



HAL
open science

Power-Law Degree Distribution in the Connected Component of a Duplication Graph

Philippe Jacquet, Krzysztof Turowski, Wojciech Szpankowski

► **To cite this version:**

Philippe Jacquet, Krzysztof Turowski, Wojciech Szpankowski. Power-Law Degree Distribution in the Connected Component of a Duplication Graph. AofA2020 - 31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, Jun 2020, Klagenfurt, Austria. hal-03123680

HAL Id: hal-03123680

<https://hal.science/hal-03123680>

Submitted on 28 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.


L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Power-Law Degree Distribution in the Connected 2 Component of a Duplication Graph

3 Philippe Jacquet


4 INRIA, Saclay – Île-de-France, France

5 philippe.jacquet@inria.fr

6 Krzysztof Turowski 

7 Theoretical Computer Science Department, Jagiellonian University, Krakow, Poland

8 krzysztof.szymon.turowski@gmail.com

9 Wojciech Szpankowski 

10 Center for Science of Information, Department of Computer Science, Purdue University, West

11 Lafayette, IN, USA

12 spa@cs.purdue.edu

13 — Abstract —

14 We study the partial duplication dynamic graph model, introduced by Bhan et al. in [3] in which a
15 newly arrived node selects randomly an existing node and connects with probability p to its neighbors.
16 Such a dynamic network is widely considered to be a good model for various biological networks
17 such as protein-protein interaction networks. This model is discussed in numerous publications
18 with only a few recent rigorous results, especially for the degree distribution. Recently Jordan [9]
19 proved that for $0 < p < \frac{1}{e}$ the degree distribution of the *connected component* is stationary with
20 *approximately* a power law. In this paper we rigorously prove that the tail is indeed a true power law,
21 that is, we show that the degree of a randomly selected node in the connected component decays
22 like C/k^β where C an explicit constant and $\beta \neq 2$ is a non-trivial solution of $p^{\beta-2} + \beta - 3 = 0$. This
23 holds regardless of the structure of the initial graph, as long as it is connected and has at least
24 two vertices. To establish this finding we apply analytic combinatorics tools, in particular Mellin
25 transform and singularity analysis.

26 **2012 ACM Subject Classification** Mathematics of computing → Random graphs; Theory of com-
27 putation → Random network models

28 **Keywords and phrases** random graphs, pure duplication model, degree distribution, tail exponent,
29 analytic combinatorics

30 **Digital Object Identifier** 10.4230/LIPIcs.AofA.2020.7

31 **Funding** This work was supported by NSF Center for Science of Information (CSoI) Grant CCF-
32 0939370, in addition by NSF Grant CCF-1524312, and National Science Center, Poland, Grant
33 2018/31/B/ST6/01294.



© Philippe Jacquet and Krzysztof Turowski and Wojciech Szpankowski;
licensed under Creative Commons License CC-BY

31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of
Algorithms (AofA 2020).

Editors: Michael Drmota and Clemens Heuberger; Article No. 7; pp. 7:1–7:15



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Recent years have seen a growing interest in dynamic graph models [10]. These models are often claimed to describe well various real-world structures, such as social networks, citation networks and various biological data. For example, protein-protein are widely viewed as driven by an internal evolution mechanism based on duplication and mutation. In this case, new nodes are added to the network as copies of existing nodes together with some random divergence. It has been claimed that graphs generated from these models exhibit many properties characteristic for real-world networks such as power-law degree distribution, the large clustering coefficient, and a large amount of symmetry [4]. However, some of these results turned out not to be correct; in particular, the power-law degree distribution was disproved in [7]. In this paper we focus on the tail distribution of the *connected component* of such networks and show rigorously the existence of a power law improving and making more precise recent result of Jordan [9].

The model analyzed in this paper is known as the *partial (pure) duplication model*, in which a new node selects an existing node and connects to its neighbors with probability p . More precisely, the model is defined formally as follows: let $0 < p \leq 1$ be the only parameter of the model. In discrete steps repeat the following procedure: first, choose a single vertex u uniformly at random. Then, add a new vertex v and for all vertices w such that uw is an edge (i.e., w is a neighbor of u) flip a coin independently at random (heads with probability p , tails with $1 - p$) and add vw edge if and only if we got heads. The partial duplication model was defined by Bhan et al. in [3] and then was further studied in [1, 4, 7, 9, 8].

The case when $p = 1$, also called the *full duplication model*, was analyzed recently in the context of graph compression in [13]. In particular, it was formally proved that the expected logarithm of the number of automorphisms (symmetries) for such graphs on n vertices is asymptotically $\Theta(n \log n)$, which in turn lead us to an asymptotically efficient compression algorithm for such case.

The *partial duplication* case $0 < p < 1$ was given much more attention, however, with very few rigorous results. It was first and foremost analyzed to find the stationary distribution of the degree, that is,

$$f_k = \lim_{n \rightarrow \infty} f_k(n) = \lim_{n \rightarrow \infty} \frac{F_k(n)}{n} = \lim_{n \rightarrow \infty} \Pr[\deg(U_n) = k],$$

where $f_k(n)$ and $F_k(n)$ are, respectively, the average fraction and the average number of vertices of degree k in a graph generated by this model and U_n is a random variable denoting a vertex chosen uniformly at random from a graph on n vertices generated from the partial duplication model. Hermann and Pfaffelhuber in [7] proved that this process $(f_k(n))_{n=n_0}^{\infty}$ converges always to the limit $f_0 = 1$ and $f_k = 0$ for all other k when $p \leq p^* = 0.57\dots$ (that is, p^* being the unique root of $pe^p = 1$), regardless of the initial graph. They have also shown that if $p > p^*$ there exists only a defective distribution of the degrees with $f_0 = c < 1$ for a certain constant c (depending on the initial graph) and $f_k = 0$ for all other k . For the average degree distribution see also [14].

This result, although it refuted the power law behavior of the whole graph claimed by [4, 2], also showed that asymptotically almost all vertices are isolated. This has still left the possibility that it might be the case that a graph generated by the partial duplication model with the isolated vertices removed exhibits such property. Note that by a simple inductive argument it is obvious that if a vertex is isolated at the time of its insertion, then it stays isolated forever, and if it was connected to other vertex, then it remains connected, so if the initial graph is connected, then there can only be one component containing all non-isolated

81 vertices. This was exactly the route pursued by Jordan in [9]. Using probabilistic tools such
 82 as the quasi-stationary distribution of a certain continuous time Markov chain embedding
 83 of the original discrete graph growth process, Jordan was able to prove that for $0 < p < \frac{1}{e}$
 84 there is an *approximate power law* behavior in the pure duplication graphs. More precisely,
 85 let us define for a vertex (denoted by U_n) picked uniformly at random from a graph on n
 86 vertices generated from the duplication model the following conditional probability

$$87 \quad a_k(n) = \Pr[\deg(U_n) = k | \deg(U_n) \neq 0] = \frac{f_k(n)}{\sum_{i=1}^{\infty} f_i(n)} = \frac{f_k(n)}{1 - f_0(n)}. \quad (1)$$

89 Jordan proved that $a_k(n) \rightarrow a_k$ as $n \rightarrow \infty$ as long as the underlying process is positive
 90 recurrent which holds for $p < \frac{1}{e}$ [9]. Moreover, Jordan showed that for $\beta(p) \neq 2$ being the
 91 solution of $p^{\beta-2} + \beta - 3 = 0$ the tail behavior of a_k is approximately a power law in the
 92 sense that it is lighter than any heavier tailed power law (with any index $\beta(p) + \varepsilon$, $\varepsilon > 0$)
 93 and heavier than any lighter tailed power law (with index $\beta(p) - \varepsilon$, $\varepsilon > 0$). This is of vital
 94 interest in this area since $\beta(p) \in (2, 3)$, which is exactly the range of the power law exponents
 95 for various real-world biological graphs, such as protein-protein networks [4].

96 It is worth noting that it partially confirmed the non-rigorous result by Ispolatov et al.
 97 from [8], who claimed that the connected component exhibits a power-law distribution both
 98 for $0 < p < \frac{1}{e}$ (with index $\beta(p)$ as above), and for $\frac{1}{e} \leq p < \frac{1}{2}$ (with index 2). Furthermore,
 99 by the virtue of (1) observe, following [9, 7], that $f_0(n) = 1 - o(1)$ and $f_k(n) = o(1)$ for $k \geq 1$
 100 which begs the question of the asymptotic behavior of $f_k(n)$ for large k and n . Certainly
 101 $f_k(n)$ does *not* grow linearly with n as suggested in some papers (cf. [2]). We conjecture
 102 that $f_k(n) = O(n^{-\alpha(p)} k^{-\beta(p)})$ for some $1 < \alpha(p) < 2$ and $\beta(p) > 2$, but this problem is left
 103 for future research.

104 In this paper we finally establish the precise behavior of the tail of the degree distribution
 105 for pure duplication model for $0 < p < \frac{1}{e}$ completing the work of Jordan [9]. More precisely,
 106 we use tools of analytic combinatorics such as the Mellin transform and singularity analysis
 107 to prove in Theorem 2 that the tail of a node degree in the connect component of the partial
 108 duplication model decays as $C/k^{\beta(p)}$ where C an explicit constant.

109 The paper is organized as follows: in Section 2 we present a formal definition of the model,
 110 introduce the tracked vertex approach, and the quasi-stationary distribution as defined by
 111 Jordan in [9]. In Section 3 we state and establish our main results using Mellin transform
 112 and singularity analysis. In concluding Section 4 we indicate a possible extension of our
 113 findings and point to some further work.

114 **2 The model and Jordan's approach**

115 We follow the standard graph-theoretical notation, e.g., from [5]. We consider only simple
 116 graphs, i.e., graphs without loops or parallel edges.

117 Let us recall first the definition of the pure duplication model. Let $G_{n_0} = (V_{n_0}, E_{n_0})$ be
 118 an initial graph with a set of vertices V_{n_0} and a set of edges E_{n_0} , such that $|V_{n_0}| = n_0 \geq 2$.
 119 Throughout the paper, let us assume that G_{n_0} is fixed and connected. For $n = n_0, n_0 + 1, \dots$
 120 we build $G_{n+1} = (V_{n+1}, E_{n+1})$ from $G_n = (V_n, E_n)$ in the following way:

- 121 1. pick a vertex $u \in V_n$ uniformly at random,
- 122 2. create a new node v_{n+1} and let $V_{n+1} = V_n \cup \{v_{n+1}\}$, $E_{n+1} = E_n$,
- 123 3. for every $w \in V_n$ such that $uw \in E_n$ add edge $v_{n+1}w$ to E_{n+1} independently at random
 124 with probability p .

125 We call the process $\mathcal{G} = (G_n)_{n=n_0}^{\infty}$ the *partial duplication graph*.

7:4 Power-Law Degree Distribution...

126 Jordan in [9] introduced the continuous-time embedding of this process, defined as
 127 following: start at time $t = 0$ with a fixed connected graph $\Gamma_0 = G_{n_0}$ and let $(\Gamma_t)_{t \geq 0}$ be
 128 a continuous time Markov chain on graphs, where each vertex is duplicated independently
 129 at times following a Poisson process of rate 1, with the rules for duplication as in the pure
 130 duplication model.

131 Jordan also defined the so called *vertex tracking approach*: we pick a vertex from Γ_0
 132 uniformly at random and then define the continuous-time process $(V_t)_{t \geq 0}$ in the following
 133 way: at time t we jump to a vertex v if and only if the vertex V_{t-} was duplicated and its
 134 „child” is v . He proved that for any $k \geq 1$ and for another continuous-time process $(U_t)_{t \geq 0}$
 135 being defined as a uniform choice of vertices over Γ_t we have

$$136 \quad \lim_{t \rightarrow \infty} \frac{\Pr[\deg(U_t) = k]}{\Pr[\deg(V_t) = k]} = 1. \quad 137$$

138 Therefore, asymptotically the behavior of a tracked vertex approximates the behavior of a
 139 random vertex in Γ_t when $t \rightarrow \infty$, and therefore in G_n when $n \rightarrow \infty$.

140 The tracked vertex approach allowed Jordan to construct the generator Q of the continuous-
 141 time Markov chain $(\deg(V_t))_{t \geq 0}$, defined over the state space \mathbb{N}_0 , with the following transitions

$$142 \quad q_{j,k} = \binom{j}{k} p^k (1-p)^{j-k} \quad \text{for } 0 \leq k \leq j-1,$$

$$143 \quad q_{j,j} = -jp - (1-p^j),$$

$$144 \quad q_{j,j+1} = jp.$$

146 Then Jordan proceeded to the analysis of the quasi-stationary distribution $(a_k)_{k=1}^\infty$, i.e.,
 147 the left eigenvector of a subset of Q , defined as before. We relate this distribution to the
 148 eigenvalue $-\lambda$ (see [11] for details of this approach) being the solution of the equation
 149 $AQ = -\lambda Q$, where $A = (a_k)_{k=1}^\infty$. This leads us to the following equation:

$$150 \quad \sum_{j=k}^{\infty} a_j \binom{j}{k} p^k (1-p)^{j-k} = -(k-1)pa_{k-1} - (\lambda - kp - 1)a_k \quad (2)$$

151 for $k = 1, 2, 3, \dots$

152 Using (2) and the generating function $A(z) = \sum_{k=0}^{\infty} a_k z^k$ Jordan found the following
 153 differential-functional equation
 154

$$155 \quad A(pz + 1 - p) = (1 - \lambda)A(z) + pz(1 - z)A'(z) + A(1 - p). \quad (3)$$

157 Notice that the above equation implies that $A(0) = 0$. Since it is a sum of limits for
 158 probability distributions, by Fatou’s lemma $|A(z)| \leq 1$ for $|z| \leq 1$. By letting $z \rightarrow 1^-$ in (3)
 159 and assuming finite $A'(1)$ we get $A(1 - p) = \lambda A(1)$.

160 Furthermore with the identity

$$161 \quad A'(z) = \frac{A(pz + 1 - p) - A(1)}{pz(1 - z)} - (1 - \lambda) \frac{A(z) - A(1)}{pz(1 - z)} \quad (4)$$

162 and letting $z \rightarrow 1^-$ Jordan found

$$163 \quad A'(1) = -A'(1) + \frac{1 - \lambda}{p} A'(1),$$

164 namely, if $A'(1)$ is non-zero and finite, then $\lambda = 1 - 2p$. Finally, using the assumptions that
 165 the distribution $(a_k)_{k=0}^\infty$ is non-degenerate (i.e., $\sum_{k=0}^{\infty} a_k = A(1) = 1$) and that the mean
 166

167 degree $A'(1)$ is finite, Jordan found that for $0 < p < \frac{1}{e}$ the quasi-stationary distribution a_k
 168 does not have q -th moment for $p^{q-2} + q - 3 < 0$.

169 In summary Jordan proved the following result.

170 ► **Theorem 1** ([9, Theorem 2.1(3)]). Assume $0 < p < \frac{1}{e}$. Let $\beta(p) > 2$ be the solution of
 171 $p^{\beta-2} + \beta - 3 = 0$. Then the tail behaviour of $(a_k)_{k=0}^\infty$ has a power law of index $\beta(p)$, in the
 172 sense that as $k \rightarrow \infty$,

173
$$\lim_{k \rightarrow \infty} \frac{a_k}{k^q} = 0 \quad \text{for } q < \beta(p),$$

174
$$\lim_{k \rightarrow \infty} \frac{a_k}{k^q} = \infty \quad \text{for } q > \beta(p).$$

 175

176 In the next section we present our refinement of this theorem and provide precise
 177 asymptotics for $(a_k)_{k=0}^\infty$.

178 **3 Main results**

179 In this section we state and prove the main result of our paper that is a refinement of
 180 Theorem 1.

181 ► **Theorem 2.** If $0 < p < \frac{1}{e}$, then the stationary distribution $(a_k)_{k=0}^\infty$ of the pure duplication
 182 model has the following asymptotic tail behavior as $k \rightarrow \infty$:

183
$$\frac{a_k}{k^{\beta(p)}} = \frac{1}{E(1) - E(\infty)} \cdot \frac{p^{-\frac{1}{2}(\beta(p) - \frac{3}{2})^2} \Gamma(\beta(p) - 2)}{D(\beta(p) - 2)(p^{-\beta(p)+2} + \ln(p))\Gamma(-\beta(p) + 1)} \left(1 + O\left(\frac{1}{k}\right)\right) \quad (5)$$

 184

185 where $\beta(p) > 2$ is the non-trivial solution of $p^{\beta-2} + \beta - 3 = 0$, $\Gamma(s)$ is the Euler gamma
 186 function and

187
$$D(s) = \prod_{i=0}^{\infty} (1 + p^{1+i-s}(s - i - 2)), \quad (6)$$

188
$$E(1) - E(\infty) = \frac{1}{2\pi i} \int_{\text{Re}(s)=c} p^{-\frac{1}{2}(s-\frac{1}{2})^2} \frac{\Gamma(s)}{D(s)} ds, \quad \text{for } c \in (0, 1).$$

 189

190 In Figure 1 numerical values of the functions involved in the formula above. It clear that
 191 that all coefficients in (5) are positive for $0 < p < \frac{1}{e}$.

192 The rest of this section is devoted to the proof of our main result. We will accomplish
 193 it by a series of lemmas. The main idea is as follows: we take (3) and apply a series of
 194 substitutions to obtain a functional equation which is in suitable form for applying Mellin
 195 transform. Observe that we cannot apply directly Mellin transform to the functional equation
 196 (5) due to the term $A(pz + 1 - p)$.

197 It is already known from [9] that $A'(1)$ is non-zero and finite, hence $\lambda = 1 - 2p$. First, let
 198 us substitute $z = 1 - v$ and $B(v) = A(1 - v)$ in (3). Thus

199
$$A(1 - pv) = 2pA(1 - v) + pv(1 - v)A'(1 - v) + A(1 - p),$$

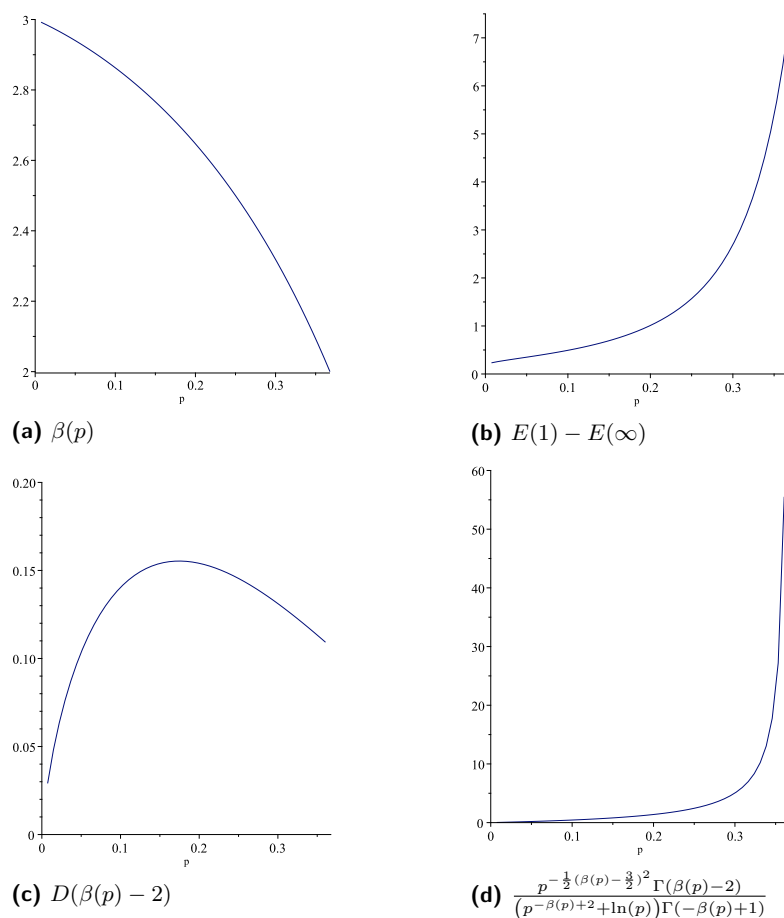
200
$$B(pv) = 2pB(v) - pv(1 - v)B'(v) + A(1 - p).$$

 201

202 Observe now that the functional equation on $B(v)$ is suitable for the Mellin transform. To
 203 ease some computation let $w = \frac{1}{v}$ and $C(w) = B\left(\frac{1}{w}\right)$. Then

204
$$B\left(\frac{p}{w}\right) = 2pB\left(\frac{1}{w}\right) - \frac{p}{w} \left(1 - \frac{1}{w}\right) B'\left(\frac{1}{w}\right) + A(1 - p),$$

7:6 Power-Law Degree Distribution...



■ **Figure 1** Numerical values of different parts of (3) for $0 < p < \frac{1}{e}$.

$$C\left(\frac{w}{p}\right) = 2pC(w) + p(w-1)C'(w) + A(1-p). \quad (7)$$

Therefore, we are essentially looking at the solution of (7) with boundary conditions $C(1) = A(0) = 0$ and $\lim_{w \rightarrow \infty} C(w) = A(1)$ (which is equal to 1, as pointed out in [9]).

Our objective is to find an asymptotic expansion for $C(w)$ when $w \rightarrow \infty$. Notice that it is equivalent to finding the asymptotic expansion of $A(z)$ when $z \rightarrow 1$ by inferior values. For this purpose we will use the Mellin transform which is a powerful tool for extracting accurate asymptotic expansions [12]. Unfortunately we cannot directly apply the Mellin transform over function $C(w)$ since the behavior of $C(w)$ for $w \rightarrow 0$ is yet unknown. To circumvent this problem we search for a similar function $E(w)$ defined by the following functional equation

$$E\left(\frac{w}{p}\right) = 2pE(w) + p(w-1)E'(w) + K \quad (8)$$

for some constant K for which we shall postulate that the Mellin transform

$$E^*(s) = \int_0^\infty w^{s-1} E(w) dw$$

exists in some fundamental strip.

221 To connect $E(w)$ with our function $C(w)$ we notice that it holds necessarily that $C(1) = 0$
 222 which corresponds to the fact that $A(0) = 0$. Clearly, if $E(w)$ is the solution of (8) with
 223 finite values of both $E(1)$ and $E(\infty) = \lim_{w \rightarrow \infty} E(w)$ (which will be shown later to be the
 224 case), then it is also true that

$$225 \quad C(w) = A(1) \frac{E(w) - E(1)}{E(\infty) - E(1)} \quad (9)$$

227 is the solution of (7) with $C(1) = 0$ which also satisfies $\lim_{w \rightarrow \infty} C(w) = A(1) = 1$.

228 Let us now proceed through definition and lemmas. We first define

$$229 \quad E^*(s) = p^{-\frac{1}{2}(s-\frac{1}{2})^2} \frac{\Gamma(s)}{D(s)} \quad (10)$$

230 for $D(s) = \prod_{j=0}^{\infty} (1 + p^{1+j-s}(s-j-2))$ defined already in (6).

231 Now notice that $D(s) = 0$ implies $1 + p^{1+j-s}(s-j-2) = 0$ for some $j \in \mathbb{N}$. This equation
 232 for $0 < p < \frac{1}{e}$ has only two solutions: $s = j + 1$ and $s = j + 1 + s^*$, where s^* is the non-trivial
 233 (i.e. other than $s = 0$) solution of $p^s + s - 1 = 0$.

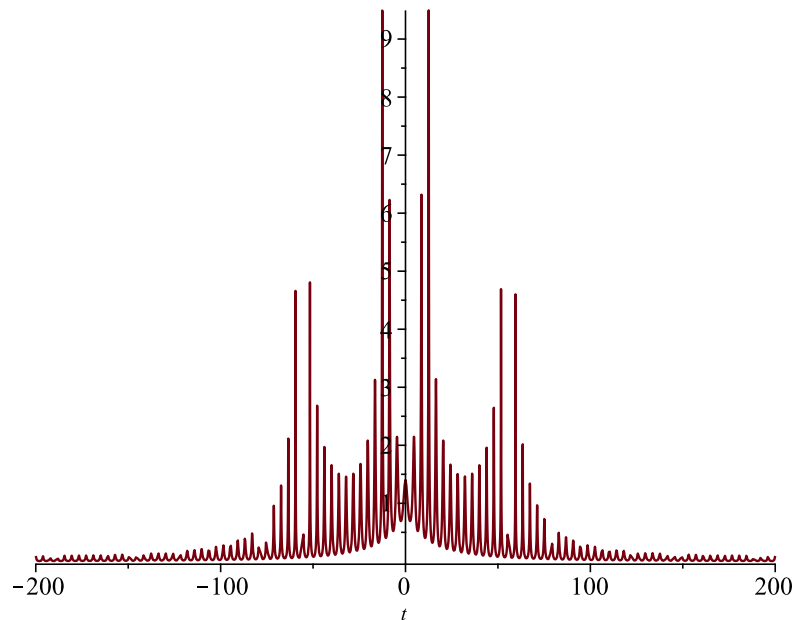
234 Therefore, $E^*(s)$ has only simple, isolated poles of three types:

- 235 ■ for $s = 0, -1, -2, \dots$, introduced by $\Gamma(s)$,
- 236 ■ for $s = 1, 2, 3, \dots$, introduced by $\frac{1}{D(s)}$,
- 237 ■ for $s = s^* + 1, s^* + 2, s^* + 3, \dots$, introduced by $\frac{1}{D(s)}$.

238 Moreover, if we omit these poles, then $D(s)$ converges to a non-zero finite value when
 239 $\text{Re}(s) < 0$ because p^{i-s} exponentially decays. We summarize it in the next lemma.

240 ► **Lemma 3.** For $\text{Re}(s) \in (-1, 0)$ and $0 < p < \frac{1}{e}$ it holds that $\frac{1}{|D(s)|}$ is absolutely convergent.

241 Due to its technical intricacies, the proof of Lemma 3 was moved to the Appendix. In
 242 Figure 2 we present an example plot of values of $\frac{1}{|D(s)|}$.



■ **Figure 2** Numerical values of $\frac{1}{|D(c+it)|}$ for $p = 0.2$ and $c = -0.5$.

7:8 Power-Law Degree Distribution...

243 ► **Lemma 4.** For $0 < p < \frac{1}{e}$ it holds that

$$244 \quad E^*(s) = \frac{p(s-1)}{p^s + ps - 2p} E^*(s-1).$$

245 **Proof.** We have the identity

$$246 \quad \frac{p^{\frac{1}{2}(s-\frac{1}{2})^2}}{\Gamma(s)} E^*(s) = \frac{p^{\frac{1}{2}(s-\frac{3}{2})^2}}{\Gamma(s-1)} E^*(s-1) \frac{1}{1+p^{1-s}(s-2)}.$$

247 Thus

$$248 \quad E^*(s) = \frac{p^{-\frac{1}{2}(s-\frac{1}{2})^2 + \frac{1}{2}(s-\frac{3}{2})^2}}{1+p^{1-s}(s-2)} \frac{\Gamma(s)}{\Gamma(s-1)} E^*(s-1) = \frac{p^{1-s}}{1+p^{1-s}(s-2)} (s-1) E^*(s-1)$$

250 since $\frac{\Gamma(s)}{\Gamma(s-1)} = s-1$. Multiplying by numerator and denominator by p^s completes the
251 proof. ◀

252 We now define for any given $c \in (-1, 0)$

$$253 \quad E(w) = \frac{1}{2\pi i} \int_{\operatorname{Re}(s)=c} E^*(s) w^{-s} ds = \frac{1}{2\pi i} \int_{\operatorname{Re}(s)=c} p^{-\frac{1}{2}(s-\frac{1}{2})^2} \frac{\Gamma(s)}{D(s)} w^{-s} ds. \quad (11)$$

254 Notice that this integral converges for any complex value of w with $\operatorname{Im}(w) \rightarrow \pm\infty$ because
255 from Lemma 3 it follows that $\frac{1}{|D(s)|}$ is bounded by a constant and $\Gamma(s)p^{-\frac{1}{2}(s-\frac{1}{2})^2}$ decays
256 faster than any polynomial. Furthermore the value of $E(w)$ does not depend on the value of
257 quantity c thanks to Cauchy's theorem.

258 ► **Lemma 5.** The function $E(w)$ has function $E^*(s)$ as its Mellin transform with its funda-
259 mental strip being $\{s : \operatorname{Re}(s) \in (-1, 0)\}$.

260 **Proof.** We have

$$261 \quad |E(w)| \leq \frac{|w|^{-c}}{2\pi} \int_{-\infty}^{+\infty} |E^*(c+it)| \exp(\arg(w)t) dt.$$

262 Now, it is easy to spot that $E(c+it) = O\left(\exp\left(-\frac{t^2}{2}\right)\right)$ since $\ln(p) < -1$, thus the integral
263 $\int_{-\infty}^{+\infty} |E^*(c+it)| \exp(\arg(w)t) dt$ absolutely converges and it follows that $E(w) = O(w^{-c})$.
264 Since it is true for any values of $c \in (-1, 0)$ when $w \rightarrow 0$ and $w \rightarrow \infty$, then the Mellin
265 transforms of function $E(w)$ exists with the fundamental strip $\{s : \operatorname{Re}(s) \in (-1, 0)\}$.

266 Furthermore, its Mellin transform is $E^*(s)$ because (11) is exactly the inverse Mellin
267 transform formula. ◀

268 ► **Lemma 6.** There exists a value K independent of w such that

$$269 \quad R(w) = -\operatorname{Res}[E^*(s-1)p(s-1)w^{-s}, s=0] = -K.$$

271 **Proof.** The expression

$$272 \quad R(w) = E\left(\frac{w}{p}\right) - 2pE(w) - p(w-1)E'(w)$$

273 can be also expressed via an integral as

$$275 \quad R(w) = \frac{1}{2\pi i} \int_{\operatorname{Re}(s)=c} E^*(s) (p^s w^{-s} - 2p w^{-s} + sp w^{-s} - sp w^{-s-1}) ds$$

276 which can be rewritten as follows

$$\begin{aligned}
 277 \quad R(w) &= \frac{1}{2\pi i} \int_{\operatorname{Re}(s)=c} E^*(s) (p^s - 2p + ps) w^{-s} ds \\
 278 \quad &\quad - \frac{1}{2\pi i} \int_{\operatorname{Re}(s)=c+1} E^*(s-1)p(s-1)w^{-s} ds \\
 279 \quad &= \frac{1}{2\pi i} \int_{\operatorname{Re}(s)=c} ((p^s + ps - 2p) E^*(s) - p(s-1)E^*(s-1)) w^{-s} ds \\
 280 \quad &\quad - \operatorname{Res}[p(s-1)E^*(s-1), s=0] \\
 281
 \end{aligned}$$

282 since

$$\int_{\operatorname{Re}(s)=c+1} p(s-1)E^*(s-1)w^{-s} ds - \int_{\operatorname{Re}(s)=c} p(s-1)E^*(s-1)w^{-s} ds$$

285 define a contour path which encircles a simple pole at $s = 0$ in the counter-clockwise (i.e.,
286 positive) direction.

287 Furthermore from Lemma 3 it follows that

$$288 \quad (p^s + ps - 2p)E^*(s) - p(s-1)E^*(s-1) = 0,$$

289 therefore the integral vanishes and finally $R(w) = -\operatorname{Res}[p(s-1)E^*(s-1), s=0] = -K$ for
290 some constant K independent of w . ◀

291 ▶ **Lemma 7.** *It holds that*

$$292 \quad K = -\frac{p^{-\frac{1}{8}}(1-2p)}{D(0)}, \quad E(\infty) = \frac{p^{-\frac{1}{8}}}{D(0)}.$$

294 Furthermore,

$$295 \quad E(\infty) - E(1) = -\frac{1}{2\pi i} \int_{\operatorname{Re}(s)=c} E^*(s) ds, \quad \text{for } c \in (0, 1). \quad (12)$$

297 **Proof.** From Lemma 6 we have

$$298 \quad K = \operatorname{Res}[p(s-1)E^*(s-1), s=0] = \frac{p^{-\frac{1}{8}}}{D(-1)}.$$

300 Moreover, from the definition $D(0) = (1-2p)D(-1)$, which establishes the first identity.

301 To find an expression for $E(\infty)$ is a little more delicate. Indeed from (11) we find

$$302 \quad E(w) = -\operatorname{Res}[E^*(s)w^{-s}, s=0] + \frac{1}{2\pi i} \int_{\operatorname{Re}(s)=c'} E^*(s)w^{-s} ds$$

303 by assuming the contour path is moved right to origin for some $c' \in (0, 1)$. It turns out that
304 0 is the simple pole encountered in the move, as $D(s) \neq 0$ for all other s with $\operatorname{Re}(s) \in (0, 1)$.

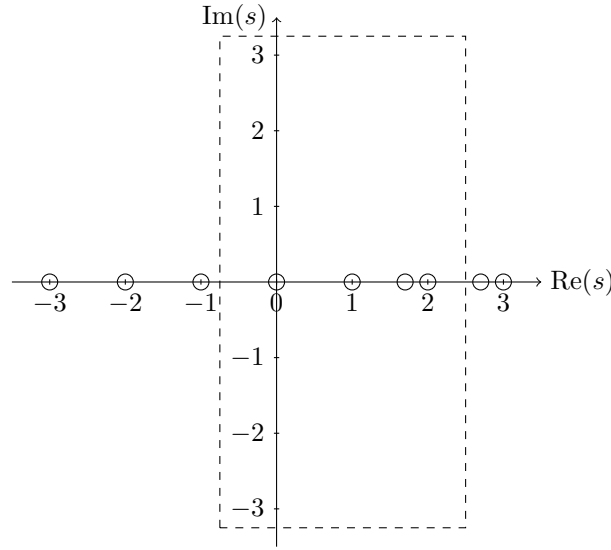
305 Furthermore, the integral on $\operatorname{Re}(s) = c'$ is in $O(w^{-c'})$ as $w \rightarrow \infty$, which allows to conclude
306 that $E(w) = -\operatorname{Res}[E^*(s)w^{-s}, s=0] + O(w^{-c'})$ with $c' \in (0, 1)$, thus

$$307 \quad E(\infty) = \lim_{w \rightarrow \infty} E(w) = -\lim_{w \rightarrow \infty} \operatorname{Res}[E^*(s)w^{-s}, s=0] = -\operatorname{Res}[E^*(s), s=0] = -\frac{p^{-\frac{1}{8}}}{D(0)}.$$

308 Finally,

$$309 \quad E(\infty) - E(1) = -\operatorname{Res}[E(s), s=0] - \frac{1}{2\pi i} \int_{\operatorname{Re}(s)=c} E^*(s) ds = -\frac{1}{2\pi i} \int_{\operatorname{Re}(s)=c'} E^*(s) ds$$

7:10 Power-Law Degree Distribution...



■ **Figure 3** Example integration area for $E^*(s)$ and $E(w)$ with $s^* = 0.7$ and $M = 2.5$.

310

311 for, respectively, $c \in (-1, 0)$ and $c' \in (0, 1)$ since

$$312 \quad \frac{1}{2\pi i} \int_{\text{Re}(s)=c'} E^*(s) ds - \frac{1}{2\pi i} \int_{\text{Re}(s)=c} E^*(s) ds = \text{Res}[E(s), s = 0].$$

313

314 This completes the proof. ◀

315 Note that $D(0) > 0$ since every element in the product is positive for $0 < p < \frac{1}{e}$. Therefore
 316 $K > 0$ and $E(\infty) < 0$.

317 Finally we proceed with the proof of the main theorem.

318 **Proof of Theorem 2.** Recall the observation that $E^*(s)$ has poles for $s \in \{1, 2, \dots\} \cup \{s^* +$
 319 $1, s^* + 2, \dots\} \cup \{0, -1, -2, \dots\}$, for s^* – the non-zero solution of $p^s + s - 1 = 0$. Note that if
 320 $0 < p < \frac{1}{e}$, then $s^* > 0$.

321 Therefore, for any $c \in (-1, 0)$ and a rectangle as presented in Figure 3, we are in position
 322 to write

$$323 \quad C(w) = \frac{1}{E(\infty) - E(1)} \frac{1}{2\pi i} \int_{\text{Re}(s)=c} E^*(s) w^{-s} ds - \frac{E(1)}{E(\infty) - E(1)}$$

$$324 \quad = -\frac{1}{E(\infty) - E(1)} (E(1) + \text{Res}[E^*(s), s = 0] + \text{Res}[E^*(s)w^{-s}, s = 1])$$

$$325 \quad - \frac{1}{E(\infty) - E(1)} (\text{Res}[E^*(s)w^{-s}, s = 2] + \text{Res}[E^*(s)w^{-s}, s = s^* + 1])$$

$$326 \quad + \frac{1}{E(\infty) - E(1)} \frac{1}{2\pi i} \int_{\text{Re}(s)=M} E^*(s) w^{-s} ds \tag{13}$$

327

328 for any number $M \in (2, 2 + s^*)$.

329 The quantity

$$330 \quad \frac{1}{2\pi i} \int_{\text{Re}(s)=M} E^*(s) w^{-s} ds = O(w^{-M})$$

331

332 since $w^{-s} = w^{-M}w^{-\text{Im}(s)}$ and the integral in $E^*(s)w^{-\text{Im}(s)}$ absolutely converge. Again this
 333 holds by a similar argument that was used in Lemma 3: $p^{-\frac{1}{2}(s-\frac{1}{2})^2}$ decays exponentially
 334 faster than $\frac{\Gamma(s)}{D(s)}w^{\text{Im}(s)}$ for complex s .

335

By virtue of the residue theorem

$$\begin{aligned}
 336 \quad C(w) &= -\frac{1}{E(\infty) - E(1)} (E(1) + \text{Res}[E^*(s), s = 0] + \text{Res}[E^*(s), s = 1]w^{-1}) \\
 337 \quad &\quad - \frac{1}{E(\infty) - E(1)} (\text{Res}[E^*(s), s = 2]w^{-2} + \text{Res}[E^*(s), s = s^* + 1]w^{-1-s^*}) \\
 338 \quad &\quad + O(w^{-M}). \tag{14}
 \end{aligned}$$

339

340 This formula gives us an asymptotic expansion of $C(w)$ up to order w^{-M} where $M \in (2, 2+s^*)$.

341

342 In fact, for more precise computations it is possible an expansion to any desired value M ,
 343 just by including all the residues of the poles in k ($k \in \mathbb{N}$) and $k + s^*$ ($k \in \mathbb{N}_+$) which are
 344 smaller than M as for $0 < p < \frac{1}{e}$ all poles are simple.

344

Next, there are computed the first residues, e.g.,

$$\begin{aligned}
 345 \quad \text{Res}[E^*(s)w^{-s}, s = 0] &= \left[p^{-\frac{1}{2}(s-\frac{1}{2})^2} \frac{w^{-s}}{D(s)} \right]_{s=0} = \frac{p^{-\frac{1}{8}}}{D(0)} = -E(\infty), \\
 346 \quad \text{Res}[E^*(s)w^{-s}, s = 1] &= \left[\frac{p^{-\frac{1}{2}(s-\frac{1}{2})^2} \Gamma(s)}{p^{1-s} - (s-2)p^{1-s} \ln(p)} \frac{w^{-s}}{D(s-1)} \right]_{s=1} \\
 347 \quad &= \frac{p^{-\frac{1}{8}} w^{-1}}{1 + \ln(p) D(0)}, \\
 348 \quad \text{Res}[E^*(s)w^{-s}, s = s^* + 1] &= \left[\frac{p^{-\frac{1}{2}(s-\frac{1}{2})^2} \Gamma(s)}{p^{1-s} - (s-2)p^{1-s} \ln(p)} \frac{w^{-s}}{D(s-1)} \right]_{s=s^*+1} \\
 349 \quad &= \frac{p^{-\frac{1}{2}(s^*+\frac{1}{2})^2} \Gamma(s^*)}{p^{-s^*} - (s^*-1)p^{-s^*} \ln(p)} \frac{w^{-s^*-1}}{D(s^*)} \\
 350 \quad &= \frac{p^{-\frac{1}{2}(s^*+\frac{1}{2})^2} \Gamma(s^*) w^{-s^*-1}}{p^{-s^*} + \ln(p) D(s^*)}.
 \end{aligned}$$

350

351

352 Observe that in the formulas above both 1 and $s^* + 1$ are not the zeros of $p^{1-s} - (s-2)p^{1-s} \ln(p)$,
 353 so all the presented expressions have finite value.

354

355 Now it is the moment to use the classic Flajolet-Odlyzko transfer theorem [6] to (9) and
 (14) and obtain

$$\begin{aligned}
 356 \quad A(z) &= 1 - \frac{1}{E(\infty) - E(1)} \frac{p^{-\frac{1}{8}}}{1 + \ln(p) D(0)} \frac{1-z}{D(0)} \\
 357 \quad &\quad - \frac{1}{E(\infty) - E(1)} \frac{p^{-\frac{1}{2}(s^*+\frac{1}{2})^2} \Gamma(s^*)}{p^{-s^*} + \ln(p) D(s^*)} \frac{(1-z)^{1+s^*}}{D(s^*)} \\
 358 \quad &\quad - \frac{1}{E(\infty) - E(1)} \text{Res}[E^*(s), s = 2](1-z)^2 \\
 359 \quad &\quad - \frac{1}{E(\infty) - E(1)} \text{Res}[E^*(s), s = s^* + 2](1-z)^{s^*+2} + o((1-z)^{2+s^*}).
 \end{aligned}$$

360

361 Finally, $(1-z)^\alpha$ for $\alpha \in \mathbb{N}$ is a polynomial and does not contribute to the asymptotics.
 362 And for $\alpha \in \mathbb{R}_+ \setminus \mathbb{N}$ [6] it holds that

$$363 \quad [z^k](1-z)^\alpha = \frac{k^{-\alpha-1}}{\Gamma(-\alpha)} \left(1 + O\left(\frac{1}{k}\right) \right),$$

$$[z^k]o(1-z)^\alpha = o(k^{-\alpha-1}).$$

This leads to the final result, which holds for large k :

$$a_k = [z^k]A(z) = -\frac{1}{E(\infty) - E(1)} \frac{p^{-\frac{1}{2}(s^* + \frac{1}{2})^2} \Gamma(s^*)}{(p^{-s^*} + \ln(p)) \Gamma(-s^* - 1)} \frac{1}{D(s^*)} k^{-s^*-2} \left(1 + O\left(\frac{1}{k}\right)\right).$$

Note that since s^* is the non-trivial real solution of $p^s + s - 1 = 0$, equivalently the exponent may be written as $\beta(p) = s^* + 2$ – the non-trivial (i.e., other than 2) real solution of the equation $p^{\beta-2} + \beta - 3 = 0$.

Putting all the results together we obtain (5) of Theorem 2. Now it is sufficient to confirm that if $0 < p < \frac{1}{e}$, then the tail exponent $\beta(p) > 2$, which means that $A'(1)$ is indeed finite. This proves Theorem 2. ◀

4 Discussion

We proved rigorously the power-law behavior for asymptotic degree distribution of the connected component of the duplication graph $0 < p < \frac{1}{e}$. There remains therefore an open question whether the similar results may be obtained for $p \geq \frac{1}{e}$.

On the one hand, recall the non-rigorous claim in [8] that for $\frac{1}{e} \leq p < \frac{1}{2}$ the index of the power law is equal to 2. Interestingly, $\beta = 2$ is the largest solution of $p^{\beta-2} + \beta - 3 = 0$ for $p \geq \frac{1}{e}$.

On the other hand, Jordan [9, Proposition 3.7] has shown that the dual Markov chain with respect to the eigenvalue $\lambda = 1 - 2p$ is transient for all $p > \frac{1}{e}$ – which suggests that the eventual proof should rely on other value of λ . This problem is left for future research.

References

- 1 Gürkan Bebek, Petra Berenbrink, Colin Cooper, Tom Friedetzky, Joseph Nadeau, and Süleyman Cenk Sahinalp. The degree distribution of the generalized duplication model. *Theoretical Computer Science*, 369(1-3):239–249, 2006.
- 2 Gürkan Bebek, Petra Berenbrink, Colin Cooper, Tom Friedetzky, Joseph Nadeau, and Süleyman Cenk Sahinalp. The degree distribution of the generalized duplication model. *Theoretical Computer Science*, 369(1-3):239–249, 2006.
- 3 Ashish Bhan, David Galas, and T. Gregory Dewey. A duplication growth model of gene expression networks. *Bioinformatics*, 18(11):1486–1493, 2002.
- 4 Fan Chung, Linyuan Lu, T. Gregory Dewey, and David Galas. Duplication models for biological networks. *Journal of Computational Biology*, 10(5):677–687, 2003.
- 5 Reinhard Diestel. *Graph Theory*. Springer, 2005.
- 6 Philippe Flajolet and Andrew Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3(2):216–240, 1990.
- 7 Felix Hermann and Peter Pfaffelhuber. Large-scale behavior of the partial duplication random graph. *ALEA: Latin American Journal of Probability and Mathematical Statistics*, 13:687–710, 2016.
- 8 I. Ispolatov, P. L. Krapivsky, and A. Yuryev. Duplication-divergence model of protein interaction network. *Phys. Rev. E*, 71:061911, 2005. doi:10.1103/PhysRevE.71.061911.
- 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 Mark Newman. *Networks: An Introduction*. Oxford University Press, 2010.
- 11 PK Pollett. Reversibility, invariance and μ -invariance. *Advances in applied probability*, 20(3):600–621, 1988.

- 410 12 Wojciech Szpankowski. *Average Case Analysis of Algorithms on Sequences*. John Wiley &
411 Sons, New York, 2001.
- 412 13 Krzysztof Turowski, Abram Magner, and Wojciech Szpankowski. Compression of Dynamic
413 Graphs Generated by a Duplication Model. In *56th Annual Allerton Conference on Commu-
414 nication, Control, and Computing*, pages 1089–1096, 2018.
- 415 14 Krzysztof Turowski and Wojciech Szpankowski. Towards degree distribution of duplication
416 graph models, 2019.

417 **A Proof of Lemma 3**

418 We now proceed to the proof of Lemma 3. First, let us introduce $f(s) = p^s + ps - 2p$, so that

$$419 \quad D(s) = \prod_{i=0}^{\infty} f(s-i)p^{-(s-i)}.$$

420
421 Observe that $f(s)$ has only two roots, given by Lambert function W , which is the inverse
422 of function xe^x : $W^{-1}(x) = xe^x$. There are only two roots for real numbers which corresponds
423 to two branches W_0 and W_{-1} of the function W . Therefore, any chosen $c < 0$ is smaller than
424 the roots of $f(s)$ and the distance between c and any root is at least 1.

425 **► Lemma 8.** For all $0 < \varepsilon < 1$ and $c < 0$ it holds that $\min_{\operatorname{Re}(s)=c} |f(s)| \geq \Theta(p^{(1-\varepsilon)(c-1)}) > 0$.

426 **Proof.** We have $f'(s) = p^s \ln(p) + p$ and $f''(s) = p^s \ln^2(p)$.

427 Let us consider a complex disk of radius $R = p^{-\varepsilon(c-1)}$ ($R < 1$) centered on s . For
428 $\theta \in (0, 2\pi)$ by virtue of Taylor-Young theorem we have:

$$429 \quad f(s + Re^{i\theta}) = f(s) + f'(s)e^{i\theta}R + \int_0^R f''(s + \rho e^{i\theta})e^{2i\theta} \rho \, d\rho.$$

431 Now observe that

$$432 \quad \left| \int_0^R f''(s + \rho e^{i\theta})e^{2i\theta} \rho \, d\rho \right| = \left| p^s \ln^2(p) e^{2i\theta} \int_0^R p^{\rho \exp(i\theta)} \rho \, d\rho \right|$$

$$433 \quad = \left| p^s \left(e^{R \exp(i\theta) \ln(p)} [R \exp(i\theta) \ln(p) - 1] + 1 \right) \right|$$

$$434 \quad = O(|p^s R^2 e^{2i\theta}|) = O(p^c R^2),$$

436 where the last line follows from the fact that asymptotically $e^x(x-1) + 1 = O(x^2)$ for $x \rightarrow 0$.

437 When θ varies the quantity $f'(s)e^{i\theta}R$ describes a circle of radius $|f'(s)|R = (-p^c \ln(p) +$
438 $O(p))R$ around $f(s)$. The error term bound implies that each point of $f(s + Re^{i\theta})$ is at
439 distance $O(p^c R^2)$ of this circle. Thus the image by f of the disk with center s and radius R
440 contains the disk of center $f(s)$ and radius

$$441 \quad R|f'(s)| - O(R^2 p^c) = -p^{-\varepsilon(c-1)} p^c \ln(p) - O(p^{1-\varepsilon(c-1)}) - O(p^{-2\varepsilon(c-1)} p^c)$$

$$442 \quad = p^{(1-\varepsilon)(c-1)} \left(-p \ln(p) - O(p^{1-c}) - O(p^{-\varepsilon(c-1)}) \right) = \Theta(p^{(1-\varepsilon)(c-1)}).$$

444 The point $s = 0$ cannot be in this disk, otherwise the function $f(s)$ would have other
445 roots than the expected ones, thus necessarily $|f(s)| \geq \Theta(p^{(1-\varepsilon)(c-1)})$. ◀

446 Let now $g(s) = p^{-s} f(s)$ so that

$$447 \quad D(s) = \prod_{i=0}^{\infty} g(s-i).$$

448 **► Lemma 9.** For t real and $c < 0$, the following inequality holds

$$|g(c+it)| \geq |1 - p^{1-c}(2-c) - p^{1-c}|t||.$$

449 **Proof.** We have

$$450 \quad |g(c+it)| = |p^{-c} f(c+it)| = |p^{it} + p^{1-c}(c-2) + p^{1-c}it|$$

$$451 \quad \geq ||p^{it}| - |p^{1-c}(c-2)| - |p^{1-c}it||.$$

453 But now observe that $|p^{it}| = 1$, which completes the proof. ◀

454 ► **Lemma 10.** For $c \in (-1, 0)$ and for all real number t outside any neighborhood of 0, for
 455 all $\varepsilon > 0$ it is true that $\frac{1}{D(c+it)} = O(\exp(-(\log_p^2 |t|/2 + O(\log |t|)))$.

Proof. From Lemmas 8 and 9, it follows that:

$$|D(c + it)| \geq \prod_{k \geq 0} \max\{Bp^{-\varepsilon(1-c)}, |1 - (|t| + 2 + k - c)p^{k+1-c}|\}$$

456 For a given real number t , we denote $k(t)$ the largest integer k such that $(|t| + 2 + k - c)p^{k+1-c} >$
 457 1, and we split the product at $k = k(t)$:

$$\begin{aligned} 458 \quad |D(c + it)| &\geq \prod_{k < k(t)} ((|t| + 2 + k - c)p^{k+1-c} - 1) \\ 459 \quad &\quad B'|t|^{-\varepsilon} \prod_{k > k(t)} (1 - (|t| + 2 + k - c)p^{k+1-c}) \\ 460 \quad &\geq \left(\prod_{k < k(t)} \left(p^{k-k(t)} \left(1 - (k(t) - k)p^{k(t)+1-c} \right) - 1 \right) \right) \\ 461 \quad &\quad B'|t|^{-\varepsilon} \prod_{k > k(t)} \left(1 - p^{k-k(t)} \left(1 - (k(t) - k)p^{k(t)-c} \right) \right) \\ 462 \end{aligned}$$

Now

$$\prod_{k > k(t)} \left(1 - p^{k-k(t)} \left(1 - (k(t) - k)p^{k(t)-c} \right) \right) \geq \prod_{k > 0} (1 - p^k).$$

Furthermore $\prod_{k < k(t)} p^{k-k(t)} \geq p^{k(t)(k(t)-1)/2}$, thus

$$\prod_{k < k(t)} \left(p^{k-k(t)} \left(1 - (k(t) - k)p^{k(t)+1-c} - 1 \right) \right) \geq p^{k(t)(k(t)-1)/2} \prod_{k > 0} (1 - p^k).$$

Finally, $p^{-k(t)} = |t|p^{-c}$ and therefore

$$|D(c + it)| \geq p^{k(t)(k(t)-1)/2} B'|t|^{-\varepsilon} \prod_{k > 0} (1 - p^k)^2 = B'' \frac{|t|^{-\varepsilon}}{(|t|p^{-c})^{(k(t)-1)/2}}.$$

463 We conclude, since $k(t) = c - \log_p |t|$. ◀

464 Notice that $D(c + it)$ tends to infinity when $|t| \rightarrow \infty$. To conclude the proof of Lemma 3
 465 it is sufficient to observe that the function $1/D(s)$ for s is any compact set containing a
 466 neighborhood of $\text{Re}(s)$ and away from the roots of $f(s)$ is naturally bounded by dominated
 467 convergence of the product.