is given in Appendix C.

\[
\textbf{Theorem 2.} \text{ For all } t \geq t_0 \text{ we have }
E[D(G_t)] = \frac{\Gamma(t+c_1)\Gamma(t+c_4)}{\Gamma(t)\Gamma(t+1)} \left( D(G_{t_0}) \frac{\Gamma(t_0+1)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)} + 2r \sum_{j=t_0}^{t-1} \frac{\Gamma(j+1)^2}{\Gamma(j+c_3+1)\Gamma(j+c_4+1)} \right).
\]
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

\[
\mathbb{E}[D(G_t)] = \begin{cases} 
\frac{2r}{t^{2r}} (1 + o(1)) & \text{if } p < \frac{1}{2} \text{ and } r > 0, \\
\frac{2r}{t^{2r}} \ln t (1 + o(1)) & \text{if } p = \frac{1}{2} \text{ and } r > 0, \\
\left( D(G_{t_0}) + \frac{2r_{t_0} + 3t_0}{t_0 + 2r_{t_0} - 2r} \right) & \text{if } p > \frac{1}{2} \text{ or } r = 0,
\end{cases}
\]

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

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

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 it holds that

\[
\Pr[D(G_t) \geq AC \log^2(t)] = O(t^{-A}) \quad \text{for } p < \frac{1}{2},
\]

\[
\Pr[D(G_t) \geq AC \log^3(t)] = O(t^{-A}) \quad \text{for } p = \frac{1}{2},
\]

\[
\Pr[D(G_t) \geq AC t^{3p-1} \log^2(t)] = O(t^{-A}) \quad \text{for } p > \frac{1}{2},
\]

for some fixed constant \( C > 0 \) and any \( A > 0 \).

Similarly, we have the behavior of the left tail:

**Theorem 4.** For \( p > \frac{1}{2} \) asymptotically it holds that

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

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

Note that since \( D(G_t) = O(\log t) \) for \( p \leq \frac{1}{2} \), 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.
Asymptotically as $t \to \infty$ it holds that

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

where $c_3$, $c_4$ are as above.

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

$$E[\deg_t(s)] = \begin{cases} pt^{2p-1} \frac{\Gamma(t_0)^2}{\Gamma(t_0 + c_3 + 1) \Gamma(t_0 + c_4 + 1)} D(G_{t_0})(1 + o(1)) & \text{if } p \leq \frac{1}{2}, r = 0, \\ \frac{2r}{1 - 2p}(1 + o(1)) & \text{if } p < \frac{1}{2}, r > 0, \\ 2pr \ln t (1 + o(1)) & \text{if } p = \frac{1}{2}, r > 0, \\ \frac{\Gamma(t_0) \Gamma(t_0 + 1)}{\Gamma(t_0 + c_3 + 1) \Gamma(t_0 + c_4 + 1)} pt^{2p-1} (1 + o(1)) & \text{if } p > \frac{1}{2}, \\ \frac{2r_{t_0}}{t_0(t_0 + 2p_0 - 2)} F_2 \left[ \begin{array}{c} t_0 + 1, t_0 + 1, 1 \\ t_0 + c_3 + 1, t_0 + c_4 + 1 \end{array} ; 1 \right] \end{cases}$$

with the same notation as in Theorem 2.

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

We are finally in a position to state the exact and asymptotic expressions for $E[\deg_t(s)]$. This we need to split in two parts: first, for the initial vertices of $G_{t_0}$ ($1 \leq s \leq t_0$) 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 $1 \leq s \leq t_0$ it is true that

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

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

Asymptotically as $t \to \infty$:

$$E[\deg_t(s)] = \begin{cases} r \ln t (1 + o(1)) & \text{if } p = 0 \text{ and } r > 0, \\ \frac{pt}{\Gamma(t_0) \Gamma(t_0 + c_1) \Gamma(t_0 + c_2)} \deg_{t_0}(s) & \text{if } p = 0 \text{ and } r > 0, \\ \frac{r \Gamma(t_0) \Gamma(t_0 + 1)}{\Gamma(t_0 + c_1 + 1) \Gamma(t_0 + c_2 + 1)} F_2 \left[ \begin{array}{c} t_0, t_0 + 1 \\ t_0 + c_1 + 1, t_0 + c_2 + 1 \end{array} ; 1 \right] (1 + o(1)) & \text{if } p > 0 \text{ or } r = 0. \end{cases}$$

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 $E[\deg_t(s)]$. 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 $t^p$. 
\[ \textbf{Theorem 7.} \quad \text{For all } t_0 < s < t \text{ it is true that} \]
\[ \mathbb{E}[\deg(s)] = \frac{\Gamma(t+c)\Gamma(t^2)}{\Gamma(t)^2} \]
\[ + \left( ps - p - r \right) \frac{\Gamma(s+c-1)\Gamma(s+c-1)}{\Gamma(s+c)\Gamma(s+c)} \]
\[ + \left( D(G_{t_0}) \frac{\Gamma(t_0)^2}{\Gamma(t_0+c)\Gamma(t_0+c)} + 2r \sum_{j=t_0}^{s-2} \frac{\Gamma(j+1)^2}{\Gamma(j+c+1)\Gamma(j+c+1)} \right) \]
\[ + \frac{r\Gamma(s)^2}{\Gamma(s+c)\Gamma(s+c)} + r \sum_{j=s}^{t-1} \frac{\Gamma(j+1)}{\Gamma(j+c+1)\Gamma(j+c+1)} \]}

where \( c_1 - c_4 \) are as above.

Asymptotically as \( t \to \infty \):

1. for \( s = O(1) \)
\[ \mathbb{E}[\deg(s)] = t^p(1+o(1)) \]
\[ \left( ps - p - r \right) \frac{\Gamma(s+c-1)\Gamma(s+c-1)}{\Gamma(s+c)\Gamma(s+c)} \]
\[ + \left( D(G_{t_0}) \frac{\Gamma(t_0)^2}{\Gamma(t_0+c)\Gamma(t_0+c)} + 2r \sum_{j=t_0}^{s-2} \frac{\Gamma(j+1)^2}{\Gamma(j+c+1)\Gamma(j+c+1)} \right) \]
\[ + \frac{r\Gamma(s)^2}{\Gamma(s+c)\Gamma(s+c)} \left( 1 + F \left[ \frac{s_{s+1}}{s^2 + ps - r} \right] \right) \]

2. for \( s = \omega(1) \) and \( s = o(t) \)
\[ \mathbb{E}[\deg(s)] = \left\{ \begin{array}{ll}
D(G_{t_0}) \frac{p\Gamma(t_0)^{2p}}{\Gamma(t_0+c)^{2p+1}} \left( \frac{1}{2} \right)^p s^{2p-1}(1+o(1)) & \text{if } p \leq \frac{1}{2}, r = 0, \\
\log \left( \frac{1}{2} \right)(1+o(1)) & \text{if } p = 0, r > 0, \\
\frac{r(1-p)}{p(1-2p)} \left( \frac{1}{2} \right)^p (1+o(1)) & \text{if } 0 < p < \frac{1}{2}, r > 0, \\
\frac{r\Gamma(s)}{\Gamma(s+c)} \log s (1+o(1)) & \text{if } p = \frac{1}{2}, r > 0, \\
\left( D(G_{t_0}) + \frac{2r(t_0-c)}{\Gamma(t_0+c)^2} \right) \frac{3F_{2}}{s_{s+1}} \left[ \frac{t_0+c}{s^2 + ps - r} \right] & \text{if } p > \frac{1}{2}, r > 0
\end{array} \right. \]

3. for \( s = ct - o(t), 0 < c \leq 1 \)
\[ \mathbb{E}[\deg(s)] = \left\{ \begin{array}{ll}
D(G_{t_0}) \frac{p\Gamma(t_0)^{2p}}{\Gamma(t_0+c)^{2p+1}} s^{2p-1}(1+o(1)) & \text{if } p \leq \frac{1}{2}, r = 0, \\
(1 - \log c)(1+o(1)) & \text{if } p = 0, r > 0, \\
\frac{r(1-p)}{p(1-2p)} \left( \frac{1}{2} \right)^p (1+o(1)) & \text{if } 0 < p < \frac{1}{2}, r > 0, \\
\frac{r\Gamma(t)}{\Gamma(t+c)} \log t (1+o(1)) & \text{if } p = \frac{1}{2}, r > 0, \\
\left( D(G_{t_0}) + \frac{2r(t_0-t)}{\Gamma(t_0+c)^2} \right) \frac{3F_{2}}{s_{s+1}} \left[ \frac{t_0+c}{s^2 + ps - r} \right] & \text{if } p > \frac{1}{2}, r > 0
\end{array} \right. \]

The theorem above shows that there is a threefold behavior with respect to the range of \( s \): small (constant), medium (growing, but slower than \( t \)), and large (when \( s \) is directly proportional to \( t \)). In the first case we observe a behavior very similar to the one.
Towards graphs compression: The degree distribution of duplication-divergence graphs

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 $\mathbb{E}[\deg_t(s)] = \Theta((\frac{1}{2})^p s^{2p-1})$. When the majority of the edges are created due to the random addition (for $r > 0$ and $p < \frac{1}{2}$), then $\mathbb{E}[\deg_t(s)] = \Theta((\frac{1}{2})^p)$. Finally, we observe a phase transition for $p = \frac{1}{2}$, $r = 0$ with $\mathbb{E}[\deg_t(s)] = \Theta((\frac{1}{2})^p \log s)$. 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 [18] and [21] 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 results:

> **Theorem 8.** Asymptotically for $s = O(1)$ it holds that

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

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

We note additionally, that since $\deg_t(s)$ is closely dependent on the degree distribution in $G_{t-1}$, it is very unlikely that for $s$ close to $t$ the analogous bounds 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 [18], 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


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

A Useful lemmas

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

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

(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 \( E[\deg_h(v)] \) and \( E[D(G_t)] \) (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 9. Let \( (G_n)_{n=n_0}^{\infty} \) be a Markov process for which \( E[f(G_{n_0})] > 0 \) 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

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

is a martingale.

Proof. Observe that

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

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

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

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

\[ = f(G_{n_0}) \sum_{k=n_0}^{n-1} g_1(k) + \sum_{j=n_0}^{n-1} g_2(j) \prod_{k=n_0}^{j} g_1(k) \]

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 10. Let $W_1(k)$, $W_2(k)$ be polynomials of degree $d$ with respective roots $a_i$, $b_i$ $(i = 1, \ldots, d)$, 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 11 (Abramowitz, Stegun [1]). For any $a, b \in \mathbb{R}$ if $n \to \infty$, then

$$\frac{\Gamma(n + a)}{\Gamma(n + b)} = n^{a-b} \sum_{k=0}^{\infty} \binom{a-b}{k} B_k(a-b+1)(a) \cdot n^{-k}$$

$$= n^{a-b} \left( 1 + \frac{(a-b)(a-b+1)}{2n} + O \left( \frac{1}{n^2} \right) \right),$$

where $B_k(t)$ 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 12. 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

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

Proof. We estimate the sum using Lemma 11 and the Euler-Maclaurin formula [19, p. 294]

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

$$= \left[ j^{a-b+1} \left( \frac{1}{a-b+1} + O \left( \frac{1}{n} \right) \right) \right]_{n_0}^{n} = n^{a-b+1} \left( \frac{1}{a-b+1} + O \left( \frac{1}{n} \right) \right) + O(1)$$

which completes the proof.

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 asymptotically

$$\sum_{j=n_0}^{n} \frac{\prod_{i=1}^{k} \Gamma(j + a_i)}{\prod_{i=1}^{k} \Gamma(j + b_i)} = \ln n + O(1)$$
Towards graphs compression: The degree distribution of duplication-divergence graphs

**Proof.** We proceed as before

\[
\sum_{j=n_0}^n \prod_{i=1}^k \frac{\Gamma(j + a_i)}{\Gamma(j + b_i)} = \sum_{j=n_0}^n \frac{1}{j} \left(1 + O \left(\frac{1}{j}\right)\right) \int_{n_0}^n \frac{1}{j} \left(1 + O \left(\frac{1}{j}\right)\right) \, dj = \ln(n + O(1))
\]

which completes the proof.

**Lemma 14.** 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

\[
\sum_{j=n}^\infty \prod_{i=1}^k \frac{\Gamma(j + a_i)}{\Gamma(j + b_i)} = \prod_{i=1}^k \frac{\Gamma(n + a_i)}{\Gamma(n + b_i)} \frac{\text{E}_k \left[ n+1, \ldots, n+b_k ; 1 \right]}{\text{E}_k \left[ n+1, \ldots, n+b_k ; 1 \right]}
\]

where \(\text{E}_k \left[ a_i ; z \right]\) is the generalized hypergeometric function. Moreover it is true that asymptotically

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

**Proof.** The proof of the first formula follows directly from the definition of the generalized hypergeometric function. Second formula follows from Lemma 11, as we know that for \(n \to \infty\):

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

\[
= \left[j^{a-b+1} \left(\frac{1}{b-a-1} + O \left(\frac{1}{j}\right)\right)\right]_n^\infty = n^{a-b+1} \left(\frac{1}{b-a-1} + O \left(\frac{1}{n}\right)\right)
\]

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 \(N_t(\text{parent}(t + 1))\) (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

\[
\text{deg}_{t+1}(t+1) \sim \text{Bin} \left(\text{deg}_t(\text{parent}(t + 1)), p\right) + \text{Bin} \left(t - \text{deg}_t(\text{parent}(t + 1)), \frac{r}{t}\right).
\]

Hence

\[
\mathbb{E}[	ext{deg}_{t+1}(t+1) \mid G_t] = \sum_{k=0}^t \Pr(\text{deg}_t(\text{parent}(t + 1)) = k) \sum_{a=0}^k \binom{k}{a} p^a (1-p)^{k-a}
\]

\[
\sum_{b=0}^{t-k} \binom{t-k}{b} \left(\frac{r}{t}\right)^b \left(1 - \frac{r}{t}\right)^{t-k-b} (a + b)
\]

\[
= \sum_{k=0}^t \Pr(\text{deg}_t(\text{parent}(t + 1)) = k) \left(p k + \frac{r}{t}(t - k)\right)
\]
\( = \left(p - \frac{r}{t}\right) \sum_{k=0}^{t} k \Pr(\text{deg}_t(\text{parent}(t+1)) = k) + r.\)

Since parent sampling is uniform, we know that \(\Pr(\text{parent}(t+1) = i) = \frac{1}{t}\) and therefore
\[
 D(G_t) = \sum_{i=1}^{t} \Pr(\text{parent}(t+1) = i) \text{deg}_t(i) = \sum_{k=0}^{t} k \Pr(\text{deg}_t(\text{parent}(t+1)) = k).\]

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

### 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 9 and 10 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 12, 13 and 14 for the respective ranges of \(p\).

Finally, we proceed to the proof of Theorems 6 and 7. First, we apply Lemma 9 with \(g_1(t) = 1 + \frac{p-t}{r} - \frac{r}{s}\) and \(g_2(t) = \frac{r}{j}\) to Eqn. (1) and we obtain aforementioned Eqn. (3). Now we combine this result with Lemma 10. First, we if we apply it for \(1 \leq s \leq t_0\) 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 \(G_t\):

\[
\mathbb{E}[\text{deg}_t(s)] = \frac{\Gamma(t+c_1)\Gamma(t+c_2)}{\Gamma(t)^2} \left( \mathbb{E}[\text{deg}_s(s)] \right) + \frac{\Gamma(s)^2}{\Gamma(s+c_1)\Gamma(s+c_2)} + \sum_{j=s}^{t-1} \frac{r\Gamma(j)\Gamma(j+1)}{\Gamma(j+c_1+1)\Gamma(j+c_2+1)}\]

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 \(\mathbb{E}[\text{deg}_t(s)].\)

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 12, 13 and 14.

### Proof of Theorem 3

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

\[\text{Lemma 15. For any } \lambda = O\left(\frac{1}{t}\right) \text{ it holds that} \]
\[\mathbb{E}[\exp(\lambda \text{deg}_{t+1}(t+1))|G_t] \leq \exp(\lambda pD(G_t)(1 + O(\lambda t)) + \lambda r(1 + O(\lambda))).\]

\[\text{Proof.} \]
\[\mathbb{E}[\exp(\lambda \text{deg}_{t+1}(t+1))|G_t] = \frac{1}{t} \sum_{i=1}^{t} \mathbb{E}\left[\exp\left(\lambda \text{Bin}(\text{deg}_t(i), p) + \lambda \text{Bin}\left(\text{deg}_t(i), \frac{r}{t}\right)\right)\right]|G_t] \]
Towards graphs compression: The degree distribution of duplication-divergence graphs

\[ \frac{1}{t} \sum_{i=1}^{t} (1 - p + p \epsilon^r)^{\text{deg}_\epsilon(i)} \left( 1 - \frac{r}{t} + \frac{r}{t} \epsilon^r \right)^{\text{deg}_\epsilon(i)}. \]

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 \( x \):

\[ \mathbb{E} [\exp(\lambda \text{deg}_{\epsilon+1}(t+1)|G_t)] \]

\[ \leq \frac{1}{t} \sum_{i=1}^{t} (1 + p\lambda(1 + O(\lambda)))^{\text{deg}_\epsilon(i)} \left( 1 + \frac{r\lambda}{t} (1 + O(\lambda)) \right)^{\text{deg}_\epsilon(i)} \]

\[ \leq \frac{1}{t} \sum_{i=1}^{t} (1 + p\lambda \text{deg}_\epsilon(i)(1 + O(\lambda t)) (1 + r\lambda(1 + O(\lambda))) \]

\[ \leq \frac{1}{t} \sum_{i=1}^{t} (1 + p\lambda \text{deg}_\epsilon(i)(1 + O(\lambda t))) \exp(r\lambda(1 + O(\lambda))) \]

\[ = (1 + p\lambda D(G_t)(1 + O(\lambda))) \exp(r\lambda(1 + O(\lambda))) \]

\[ \leq \exp(\lambda p D(G_t)(1 + O(\lambda t)) + \lambda r(1 + O(\lambda))). \]

Now we are ready to finally prove the theorem.

\[ \mathbb{E} [\exp(\lambda_{t+1} D(G_{t+1}) | G_t)] = \mathbb{E} \left[ \exp \left( \frac{\lambda_{t+1}}{t+1} D(G_t) \right) \left( 1 + O(\lambda_{t+1}) \right) \left( 1 + \frac{2r\lambda_{t+1}}{t+1} \right) \right] \]

\[ \leq \exp \left( \lambda_{t+1} D(G_t) \left( 1 - \frac{2p-1}{t+1} \right) (1 + O(\lambda_{t+1})) + \frac{2r\lambda_{t+1}}{t+1} (1 + o(t^{-1})) \right). \]

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

\[ \lambda_{t_0} \in \left[ \lambda_t \prod_{k=t_0}^{t-1} \left( 1 + \frac{2p-1}{k+1} \right), \lambda_t \prod_{k=t_0}^{t-1} \left( 1 + \left( \frac{2p-1}{k+1} \right) (1 + O(\varepsilon_t)) \right) \right] \]

\[ \subseteq \left[ \lambda_t \left( \frac{t}{t_0} \right)^{2p-1} (1 + o(1)), \lambda_t \left( \frac{t}{t_0} \right)^{(2p-1)(1+O(\varepsilon_t))} (1 + o(1)) \right] \]

It follows that

\[ \mathbb{E} [\exp(\lambda_{t} D(G_t))] \leq \exp(\lambda_{t_0} D(G_{t_0})) \prod_{k=t_0}^{t-1} \exp \left( \frac{2r\lambda_{k+1}}{k+1} (1 + o(k^{-1})) \right) \]
We start the proof by obtaining a simple lemma, analogous to Lemma 15:

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

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

**Proof.**

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

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

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

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 bounds it follows that

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

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

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

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 15:

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

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

**Proof.**

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

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

\[
\mathbb{E} \left[ \exp(\lambda \deg_{t+1}(t+1)|G_t) \right] \leq \frac{1}{t} \sum_{i=1}^{t} (1 + p\lambda(1 + O(\lambda))^{\deg_t(i)} \left( 1 + \frac{r\lambda}{t} (1 + O(\lambda)) \right)^{t - \deg_t(i)}
\]
Towards graphs compression: The degree distribution of duplication-divergence graphs

\[ \leq \frac{1}{t} \sum_{i=1}^{t} \left( 1 + 2p \lambda \deg_t(i) (1 + O(\lambda)) \right) (1 + 2r \lambda (1 + O(\lambda))) \]

\[ \leq \frac{1}{t} \sum_{i=1}^{t} \left( 1 + 2p \lambda \deg_t(i) (1 + O(\lambda)) \right) \exp(2r(1 + O(\lambda))) \]

\[ = \left( 1 + 2p \lambda D(G_t)(1 + O(\lambda)) \right) \exp(2r(1 + O(\lambda))) \]

\[ \leq \exp(2\lambda p D(G_t)(1 + O(\lambda)) + 2\lambda r(1 + O(\lambda))). \]

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

**Lemma 17.** Let \( p > \frac{1}{2} \). For sufficiently large \( t \) and all \( k < t \) it is true that

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

for some fixed constant \( A > 0 \).

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

Now, we note that from Lemma 15 for any \( \lambda = o(1) \)

\[ \mathbb{E} \left[ \exp(\lambda (D(G_{(k+1)t}) - D(G_{kt})) \bigg| G_t, \neg B_i \right] \]

\[ \leq \mathbb{E} \left[ \exp \left( \frac{2\lambda p}{t+1} \deg_{t+1}(i)(t+1) \bigg| G_t, \neg B_i \right] \]

\[ \leq \exp \left( \frac{2\lambda p}{t+1} D(G_t)(1 + O(\lambda)) + \frac{2\lambda r}{t+1} (1 + O(\lambda)) \bigg| \neg B_i \right] \]

\[ \leq \exp(\lambda (A + 1)C_2 2^{p-2} log^2(t)(1 + o(1))) \]

for a certain constant \( C_2 \).

Now we proceed as following:

\[ \Pr[D(G_{(k+1)t}) - D(G_{kt}) \geq d|G_{kt}] \]

\[ \leq \Pr[D(G_{(k+1)t}) - D(G_{kt}) \geq d|G_{kt}, \neg A_k] \Pr[\neg A] + \Pr[A_k] \]

\[ \leq \exp(-\lambda d) \sum_{i=kt}^{(k+1)t-1} \mathbb{E} \left[ \exp(\lambda (D(G_{(i+1)t}) - D(G_{it})) \bigg| G_{it}, \neg B_i \right] + O(t^{-A}) \]

\[ \leq \exp(-\lambda d) \sum_{i=kt}^{(k+1)t-1} \exp(\lambda (A + 1)C_2 2^{p-2} log^2(i)(1 + o(1))) + O(t^{-A}) \]

\[ \leq \exp(-\lambda d) \left( \sum_{i=kt}^{(k+1)t-1} \lambda (A + 1)C_3 2^{p-2} log^2(t)(1 + o(1)) \right) + O(t^{-A}) \]

\[ \leq \exp(-\lambda d) \left( \sum_{i=kt}^{(k+1)t-1} \lambda (A + 1)C_3 ((k + 1)^{2p-1} - k^{2p-1})^2 log^2(t) \right) + O(t^{-A}) \]

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 \( d = AC_4((k + 1)^{2p-1} - k^{2p-1}) log^2(t) \) for sufficiently large \( C_4 \) to obtain the final result. \( \blacksquare \)
Now we may return to the main theorem. Let $Y_k = D(G_{(k+1)t}) - D(G_{kt})$. We know that for $p > \frac{1}{2}$

\[ EY_k = E D(G_{(k+1)t}) - E D(G_{kt}) = C_1 \left((k+1)^{2p-1} - k^{2p-1}\right) t^{2p-1}(1 + o(1)) \]

for some constant $C_1$.

Let now define the following events:

\[ A_1 = \left[Y_k \leq \frac{t^{2p-1}}{f(t)} \right] \]
\[ A_2 = \left[\frac{t^{2p-1}}{f(t)} < Y_k \leq C_2((k+1)^{2p-1} - k^{2p-1}) t^{2p-1} \log^2(t) \right] \]
\[ A_3 = \left[Y_k > C_2((k+1)^{2p-1} - k^{2p-1}) t^{2p-1} \log^2(t) \right] \]

for a constant $C_2$ such that (from the lemma above) $\Pr[A_3] = O(t^{-2})$. Here $f(t)$ is any (monotonic) function such that $f(t) \to \infty$ as $t \to \infty$.

We know that

\[ EY_k = E \left[Y_k | A_1\right] \Pr[A_1] + E \left[Y_k | A_2\right] \Pr[A_2] + E \left[Y_k | A_3\right] \Pr[A_3] \]
\[ EY_k \geq C_1 \left((k+1)^{2p-1} - k^{2p-1}\right) t^{2p-1} \]

\[ E \left[Y_k | A_1\right] \leq \frac{t^{2p-1}}{f(t)} \]
\[ E \left[Y_k | A_2\right] \leq C_2((k+1)^{2p-1} - k^{2p-1}) t^{2p-1} \log^2(t) \]
\[ E \left[Y_k | A_3\right] \leq (k+1) t \]

and therefore for sufficiently large $t$ it holds that

\[ \Pr[A_1] \leq \frac{C_2 \left((k+1)^{2p-1} - k^{2p-1}\right) \log^2(t) - C_1 \left((k+1)^{2p-1} - k^{2p-1}\right)}{C_2 \left((k+1)^{2p-1} - k^{2p-1}\right) \log^2(t) - \frac{1}{f(t)}} \]
\[ \leq 1 - \frac{C_1}{2C_2 \log^2(t)} \]

Let now $\tau = kt$.

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

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

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

and finally

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

for some constant $C_3$ and any $\varepsilon > 0$. 
Proof of Theorem 8

\[ E \left[ \exp \left( \lambda_{t+1} \deg_{t+1}(s) \right) \mid G_s \right] = \]
\[ = \left( \frac{\deg_G(s)}{t} + \frac{t - \deg_G(s)}{t} \right) \exp \left( \lambda_{t+1} \left( \deg_G(s) + 1 \right) \right) \]
\[ + \left( \frac{\deg_G(s)}{t} \right) \left( 1 - p \right) + \frac{t - \deg_G(s)}{t} \left( 1 - \frac{r}{t^2} \right) \exp \left( \lambda_{t+1} \deg_G(s) \right) \]
\[ = \exp \left( \lambda_{t+1} \deg_G(s) \right) \]
\[ \left( \frac{\deg_G(s)}{t} \right) \left( 1 - p + p \exp \left( \lambda_{t+1} \right) \right) + \frac{t - \deg_G(s)}{t} \left( 1 - \frac{r}{t^2} \exp \left( \lambda_{t+1} \right) \right) \]
\[ \leq \exp \left( \lambda_{t+1} \deg_G(s) \right) \left( 1 + \left( \frac{p \deg_G(s)}{t} + \frac{r (t - \deg_G(s))}{t^2} \right) \left( \lambda_{t+1} + \lambda_{t+1}^2 \right) \right) \]
\[ \leq \exp \left( \lambda_{t+1} \deg_G(s) \right) \left( 1 + \left( \frac{p \deg_G(s)}{t} + \frac{r (t - \deg_G(s))}{t^2} \right) \left( \lambda_{t+1} + \lambda_{t+1}^2 \right) \right) \]
\[ \left( 1 + \left( \frac{p \deg_G(s)}{t} + \frac{r (t - \deg_G(s))}{t^2} \right) \left( 1 + \lambda_{t+1} \right) \right) \exp \left( \lambda_{t+1} \left( 1 + \lambda_{t+1} \right) \right). \]

Let us assume that \( \lambda_k = o(1) \) for all \( s \leq k \leq t \). Then for all \( k = s, s+1, \ldots, t \) we have
\[ \lambda_k = \lambda_{k+1} \left( 1 + \left( \frac{p}{k} \right) \left( 1 + \lambda_{k+1} \right) \right) \leq \lambda_{k+1} \left( 1 + \left( \frac{p}{k} \right) \left( 1 + \varepsilon_t \right) \right) \]
which lead us to
\[ \lambda_s \leq \lambda_t \prod_{k=s}^{t-1} \left( 1 + \left( \frac{p}{k} \right) \left( 1 + \varepsilon_t \right) \right) \leq \lambda_t \exp \left( \left( 1 + \varepsilon_t \right) \sum_{k=s}^{t-1} \left( \frac{p}{k} \right) \left( 1 + \varepsilon_t \right) \right) \]
\[ \leq \lambda_t \exp \left( \left( 1 + \varepsilon_t \right) \int_s^t \left( \frac{p}{k} \right) \frac{r}{k^2} \, dk \right) = \lambda_t \exp \left( \left( 1 + \varepsilon_t \right) \left( p \ln \frac{t}{s} + r \left( 1 + \frac{1}{s} \right) \right) \right) \]
\[ \leq \lambda_t \left( \frac{t}{s} \right)^{p(1+\varepsilon_t)} \exp \left( \frac{r}{t} \left( 1 + \varepsilon_t \right) \right). \]

It follows that
\[ E \left[ \exp \left( \lambda_t \deg_G(s) \right) \mid G_s \right] \leq \exp \left( \lambda_s \deg_G(s) \right) \prod_{k=s}^{t-1} \exp \left( \lambda_{k+1} \left( 1 + \lambda_{k+1} \right) \right) \]
\[ \leq \exp \left( \lambda_s \deg_G(s) \right) \exp \left( \varepsilon_t \left( 1 + \varepsilon_t \right) \frac{r}{t} \ln \frac{t}{s} \right) \leq \exp \left( \lambda_s \deg_G(s) \right) \exp \left( \frac{r}{s} (1+\varepsilon_t) \right) \]

Now, let \( \lambda_t = \varepsilon_t \left( \frac{t}{s} \right)^{-p(1+\varepsilon_t)} \exp \left( -\frac{r}{t} \left( 1 + \varepsilon_t \right) \right) \) so that \( \lambda_s \leq \varepsilon_t \). Then, from Chernoff bound it follows that
\[ \Pr[\deg_G(s) \geq \alpha \mathbb{E}[\deg_G(s)] \mid G_s] = \Pr[\exp(\deg_G(s) - \alpha \mathbb{E}[\deg_G(s)]) \geq 1 \mid G_s] \]
\[ \leq \exp \left( -\alpha \lambda_s \mathbb{E}[\deg_G(s)] \right) \mathbb{E}[\exp \left( \lambda_s \deg_G(s) \right) \mid G_s] \]
\[ \leq \exp \left( -\alpha \lambda_s \mathbb{E}[\deg_G(s)] \right) \exp \left( \lambda_s \deg_G(s) \right) \left( \frac{t}{s} \right)^{r \varepsilon_t (1+\varepsilon_t)} \).

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 \( \mathbb{E}[\deg_G(s)] = C_1 t^p \) and therefore
\[ \Pr[\deg_G(s) \geq \alpha C_1 t^p \mid G_s] \leq \exp \left( -\alpha C_2 \varepsilon_t t^{-p(1+\varepsilon_t)} \right) \exp \left( \varepsilon_t \deg_G(s) \right) \left( \frac{t}{s} \right)^{r \varepsilon_t (1+\varepsilon_t)} \]
\[ \leq \exp \left( \frac{-\alpha C_3}{\ln t} \frac{\deg_t(s)}{\ln t} \exp(2r) \right) \]

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.

\[ \text{G} \quad \text{Proof of Theorem} \]

We start by showing two lemmas, giving us the crude lower bound on the degree of a given vertex:

\[ \text{Lemma 18. Let } s = O(1). \text{ Then asymptotically as } t \to \infty, \text{ if } r > 0, \text{ then} \]

\[ \Pr \left[ \deg_t(s) \leq \frac{C}{\sqrt{A}} \ln t \right] = O(t^{-A}) \]

for some constant \(C\) and any \(A > 0\).

**Proof.** Let \(X_k \sim \text{Bin}(k, \frac{r}{t})\), \(Y_i = \sum_{k=t+1}^{t} X_k\). Then

\[ \mathbb{E} Y_i = \sum_{k=t+1}^{t} \mathbb{E} X_k = r \ln \frac{t}{l} + O(1). \]

We note that if \(l = \min \{ s, \frac{r}{p} \}\), we have \(\deg_t(s) \geq \deg_t(s) - \deg_t(s) \geq Y_i\). Therefore, from the Chernoff bound

\[ \Pr[\deg_t(s) \leq r(1 - 2\delta) \ln t] \leq \Pr[\deg_t(s) \leq r(1 - \delta) \ln \frac{t}{l}] \leq \Pr[Y_i \leq r(1 - \delta) \ln \frac{t}{l}] \]

\[ \leq \exp \left( -\frac{\delta^2}{2} \mathbb{E} Y_i \right) = O \left( t^{-\frac{2\delta^2}{t}} \right), \]

and it is sufficient to pick \(\delta = \sqrt{\frac{2A}{t}}\) to finish the proof. \(\blacksquare\)

Now we may go to the proof of the theorem.

\[ \mathbb{E} \left[ \exp \left( -\mu_{t+1} \deg_{t+1}(s) \right) \mid G_t \right] \]

\[ = \left( \frac{\deg_t(s)}{l} (1 - p) + \frac{t - \deg_t(s)}{l} r \right) \exp \left( -\mu_{t+1} \frac{\deg_t(s)}{l} \right) \]

\[ + \left( \frac{\deg_t(s)}{l} (1 - p) + \frac{t - \deg_t(s)}{l} (1 - \frac{r}{l}) \right) \exp \left( -\mu_{t+1} \deg_t(s) \right) \]

\[ = \exp \left( -\mu_{t+1} \deg_t(s) \right) \]

\[ \left( \frac{\deg_t(s)}{l} (1 - p + r \exp(-\mu_{t+1})) + \frac{t - \deg_t(s)}{l} \left( 1 - \frac{r}{l} + \frac{r}{l} \exp(-\mu_{t+1}) \right) \right) \]

\[ \leq \exp \left( -\mu_{t+1} \deg_t(s) \right) \left( 1 + \frac{p \deg_t(s)}{l} + \frac{r (t - \deg_t(s))}{l^2} \right) \left( -\mu_{t+1} + \mu_{t+1}^2 \right) \]

\[ \leq \exp \left( -\mu_{t+1} \deg_t(s) + \frac{p \deg_t(s)}{l} + \frac{r (t - \deg_t(s))}{l^2} \right) \left( -\mu_{t+1} + \mu_{t+1}^2 \right) \]

\[ = \exp \left( -\mu_{t+1} \deg_t(s) \left( 1 + \left( \frac{p}{l} - \frac{r}{l^2} \right) \left( 1 - \mu_{t+1} \right) \right) \right) \exp \left( -\mu_{t+1} (1 - \mu_{t+1}) \frac{r}{l} \right). \]
Let us assume that \( \mu_k \leq \varepsilon t = o(1) \) for all \( s \leq k \leq t \). Then for all \( k = s, s + 1, \ldots, t \) we have

\[
\mu_k = \mu_{k+1} \left(1 + \left( \frac{p}{k} - \frac{r}{k^2} \right) (1 - \mu_{k+1}) \right) \geq \mu_{k+1} \left(1 + \left( \frac{p}{k} - \frac{r}{k^2} \right) (1 - \varepsilon t) \right)
\]

which lead us to

\[
\mu_s \geq \mu_t \prod_{k=s}^{t-1} \left(1 + \left( \frac{p}{k} - \frac{r}{k^2} \right) (1 - \varepsilon t) \right) \geq \mu_t \exp \left( (1 - \varepsilon t) \sum_{k=s}^{t-1} \left( \frac{p}{k} - \frac{r}{k^2} \right) \right)
\]

\[
\geq \mu_t \exp \left( (1 - \varepsilon t) t \left( \frac{p}{s} - \frac{r}{s^2} \right) \right) = \mu_t \exp \left( (1 - \varepsilon t) \left( p \ln \frac{t}{s} + r \left( \frac{1}{t} - \frac{1}{s} \right) \right) \right)
\]

\[
\geq \mu_t \left( \frac{t}{s} \right)^{p(1-\varepsilon t)} \exp \left( \frac{r}{t} (1 - \varepsilon t) \right)
\]

and

\[
\mu_s \leq \mu_t \prod_{k=s}^{t-1} \left(1 + \left( \frac{p}{k} - \frac{r}{k^2} \right) \right) \leq \mu_t \left( \frac{t}{s} \right)^p \exp \left( \frac{r}{t} \right).
\]

It follows that for any \( s \leq v \leq t \)

\[
E \left[ \exp \left( -\mu_s \deg_t(s) \right) \right] \leq E \left[ \exp \left( -\mu_v \deg_v(s) \right) \right] \leq E \left[ \exp \left( -\mu_{t+1} \right) \right] \leq E \left[ \exp \left( -\mu_s \deg_s(s) \right) \right] \leq E \left[ \exp \left( -\mu_{s^*} \deg_{s^*}(s) \right) \right].
\]

Now, let \( \mu_t = \varepsilon t \left( \frac{1}{s} \right)^p \exp \left( -\frac{r}{t} \right) \) so that \( C \varepsilon t \leq \varepsilon t \left( \frac{1}{s} \right)^p \exp \left( -\frac{r}{t} \right) \leq \mu_{s^*} \leq \varepsilon t \). Then, from Chernoff bound it follows that

\[
Pr[\deg_t(s) \leq \beta \varepsilon t \deg_t(s) | G_s] = Pr[\exp (\beta \varepsilon t \deg_t(s) - \deg_t(s)) \geq 1 | G_s] \leq \exp (\beta \mu_t \varepsilon t \deg_t(s) | G_s) \leq \exp (\beta \mu_t \varepsilon t \deg_t(s) | G_s).
\]

Now we know that

\[
E \left[ \exp \left( -\mu_{s^*} \deg_{s^*}(s) \right) \right] \leq \exp \left( -\mu_{s^*} \delta s \ln \frac{s^*}{s} \right) \Pr \left[ \deg_t(s) > \delta s \ln \frac{s^*}{s} \right] + \Pr \left[ \deg_t(s) \leq \delta s \ln \frac{s^*}{s} \right]
\]

\[
\leq \left( \frac{s^*}{s} \right)^{-\mu_{s^*} \delta s} + O \left( \left( \frac{s^*}{s} \right)^{-\frac{\delta s}{2}} \right)
\]

Let’s assume \( s^* = t^\gamma \) and \( \varepsilon_t = \frac{\lambda}{\gamma \ln \ln t} \). For \( p > 0 \) we may proceed further:

\[
Pr[\deg_t(s) \leq \beta \varepsilon t \deg_t(s) | G_s] \leq \exp \left( \beta \varepsilon t \left(1-\gamma\right) \left( \deg_t(s) + C_2(p, r) \right) \left( \frac{t}{s} \right)^p \right) E \left[ \exp \left( -\mu_{s^*} \deg_{s^*}(s) \right) \right] \leq \exp \left( \beta \varepsilon t \left( \deg_t(s) + C_2(p, r) \right) \left( \frac{t}{s} \right)^p \right) \left( \left( \frac{t}{s} \right)^{-\frac{\lambda \gamma \ln \ln t}{2}} + O \left( \left( \frac{t}{s} \right)^{-\frac{\lambda \gamma \ln \ln t}{2}} \right) \right)
\]
\[ \leq \exp \left( A \frac{\beta}{Cr^\gamma \ln \ln t} \left( \deg_s(s) + C_2(p,r) \frac{t^\delta}{s^p} \right) \left( t^{-\frac{4}{3r^2}} + O \left( t^{-\frac{2}{3r^2}} \right) \right) \right) \]

Let now \( \beta = t^{-p\gamma} \). Then

\[ \Pr[\deg_t(s) \leq t^{-p\gamma} \deg_t(s) | G_s] \]

\[ \leq \exp \left( \frac{A(\deg_s(s) + C_2(p,r))}{Cr^\gamma s^p \ln \ln t} \right) \left( t^{-\frac{4}{3r^2}} + O \left( t^{-\frac{2}{3r^2}} \right) \right) \]