



*Citation for published version:*

Betken, C, Döring, H & Ortgiese, M 2019, 'Fluctuations in a general preferential attachment model via Stein's method', *Random Structures and Algorithms*, vol. 55, no. 4, pp. 808-830. <https://doi.org/10.1002/rsa.20852>

*DOI:*

[10.1002/rsa.20852](https://doi.org/10.1002/rsa.20852)

*Publication date:*

2019

*Document Version*

Peer reviewed version

[Link to publication](#)

This is the peer reviewed version of the following article: Betken, C, Döring, H, Ortgiese, M. Fluctuations in a general preferential attachment model via Stein's method. *Random Struct Alg.* 2019; 1– 23, which has been published in final form at <https://doi.org/10.1002/rsa.20852>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

**University of Bath**

**Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Fluctuations in a general preferential attachment model via Stein's method

CARINA BETKEN<sup>1</sup>, HANNA DÖRING<sup>1</sup> and MARCEL ORTGIESE<sup>2</sup>

## Abstract

We consider a class of dynamic random graphs known as preferential attachment models, where the probability that a new vertex connects to an older vertex is proportional to a sublinear function of the indegree of the older vertex at that time. It is well known that the distribution of a uniformly chosen vertex converges to a limiting distribution. Depending on the parameters, the tail of the limiting distribution may behave like a power law or a stretched exponential. Using Stein's method we provide rates of convergence to zero of the total variation distance between the finite distribution and its limit. Our proof uses the fact that the limiting distribution is the stationary distribution of a Markov chain together with the generator method of Barbour.

*2010 Mathematics Subject Classification:* Primary 05C80, Secondary 60B12.

*Keywords:* random graphs; preferential attachment; Stein's method; coupling; rates of convergence

## 1 Introduction and main results

### 1.1 Introduction

Preferential attachment models, as popularized by Barabási and Albert [BA99], rely only on a simple mechanism to explain the occurrence of power-law degree distributions often observed in real-world networks: the network is modelled as a growing sequence of random graphs, where the probability of connecting a new vertex to an old vertex is proportional to its degree. Here we study a general preferential attachment model, based on the model introduced by Dereich and Mörters [DM09], where the connection probabilities are given by a general (sub-linear) function of the old degree. This framework is sufficiently general to lead to typical degree distributions that can be power-laws, but also stretched exponential distributions. Our main result gives rates of convergence for the total variation distance between the indegree of a uniformly chosen vertex and the limiting degree distribution.

---

<sup>1</sup>Universität Osnabrück, Institut für Mathematik, D-49076 Osnabrück, Germany, [carina.betken@uni-osnabrueck.de](mailto:carina.betken@uni-osnabrueck.de), [hanna.doering@uni-osnabrueck.de](mailto:hanna.doering@uni-osnabrueck.de).

<sup>2</sup>Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom, [m.ortgiese@bath.ac.uk](mailto:m.ortgiese@bath.ac.uk).

For the model introduced in [DM09] we start with  $\mathcal{G}_1$  being one vertex with no edges. In each step we add one vertex. The graph  $\mathcal{G}_n$  consists of  $n$  vertices, but loops or multiple edges are not allowed. The new vertex connects to each of the former vertices  $j$  independently with probability

$$\mathbb{P}(n+1 \text{ connects to } j \mid \mathcal{G}_n) = \frac{f(\deg_n^-(j))}{n},$$

where  $\deg_n^-(j)$  is the indegree of vertex  $j$  at time  $n$  and the function  $f : \mathbb{N}_0 \rightarrow (0, \infty)$  is such that  $f(n) \leq n+1$ . If  $f$  is a linear function, we refer to the model as linear preferential attachment. Here the connections to old vertices are sampled independently, so that the outdegree is random, however our framework also allows for models with fixed outdegree (see below) that are closer to the original mathematical formulation of the Barabási-Albert model due to [BRST01].

Here, we analyse the rates of convergence in total variation between the indegree of a random uniformly chosen vertex, resp. the outdegree of vertex  $n$ , and the limiting distribution with Stein's method. Let  $W_n$  denote the indegree of a uniformly chosen vertex at time  $n$  in the model described above. Then, our main theorem shows under fairly weak conditions on  $f$  (essentially that eventually  $f(k) \leq k$ ): if  $W$  has the asymptotic degree distribution identified in [DM09] (see also (3) below), then for all  $n \geq 2$

$$d_{\text{TV}}(W_n, W) \leq C \frac{\log(n)}{n}. \quad (1)$$

We also give a weaker result if  $f$  is of the form  $f(k) = k + \gamma$  for some  $\gamma \in (0, 1)$ . Moreover, for the model with random outdegree, [DM09], we can also obtain rate of convergence results for the outdegree when compared with a Poisson distribution.

For linear  $f$ , a result analogous to ours was shown in [For09], [PRR13a] and [Ros13]. In the thesis [For09] various versions of the preferential attachment model (including random and fixed outdegree) were studied using couplings and Stein's method. However, in her work only the former was used successfully to derive rates of convergence, which in the best case are of the form (1). For the linear model with fixed outdegree and where connections are made with probability proportional to the degree plus a constant  $\delta$ , there is an alternative proof of (1) first in [PRR13a] (for  $\delta = 0$ ) and then in [Ros13]. These proofs use Stein's method, but crucially rely on the fact that the limiting distribution can be represented as a mixture of geometric (resp. negative binomially distributed for  $\delta \neq 0$ ) random variables.

The first mathematical papers on preferential attachment models showed that the proportion of vertices with a particular degree is close to a power law including error bounds, see [BRST01]. The approach in [BL12] also works for large degrees and [Hof17] provides an excellent overview. These results are often proved by first showing concentration of the empirical degree distribution and then showing

that the expected degree sequence converges. However, in these cases the bounds on the expected degree sequence are not strong enough to give (1), see also the discussion in [Ros13].

Other questions that have been studied include error bounds for the degree of a fixed vertex. Peköz, Röllin and Ross [PRR13a] apply Stein’s method successfully to prove a rate of convergence for the Kolmogorov distance between the indegree distribution of any fixed vertex and a power law distribution by comparing it to a mixed negative binomial distribution, whereas in [PRR13b] they provide rates of convergence for the total variation distance between a random vertex in a uniform attachment graph and the geometric distribution. In [PRR17] the authors prove a rate of convergence in the multidimensional case for the joint degree distribution. These articles consider models where every vertex has fixed outdegree.

Our results are based on a new application of Stein’s method. Compared to previous results our methods apply to a generalisation of preferential attachment models having very different limit laws, including power law, but also stretched exponential, tail behaviour. This is remarkable, considering that Stein’s method (at least in a first step) requires the characterization of the limiting distribution in terms of the Stein operator. Secondly, our method is direct in that it gives a Stein operator for the limiting distribution (unlike in [Ros13] where it is exploited that the limiting distribution can be written as the mixture of negative binomial distributions). Our proof relies on the fact that the limiting distribution is the invariant distribution of a continuous-time Markov chain and therefore it allows to use Barbour’s generator method [Bar88] to take care of the first part of Stein’s method. Finally, our method is robust in that it only depends on the marginal distribution of a new vertex being connected to a particular old vertex, and so allows to treat models with fixed and random outdegree at the same time.

Throughout we will use the following notation:  $\mathbb{N} = \{1, 2, \dots\}$  denotes the natural numbers,  $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$  and  $[n] := \{1, \dots, n\}$ . For any two probability measures  $\mu, \nu$  on  $(E, \mathcal{F})$  we will denote their total variation distance by

$$d_{\text{TV}}(\mu, \nu) = \sup\{|\mu(A) - \nu(A)| : A \in \mathcal{F}\}.$$

Moreover, for any random variables  $X, Y$  taking values in  $E$ , we write  $d_{\text{TV}}(X, Y) := d_{\text{TV}}(\mathcal{L}(X), \mathcal{L}(Y))$ .

## 1.2 Main results

Our methods do not rely on the fine details of the model under consideration, so that we state our assumptions first in all generality and then highlight some of the models that fit in this class.

**Assumptions (A).** Fix  $d_0 \in \mathbb{N}_0$  and let  $f : \mathbb{N}_0 \rightarrow (0, \infty)$  such that  $f(n) \leq \max\{n + 1 - d_0, 1\}$ . We assume that  $(\mathcal{G}_n)_{n \geq 1}$  is a sequence of directed random graphs with vertex set  $[n] := \{1, \dots, n\}$ . The initial graph  $\mathcal{G}_1$  consists of vertex 1

and  $d_0$  (directed) self-edges. For any  $n \geq 1$ , at time  $n + 1$  we add vertex  $n + 1$  to the vertex set and insert at most one directed edge from  $n + 1$  to  $j \in [n]$  such that

$$\mathbb{P}(n + 1 \text{ connects to } j \mid \mathcal{G}_n) = \frac{f(\deg_n^-(j))}{n}. \quad (2)$$

Here  $\deg_n^-(j)$  denotes the indegree of vertex  $j$  after the  $n$ th vertex has been inserted.

Note that we have by construction that  $\deg_n^-(j) \leq d_0 + n - 1$ , so that by the condition on  $f$  the right hand side of (2) is indeed  $\leq 1$ .

These assumptions do not completely specify the model: they allow for deterministic as well as random outdegree and also only impose conditions on the marginal probabilities of connecting  $n + 1$  to  $j$ . In particular, the following models are included.

**Example 1.1** (Preferential attachment with random outdegree).

In the case  $d = 0$  and with (conditionally) independent outgoing edges, assumption (A) describes the model in [DM09]: Suppose  $f(n) \leq n + 1$  for all  $n \in \mathbb{N}_0$ . We start with the graph  $\mathcal{G}_1$  consisting of one vertex and no edges. At time  $n + 1$ , we add vertex  $n + 1$  to the vertex set and independently for each  $k \in [n]$  add a directed edge from  $n + 1$  to  $k$  with probability

$$\frac{f(\deg_n^-(k))}{n}.$$

**Example 1.2** (Preferential attachment with fixed outdegree). Start with  $\mathcal{G}_1$  consisting of vertex 1 and a (directed) self-loop. At time  $n + 1$  insert vertex  $n + 1$  and connect it to exactly one previous vertex  $j \in [n]$  with probability

$$\frac{\deg_n(j) + \delta}{n(2 + \delta)},$$

where  $\deg_n(j)$  denotes the total degree of vertex  $j$  at time  $n$  and  $\delta > -1$  is a parameter. Noticing that  $\deg_n(j) = \deg_n^-(j) + 1$ , this fits into our framework with  $f(k) = \frac{k + (1 + \delta)}{2 + \delta}$  and  $d_0 = 1$ . This model is almost the model proposed in [BRST01] (where however  $\delta = 0$  and we do not allow for self-loops of vertex  $n + 1$ ) and it is very closely related to what is referred to as model (b) in [Hof17, Chapter 8.2].

**Example 1.3** (Spatial preferential attachment model). In [ACJP08], the authors introduce the following spatial random graph model. Let  $S$  be the unit hypercube in  $\mathbb{R}^m$ . The initial graph consists of vertex 1 that is placed uniformly at random into  $S$  and no edges. For each vertex  $i$  we define *the sphere of influence*  $S(i, n)$  of  $i$  as the ball (in the torus metric induced by the Euclidean metric) that has volume  $\frac{A_1 \deg_n^i(i) + A_2}{n}$  centered at the position of  $i$ , where  $A_1, A_2 \geq 0$ . Fix a parameter  $p \in [0, 1]$ . Then, at time  $n + 1$ , we insert vertex  $n + 1$  at a position that is chosen uniformly at random in  $S$ . Now, independently for each vertex  $j$  such that the

position of  $n + 1$  is in  $S(j, n)$ , insert an edge from  $n + 1$  to  $j$  with probability  $p$ . In particular, we get

$$\mathbb{P}(n + 1 \text{ connects to } j \mid \mathcal{G}_n) = p \cdot \frac{A_1 \deg_n^-(j) + A_2}{n}.$$

Thus, this model fits into our framework if we choose  $f(k) = pA_1k + pA_2$  and where we assume that the constants are chosen such that  $pA_1, pA_2 \leq 1$ .

The limiting distribution of the asymptotic distribution of the indegree of a uniform vertex in the above examples is known (see e.g. [DM09, Thm. 1.1], [ACJP08, Thm. 1.1]) and it is given as  $\mu = (\mu_k)_{k \in \mathbb{N}_0}$  with

$$\mu_k = \frac{1}{1 + f(k)} \prod_{i=0}^{k-1} \frac{f(i)}{1 + f(i)}, \quad k \in \mathbb{N}_0. \quad (3)$$

This is a probability distribution (see e.g. Lemma 2.1 below) and moreover, by varying  $f$ , the limiting distribution can exhibit very different tail behaviour. As discussed in Examples 1.3 and 1.4 in [DM09] we obtain for  $f(k) = \gamma k + \beta$

$$\mu_k \sim \frac{\Gamma(\frac{\beta+1}{\gamma})}{\gamma \Gamma(\frac{\beta}{\gamma})} k^{-(1+\frac{1}{\gamma})}, \quad \text{for } k \rightarrow \infty.$$

Therefore, our framework allows for models with power-law distribution with tail exponent  $1 + 1/\gamma \in [2, \infty)$ . Furthermore, if  $f(k) \sim \gamma k^\alpha$  with  $0 < \alpha < 1$ ,  $\gamma > 0$ , then

$$\log \mu_k \sim \frac{1}{\gamma(1-\alpha)} k^{1-\alpha}, \quad \text{for } k \rightarrow \infty,$$

so that we obtain a limiting distribution with stretched exponential tails.

We can now state our main theorem.

**Theorem 1.4.** *Let  $W_n$  denote the indegree of a uniformly chosen vertex at time  $n$  in a preferential attachment model satisfying assumptions (A). Suppose further that there exists  $k_* \in \mathbb{N}_0$  such that  $f(k) > k$  for all  $k < k_*$  and  $f(k) \leq k$  for all  $k \geq k_*$ . Then, there exists a constant  $C > 0$  such that for all  $n \geq 2$*

$$d_{\text{TV}}(W_n, W) \leq C \frac{\log(n)}{n}, \quad (4)$$

where  $W \sim \mu$  and  $\mu$  as in (3).

The condition that there exists  $k_* \in \mathbb{N}_0$  such that  $f(k) > k$  for all  $k < k_*$  and  $f(k) \leq k$  for all  $k \geq k_*$  is for example fulfilled for all sublinear models such that  $\max_k (f(k+1) - f(k)) < 1$ , which is a popular condition in the setting of Example 1.1, see e.g. [DM13].

The next result gives a weaker result in the regime, when  $f(k) \in [k, k + \gamma]$  for all  $k$ . This is not surprising since for example in the case  $f(k) = k + \gamma$ , the distribution has power law exponent 2 and no longer has a finite mean.

**Theorem 1.5.** *Let  $W_n$  denote the indegree of a uniformly chosen vertex at time  $n$  in a model satisfying assumptions (A). Suppose further that  $f(k) \in [k, k + \gamma]$  for all  $k \in \mathbb{N}_0$  for some  $\gamma \in (0, 1)$ . Then, there exists a constant  $C > 0$  such that for all  $n \geq 1$ ,*

$$d_{\text{TV}}(W_n, W) \leq C n^{-(1-\gamma)},$$

where  $W \sim \mu$  and  $\mu$  as in (3).

Theorem 1.4 as well as Theorem 1.5 deal with the indegree of a uniformly chosen vertex. However, for the model described in Example 1.1 it also makes sense to ask about the distribution of the random outdegree. [DM09] show that the outdegree is approximately Poisson, the next theorem gives an error bound on this approximation.

**Theorem 1.6.** *Let  $D_n$  denote the outdegree of vertex  $n$  in the model described in Example 1.1 and suppose for some  $\gamma \in (0, 1)$ , we have  $f(k) \leq \gamma k + 1$  for all  $k \in \mathbb{N}_0$ . Then there exists  $C > 0$  such that*

$$d_{\text{TV}}(D_n, \text{Po}(\lambda_n)) \leq C \begin{cases} \frac{1}{n+1}, & \text{for } 0 < \gamma < \frac{1}{2}, \\ \frac{\log(n)}{n}, & \text{for } \gamma = \frac{1}{2}, \\ n^{-2(1-\gamma)}, & \text{for } \frac{1}{2} < \gamma < 1, \end{cases}$$

where  $\text{Po}(\lambda_n)$  denotes the Poisson distribution with parameter  $\lambda_n = \mathbb{E}[f(W_{n-1})]$  and  $W_{n-1}$  has the distribution of the indegree of a uniformly chosen vertex at time  $n - 1$ . Moreover,  $\lambda_n \rightarrow \lambda := \mathbb{E}[f(W)]$ , where  $W \sim \mu$  as in (3). Finally, if  $f(k) = \gamma k + \beta$  for  $\gamma \in (0, 1), \beta \in [0, 1]$ , then

$$|\lambda_n - \mathbb{E}[f(W)]| \leq n^{-1+\gamma}.$$

**Remark 1.7.** The assumptions on the initial graph as formulated in the Assumptions (A) are fairly restrictive. For example, the model (b) described in [Hof17, Chapter 8.2] does not fit exactly, as in that model we start with a graph of two vertices connected by two edges. Indeed, if we consider the model with fixed outdegree and change the initial conditions, then the normalization might not be an exact multiple of  $n$ . However, our proofs will also go through, if we change the assumption on the marginal connection probabilities to

$$\mathbb{P}(n + 1 \text{ connects to } j | \mathcal{G}_n) = \frac{f(\text{deg}_n^-(j))}{n + \alpha},$$

for some  $\alpha \geq 0$ . With this modification and a corresponding specification of the initial conditions, we can also deal with models with arbitrary starting configurations in the setting of Example 1.2, such as model (b) in [Hof17]. Another example that would fit into this set-up is the model  $\bar{I}$  described in [For09], where  $f$  is linear and connections are made independently to each previous vertex, but the normalization is  $n + \alpha$ . In particular, using the coupling arguments in [For09],

our results also transfer to the other models described there. However we cannot directly transfer the proofs to models with fitness, where  $\alpha$  would depend on  $n$ . In the following, in order to keep the notation simple, we restrict ourselves to the case  $\alpha = 0$ .

**Remark 1.8.** Our method does not cover directly the Barabási-Albert model in the form introduced in [BRST01] (and so neither the generalisation in [Hof17]), because they include the possibility of a self-loop for the incoming vertex. It is possible to couple these models with model (b) (from the previous remark). Unfortunately this coupling gives an error of order  $\frac{\log(n)^3}{n}$ . Therefore, we cannot recover the results of [PRR13b] and [Ros13], who obtain the same error rates as in our Theorem 1.4 for these models.

Other generalisations that we cannot deal with are the models with fixed outdegree where the probability to connect to an old vertex is proportional to a non-linear function as in [RTV07], which leads to a random normalization. Since this problem arises in a wide class of modifications of preferential attachment models, dealing with this problem is an interesting topic of future research.

### 1.3 Stein's method and overview of proof

Stein's method is a well-known tool for deriving rates of convergence of a random variable to a known target distribution. First developed for the normal distribution in [Ste72], it has been adapted to several other cases of target distributions, including Poisson in [Che75] but also many others. For a detailed overview of Stein's method see for example [Ros11].

The method can be divided into three main steps: first one has to find a characterizing operator  $\mathcal{A}$  of the target distribution  $\mu$ , in the sense that for all functions  $g$  in the domain of  $\mathcal{A}$  and any random variable  $W$ ,

$$\mathbb{E}[\mathcal{A}g(W)] = 0 \Leftrightarrow W \sim \mu. \quad (5)$$

The second step now consists of finding a solution  $g := g_h$  to the Stein-equation

$$h(k) - \int h d\mu = \mathcal{A}g(k) \quad (6)$$

for each  $h$  in a convergence-determining class of functions  $\mathcal{F}$ . This then yields for any random variable  $X$  and a random variable  $W \sim \mu$  that

$$d_{\mathcal{F}}(W, X) := \sup_{h \in \mathcal{F}} |\mathbb{E}[h(W)] - \mathbb{E}[h(X)]| = \sup_{h \in \mathcal{F}} \mathbb{E}[\mathcal{A}g_h(X)]. \quad (7)$$

Depending on  $\mathcal{F}$  we get different distances on probability measures. Here we are interested in the total variation distance so that we choose  $\mathcal{F} = \{\mathbb{1}_A(\cdot), A \in \text{Borel}(\mathbb{R})\}$ . In the last step one has to bound the right-hand side of (7) to get bounds on the distributional distance of  $W$  and  $X$ .



For the proofs of Theorems 1.4 and 1.5, in order to deal with steps 1 and 2, we will follow the generator approach introduced by Barbour [Bar88]. This approach applies when  $\mathcal{A}$  is the (infinitesimal) generator of a continuous-time Markov chain  $(Z_t)_{t \geq 0}$ . In particular, we have that (5) holds if and only if  $\mu$  is the stationary distribution of the Markov chain  $(Z_t)_{t \geq 0}$ . In Lemma 2.1, we will see that the Stein operator for our target distribution  $\mu$  in (3) is given as

$$\mathcal{A}g(k) = f(k)\Delta g(k) + g(0) - g(k), \quad (8)$$

where  $g : \mathbb{N}_0 \rightarrow \mathbb{R}$ ,  $\Delta g(k) := g(k+1) - g(k)$  and  $g$  such that  $\mathbb{E}[g(Y)] < \infty$  for  $Y \sim \mu$ . So the corresponding Markov process is a jump process  $(Z_t)_{t \geq 0}$  on  $\mathbb{N}_0$  that jumps from state  $k$  to  $k+1$  at rate  $f(k)$  and from any state at rate 1 back to the state 0.

Applying standard results for Markov chains also allows to take care of step 2 and we will show in Lemma 2.2 that  $g_h$  is given as

$$g_h(k) = - \int_0^\infty \left( \mathbb{E}_k h(Z_t) - \int h d\mu \right) dt, \quad (9)$$

for any  $h \in \mathcal{F}$ .

In the following we will explain the main parts of step 3, i.e. how to derive the claimed error bounds from the right hand side of (7). The first part is independent of preferential attachment models and we will show in Section 2.2 the following smoothness estimate for the Stein solution: for any  $g_A := g_{\mathbb{1}_A}$  and  $A \subset \mathbb{N}_0$ , we have that

$$f(k)\Delta g_A(k) \leq 1, \quad \text{for all } k \in \mathbb{N}_0. \quad (10)$$

In the second part, we use the following dynamic way of generating a uniform random variable on  $[n]$  (cf. [For09]): let  $J_n$  be a Markov chain with  $J_1 = 1$  and such that

$$\mathbb{P}(J_{n+1} = J_n | J_n) = \frac{n}{n+1} \quad \text{and} \quad \mathbb{P}(J_{n+1} = n+1 | J_n) = \frac{1}{n+1}.$$

Then, we have that  $J_n$  is uniformly distributed on  $[n]$  for every  $n$  (cf. also the proof of Lemma 2.5). In particular, we know that we can generate the indegree of a uniform vertex as  $X_n := \text{deg}_n^-(J_n)$  and moreover in Lemma 2.5 we show that  $(X_n)$  is a time-inhomogeneous Markov chain, with discrete-time generator

$$\frac{f(k)}{n+1}\Delta g(k) + \frac{1}{n+1}(g(0) - g(k)),$$

for suitable functions  $g : \mathbb{N}_0 \rightarrow \mathbb{R}$ . Note the similarities to the generator (8) of the homogeneous Markov chain  $(Z_t)_{t \geq 0}$ .

If we additionally assume that  $d_0 = 0$  in the Assumptions (A), then as a first step we use this Markov structure to show in Lemma 2.6 that

$$\mathbb{E}[Ag_A(X_{n+1})] = \frac{1}{n+1} \left( \sum_{\ell=1}^n \sum_{k=0}^{\ell-1} \Delta v_A(k) h(k, \ell) + v_A(0) \right), \quad (11)$$

where

$$v_A(k) := f(k)\Delta g_A(k) \text{ and } h(k, \ell) := f(k)\mathbb{P}(X_\ell = k) - \mathbb{P}(X_\ell \geq k + 1).$$

Then, since we have the smoothness estimate (10), as a last step it remains to analyse  $h(k, \ell)$  and show that these terms are small (in fact they are zero for  $X_\ell \sim \mu$ ). The corresponding analysis is carried out in Proposition 2.7. We first show inductively, that under the conditions of Theorems 1.4 and 1.5, for fixed  $\ell$  the functions  $k \mapsto h(k, \ell)$  are first increasing and then decreasing, which allows ultimately to deal with the inner sum over  $k$ . It is here that we used the specific assumptions on  $f$ . Then, finally we show that for suitable constants  $C > 0$ ,  $k \leq \ell - 1$ ,

$$h(k, \ell) \leq \begin{cases} C \ell^{-1} & \text{under the assumptions of Thm. 1.4,} \\ C \ell^{-(1-\gamma)} & \text{under the assumptions of Thm. 1.5.} \end{cases}$$

These bounds then lead via (11) to the error bounds of  $\log(n)/n$  and  $n^{-(1-\gamma)}$  in Theorems 1.4 and 1.5. The case  $d_0 > 0$  will be shown using an easy coupling argument.

In contrast to the above, Theorem 1.6, which considers the random outdegree in the model given in Example 1.1, can be proven by classical arguments for the sum of indicator functions.

## 2 Proofs: Stein for the indegree distribution

We will follow the strategy outlined in Section 1.3.

### 2.1 Deriving the Stein operator

In this section, we will deal with the Stein operator for  $\mu$  defined in (3) and show how to solve the Stein equation (6). This corresponds to steps 1 and 2 of the general strategy outlined in Section 1.3.

**Lemma 2.1.** *Let  $\mu$  be given by (3). Then,  $\mu$  is a probability distribution and any  $\mathbb{N}_0$ -valued random variable  $W$  satisfies  $W \sim \mu$  if and only if*

$$\mathbb{E}[Ag(W)] = 0, \tag{12}$$

for all  $g : \mathbb{N}_0 \rightarrow \mathbb{R}$  such that  $\mathbb{E}[g(W)] < \infty$ , where

$$Ag(k) := f(k)(g(k+1) - g(k)) + g(0) - g(k).$$

*Proof.* Let  $N_t$  be the process that starts in 0 and jumps from  $i$  to  $i + 1$  at rate  $f(i)$ . Then, the time of the  $k$ th jump is in distribution given by  $S_k = \sum_{i=0}^{k-1} \frac{1}{f(i)} E_i$ , where  $E_0, E_1, \dots$  is an i.i.d. sequence of exponentials with rate 1. Following Cor. 50 in [Bha07]), let  $Y$  be an independent exponential random variable, so that

$$\mathbb{P}(N_Y \geq k) = \mathbb{P}(Y \geq S_k).$$

By first conditioning on  $S_k$ , we get

$$\begin{aligned} \mathbb{P}(Y \geq S_k) &= \mathbb{E}[\mathbb{E}[\mathbb{1}\{Y \geq S_k\} | S_k]] = \mathbb{E}[e^{-S_k}] = \mathbb{E}[e^{-\sum_{i=0}^{k-1} \frac{1}{f(i)} E_i}] \\ &= \prod_{i=0}^{k-1} \mathbb{E}[e^{-E_i/f(i)}] = \prod_{i=0}^{k-1} \frac{f(i)}{1 + f(i)}, \end{aligned} \quad (13)$$

using that  $\mathbb{E}[e^{-\lambda E_i}] = \frac{1}{1+\lambda}$  in the last step. This implies

$$\begin{aligned} \mathbb{P}(N_Y = k) &= \prod_{i=0}^{k-1} \frac{f(i)}{1 + f(i)} - \prod_{i=0}^k \frac{f(i)}{1 + f(i)} \\ &= \left(1 - \frac{f(k)}{1 + f(k)}\right) \prod_{i=0}^{k-1} \frac{f(i)}{1 + f(i)} = \frac{1}{1 + f(k)} \prod_{i=0}^{k-1} \frac{f(i)}{1 + f(i)} = \mu_k. \end{aligned}$$

In particular this shows that  $\mu$  defines a probability measure on  $\mathbb{N}_0$  and (13) gives

$$\mu([k, \infty)) = \prod_{i=0}^{k-1} \frac{f(i)}{1 + f(i)} = f(k-1)\mu_{k-1}. \quad (14)$$

Using this connection to a jump process we obtain by integration by parts

$$\begin{aligned} \sum_{k=1}^{\infty} g(k)\mu_k &= \mathbb{E}[g(N_Y)] = \int_0^{\infty} \mathbb{E}[g(N_s)]e^{-s} ds \\ &= g(0) + \int_0^{\infty} \mathbb{E}\left[g\left((N_s + 1) - g(N_s)\right)f(N_s)\right]e^{-s} ds \\ &= g(0) + \mathbb{E}\left[g\left((N_Y + 1) - g(N_Y)\right)f(N_Y)\right]. \end{aligned}$$

In particular, if  $W \sim \mu$ , then  $W = Z_Y$  and we have just shown that in this case (12) holds. To show the converse statement suppose  $W$  with distribution  $(p_k)_{k \in \mathbb{N}_0}$  satisfies (12). Then, by applying (12) with  $h = \mathbb{1}_\ell$ , we obtain the following recursion

$$(1 + f(0))p_0 = 1, \quad \text{and} \quad (1 + f(\ell))p_\ell = f(\ell - 1)p_{\ell-1}, \quad \text{for } \ell \in \mathbb{N}.$$

Solving the recursion yields  $p_k = \mu_k$  for all  $k \in \mathbb{N}_0$ .  $\square$

Following the generator method introduced by Barbour in [Bar88], the solution is formally given by

$$g_h(k) = - \int_0^\infty \left( \mathbb{E}_k h(Z_t) - \int h d\mu \right) dt, \quad (15)$$

for all  $h$ , when the integral exists. Here  $(Z_t)_{t \geq 0}$  denotes the Markov process with infinitesimal generator  $\mathcal{A}$ ,  $\mathbb{E}_k$  its distribution conditional on  $Z_0 = k$ , and  $\mu$  its equilibrium distribution.

**Lemma 2.2.** *The unique solution of the Stein equation for  $\mu$ , i.e.*

$$\mathcal{A}g = h - \mu(h), \quad (16)$$

for any  $h = \mathbb{1}_A$ , with  $A \subset \mathbb{N}_0$  is given by

$$g_h(k) = - \int_0^\infty \left( \mathbb{E}_k h(Z_t) - \int h d\mu \right) dt, \quad (17)$$

where  $(Z_t)_{t \geq 0}$  is the continuous-time Markov process with generator  $\mathcal{A}$  started in  $k$  under  $\mathbb{P}_k$  (resp.  $\mathbb{E}_k$ ).

*Proof.* Let  $h = \mathbb{1}_A$  for some  $A \subset \mathbb{N}_0$ . One can check that the Markov chain  $(Z_t)$  with generator  $\mathcal{A}$  is irreducible, non-explosive and has invariant distribution  $\mu$ . It follows that  $\mathbb{E}_k h(Z_t) \rightarrow \mu(h)$  as  $t \rightarrow \infty$ .

By Kolmogorov's backward equation for the Markov chain  $(Z_t)_{t \geq 0}$ , see e.g. [Nor98, Thm. 2.8.4.], we have for any  $k, j \in \mathbb{N}_0$ ,

$$\begin{aligned} \mathbb{P}_k(Z_t = j) - \mathbb{1}_{\{k\}}(j) &= \mathbb{P}_k(Z_t = j) - \mathbb{P}_k(Z_0 = j) \\ &= \int_0^t \left\{ f(k)(\mathbb{P}_{k+1}(Z_s = j) - \mathbb{P}_k(Z_s = j)) + \mathbb{P}_0(Z_s = j) - \mathbb{P}_k(Z_s = j) \right\} ds. \end{aligned}$$

In particular, as  $h$  as above is bounded, we have by Fubini

$$\mathbb{E}_k[h(Z_t)] - h(k) = \int_0^t f(k)(\mathbb{E}_{k+1}[h(Z_s)] - \mathbb{E}_k[h(Z_s)]) + \mathbb{E}_0[h(Z_s)] - \mathbb{E}_k[h(Z_s)] ds.$$

Now, define

$$g_h^t(k) := - \int_0^t (\mathbb{E}_k h(Z_s) - \mu(h)) ds.$$

Thus, we can deduce that

$$\mathbb{E}_k[h(Z_t)] - h(k) = -(\mathcal{A}g_h^t)(k).$$

We note that the left hand side converges to  $\mu(h) - h(k)$  as  $\mu$  is the invariant distribution of  $(Z_t)_{t \geq 0}$ . By the definition of  $\mathcal{A}$  it is straightforward to see that the

operator  $\mathcal{A}$  is closed under pointwise convergence. Hence, it suffices to show that  $g_h^t(k)$  converges to  $g_h(k)$  as defined in (17) in order to conclude that  $g_h$  solves (16).

To see the latter, we can estimate since  $h = \mathbb{1}_A$ ,

$$\begin{aligned} \int_0^\infty |\mathbb{E}_k[h(Z_t)] - \mu(h)| dt &= \int_0^\infty \left| \mathbb{E} \left[ h(Z_t^{(k)}) \right] - \mathbb{E} \left[ h(Z_t^{(\mu)}) \right] \right| dt \\ &\leq \int_0^\infty \min_{(X_t, Y_t) \text{ coupling of } (Z_t^{(k)}, Z_t^{(\mu)})} \mathbb{P}(X_t \neq Y_t) dt, \end{aligned}$$

where we denote by  $Z_t^{(k)}$  and  $Z_t^{(\mu)}$  the Markov chains started in  $k$  and  $\mu$  respectively. We will now construct a coupling  $(X_t, Y_t)$  of  $Z_t^{(k)}$  and  $Z_t^{(\mu)}$  in the following way: let  $X_0 = k$  and choose  $Y_0$  according to  $\mu$ . We let both chains evolve independently until  $Y$  falls to zero at a random time  $\tau$ . We then force  $X$  to fall to zero as well and from that point on the two chains evolve together, so that  $X_t = Y_t$  for  $t \geq \tau$ . One can easily check that this defines a coupling of  $Z_t^{(k)}$  and  $Z_t^{(\mu)}$ . As  $Y$  falls down to zero at rate one we get

$$\mathbb{P}(X_t \neq Y_t) \leq \mathbb{P}(\tau \geq t) = e^{-t}$$

so that

$$\int_0^\infty |\mathbb{E}_k[h(Z_t)] - \mu(h)| dt \leq \int_0^\infty e^{-t} dt = 1.$$

Hence,  $g_h^t(k) \rightarrow g_h(k)$  by dominated convergence, which completes the proof of the lemma.  $\square$

## 2.2 Bounding the Stein solution

In this section, we exploit the connection to the Markov process with generator  $\mathcal{A}$  to find bounds on the Stein solution  $g_A := g_{\mathbb{1}_A}$  for  $A \subset \mathbb{N}_0$ . This part of the proof only depends on the limiting distribution  $\mu$  and does not make use of the preferential attachment setting. More precisely, we would like to show that  $v_A(k) := f(k)\Delta g_A(k)$  is uniformly bounded in  $k$ . As we will see later on,  $g_A$  can be written as a linear combination of  $g_j := g_{\mathbb{1}_{\{j\}}}$ . Therefore, we start by calculating the latter.

**Lemma 2.3.** *For  $g_j := g_h$  for  $h = \mathbb{1}(\cdot = j)$  as in (17) we have*

$$\Delta g_j(k) = g_j(k+1) - g_j(k) = \begin{cases} -\frac{\mu_j}{\mu_k} \frac{1}{f(k)(1+f(k))}, & \text{for } j \geq k+1, \\ \frac{1}{1+f(j)}, & \text{for } j = k, \\ 0, & \text{for } j \leq k-1, \end{cases} \quad (18)$$

where  $\mu$  is the limit distribution given in (3).

*Proof.* We apply the techniques used in [BX01] and adapt them to our Markov process  $Z_t$ . Define

$$\tau_{k,k+1} = \inf\{t : Z_t^{(k)} = k + 1\},$$

where  $Z_t^{(k)}$  denotes a Markov process with generator  $\mathcal{A}$  started in  $k$ . Then, for  $k \leq j - 1$  we obtain via the representation (17) of the Stein solution,

$$\begin{aligned} g_j(k) &= - \int_0^\infty (\mathbb{E}[\mathbb{1}_{\{j\}}(Z_t^{(k)})] - \mu_j) dt \\ &= -\mathbb{E}\left[\int_0^{\tau_{k,k+1}} (\mathbb{1}_{\{j\}}(Z_t^{(k)}) - \mu_j) dt\right] - \mathbb{E}\left[\int_{\tau_{k,k+1}}^\infty (\mathbb{1}_{\{j\}}(Z_t^{(k)}) - \mu_j) dt\right] \\ &= \mu_j \mathbb{E}[\tau_{k,k+1}] + g_j(k+1), \end{aligned}$$

where the last equality uses the strong Markov property of  $Z_t$ . Rearranging yields

$$g_j(k) - g_j(k+1) = \mu_j \mathbb{E}[\tau_{k,k+1}] \geq 0$$

for  $k \leq j - 1$ . Following the same procedure for  $k \geq j + 1$  and  $\tau_{k,0} = \inf\{t : Z_t^{(k)} = 0\}$  we get

$$\begin{aligned} g_j(k) &= - \int_0^\infty (\mathbb{E}[\mathbb{1}_{\{j\}}(Z_t^{(k)})] - \mu_j) dt \\ &= -\mathbb{E}\left[\int_0^{\tau_{k,0}} (\mathbb{1}_{\{j\}}(Z_t^{(k)}) - \mu_j) dt\right] - \mathbb{E}\left[\int_{\tau_{k,0}}^\infty (\mathbb{1}_{\{j\}}(Z_t^{(k)}) - \mu_j) dt\right] \\ &= \mu_j \mathbb{E}[\tau_{k,0}] + g_j(0) \end{aligned}$$

and thus

$$g_j(k) - g_j(0) = \mu_j \mathbb{E}[\tau_{k,0}]$$

for  $k \geq j + 1$ . Noticing that  $\mathbb{E}[\tau_{k,0}] = 1$  for all  $k$ , since the rate by which the process  $Z_t$  falls down to 0 is 1, independent of the current state of the process, this simplifies to

$$g_j(k) - g_j(0) = \mu_j.$$

This means that  $g_j(k)$  is constant for  $k \geq j + 1$ , so that

$$g_j(k+1) - g_j(k) = 0, \text{ for } k \geq j + 1.$$

Furthermore, we have

$$\begin{aligned} g_j(j+1) &= - \int_0^\infty (\mathbb{E}[\mathbb{1}_{\{j\}}(Z_t^{(j+1)})] - \mu_j) dt \\ &= -\mathbb{E}\left[\int_0^{\tau_{j+1,j}} (\mathbb{1}_{\{j\}}(Z_t^{(j+1)}) - \mu_j) dt\right] - \mathbb{E}\left[\int_{\tau_{j+1,j}}^\infty (\mathbb{1}_{\{j\}}(Z_t^{(j+1)}) - \mu_j) dt\right] \\ &= \mu_j \mathbb{E}[\tau_{j+1,j}] + g_j(j) \\ &= \mu_j (\mathbb{E}[\tau_{j+1,0}] + \mathbb{E}[\tau_{0,j}]) + g_j(j) \\ &= \mu_j (1 + \mathbb{E}[\tau_{0,j}]) + g_j(j), \end{aligned}$$

yielding

$$g_j(j+1) - g_j(j) = \mu_j (1 + \mathbb{E}[\tau_{0,j}]) = \mu_j \mathbb{E}[\tau_{j,j}] \geq 0,$$

where  $\tau_{j,j}$  defines the first return time to  $j$  of the Markov chain started at  $j$ .

Thus the following equations hold for the Stein solution  $g_j$ :

$$g_j(k+1) - g_j(k) = \begin{cases} -\mu_j \mathbb{E}[\tau_{k,k+1}], & \text{for } j \geq k+1, \\ \mu_j \mathbb{E}[\tau_{j,j}], & \text{for } j = k, \\ 0, & \text{for } j \leq k-1. \end{cases} \quad (19)$$

We can simplify this expression further as follows. Let  $S_k$  be the first jump time of  $Z_t^{(k)}$ , then by definition of the Markov chain, we have that  $S_k \sim \text{Exp}(1 + f(k))$ . Since  $Z_t^{(k)}$  jumps to  $k+1$  at rate  $f(k)$  and to 0 at rate 1 we obtain

$$\mathbb{E}[\tau_{k,k+1}] = \mathbb{E}[S_k] + \frac{1}{1+f(k)} \mathbb{E}[\tau_{0,k+1}] = \frac{1}{1+f(k)} (1 + \mathbb{E}[\tau_{0,k}] + \mathbb{E}[\tau_{k,k+1}])$$

and thus

$$\mathbb{E}[\tau_{k,k+1}] = \frac{1}{f(k)} (1 + \mathbb{E}[\tau_{0,k}]) = \frac{1}{f(k)} \mathbb{E}[\tau_{k,k}]. \quad (20)$$

A standard result for Markov chains, see e.g. [Nor98, Thm. 3.6.3], yields that

$$\mu_j = \frac{1}{(1+f(j)) \mathbb{E}[\tau_{j,j}]} \quad (21)$$

Rearranging (21) and combining it with (19) and (20) yields the statement of the lemma.  $\square$

**Proposition 2.4.** *For any  $k \in \mathbb{N}_0$  and  $A \subset \mathbb{N}_0$ , we have*

$$|v_A(k)| \leq 1.$$

*Proof.* First note that

$$\begin{aligned} g_A(k) &= - \int_0^\infty (\mathbb{E}[\mathbb{1}(Z_t^{(k)} \in A)] - \mu(A)) dt \\ &= - \int_0^\infty \left( \mathbb{E} \left[ \sum_{j \in A} \mathbb{1}(Z_t^{(k)} = j) \right] - \sum_{j \in A} \mu_j \right) dt \\ &= \sum_{j \in A} \left( - \int_0^\infty (\mathbb{E}[\mathbb{1}(Z_t^{(k)} = j)] - \mu_j) dt \right) = \sum_{j \in A} g_j(k). \end{aligned} \quad (22)$$

Consequently, we have by Lemma 2.3

$$v_A(k) = \sum_{j \in A} f(k) \Delta g_j(k) = - \frac{f(k)}{\mu_k f(k) (1+f(k))} \mu(A \cap [k, \infty)) + \frac{f(k)}{1+f(k)} \mathbb{1}_{k \in A}.$$

Using that by (14)  $\mu([k, \infty)) = \prod_{i=0}^{k-1} \frac{f(i)}{1+f(i)} = (1+f(k))\mu_k$ , we obtain

$$v_A(k) = -\frac{1}{\mu([k, \infty))}\mu(A \cap [k, \infty)) + \frac{f(k)}{1+f(k)}\mathbb{1}_{k \in A},$$

so that the proposition follows immediately.  $\square$

### 2.3 Applying Stein's method to the preferential attachment model

We use the following Markov chain to describe the evolution of the indegree of a uniform vertex. Similar ideas were also used in [DM09, For09].

**Lemma 2.5.** *Let  $X_n$  be a Markov chain with  $\mathbb{P}(X_1 = 0) = 1$  and transition probabilities given for any  $i \geq 1$  as*

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \begin{cases} \frac{f(i)}{n+1} & \text{if } j = i + 1, \\ \frac{n-f(i)}{n+1} & \text{if } j = i, \\ \frac{1}{n+1} & \text{if } j = 0, \end{cases}$$

and

$$\mathbb{P}(X_{n+1} = j | X_n = 0) = \begin{cases} \frac{f(0)}{n+1} & \text{if } j = 1, \\ 1 - \frac{f(0)}{n+1} & \text{if } j = 0. \end{cases}$$

Then  $\mathcal{L}(X_n) = \mathcal{L}(W_n)$ , where  $W_n$  denotes the indegree of the uniformly chosen vertex in any preferential attachment model at time  $n$  satisfying Assumptions (A) with  $d_0 = 0$ .

Note that the Markov chain starts at time 1 to match the index of the random graph.

*Proof.* Recall the Markov chain  $(J_n)_{n \in \mathbb{N}}$  started with  $J_1 = 1$  and such that for  $n \geq 1$

$$\mathbb{P}(J_{n+1} = J_n | J_n) = \frac{n}{n+1} \quad \text{and} \quad \mathbb{P}(J_{n+1} = n+1 | J_n) = \frac{1}{n+1}.$$

Then it is straight-forward to check by induction that  $J_n$  is uniformly distributed on  $\{1, \dots, n\}$ . We now set  $X_n := \deg_n^-(J_n)$ , so that in particular  $\mathcal{L}(W_n) = \mathcal{L}(X_n)$ . Then using the dynamics of the preferential attachment model and the tower property, the following transition probabilities hold for  $X_n$  and  $1 \leq j \leq n$

$$\begin{aligned} \mathbb{P}(X_{n+1} = j+1 | X_n = j) &= \frac{f(j)}{n} \times \frac{n}{n+1} = \frac{f(j)}{n+1}, \\ \mathbb{P}(X_{n+1} = j | X_n = j) &= \left(1 - \frac{f(j)}{n}\right) \times \frac{n}{n+1} = \frac{n-f(j)}{n+1}, \\ \mathbb{P}(X_{n+1} = 0 | X_n = j) &= \frac{1}{n+1}. \end{aligned}$$



Moreover, for  $j = 0$ ,

$$\begin{aligned}\mathbb{P}(X_{n+1} = 1 | X_n = 0) &= \frac{f(0)}{n} \times \frac{n}{n+1} = \frac{f(0)}{n+1}, \\ \mathbb{P}(X_{n+1} = 0 | X_n = 0) &= \left(1 - \frac{f(0)}{n}\right) \times \frac{n}{n+1} + \frac{1}{n+1} = 1 - \frac{f(0)}{n+1}.\end{aligned}$$

This completes the proof of the lemma.  $\square$

Following the discussion above in Section 1.3, the next step is to find a bound on

$$\begin{aligned}\mathbb{E}[\mathcal{A}g_A(W_{n+1})] &= \mathbb{E}[\mathcal{A}g_A(X_{n+1})] \\ &= \mathbb{E}[f(X_{n+1})\Delta g_A(X_{n+1}) + (g_A(0) - g_A(X_{n+1}))],\end{aligned}$$

where we recall that  $\Delta g_A(k) := g_A(k+1) - g_A(k)$ .

**Lemma 2.6.** *For the Markov chain  $(X_n)_{n \geq 0}$  defined in Lemma 2.5 and  $v_A(k) := f(k)\Delta g_A(k)$ , with  $A \subset \mathbb{N}_0$ , we have*

$$\begin{aligned}\mathbb{E}[\mathcal{A}g_A(X_{n+1})] &= \frac{1}{n+1} \left( \left( \sum_{\ell=1}^n \sum_{k=0}^{\ell-1} \Delta v_A(k) (f(k)\mathbb{P}(X_\ell = k) - \mathbb{P}(X_\ell \geq k+1)) \right) + v_A(0) \right).\end{aligned}\quad (23)$$

Note that according to (14) we have  $f(k)\mathbb{P}(X = k) - \mathbb{P}(X \geq k+1) = 0$  if  $X$  has the limiting distribution  $\mu$ . In Section 2.4 we will show that  $f(k)\mathbb{P}(X_\ell = k) - \mathbb{P}(X_\ell \geq k+1)$  is close to zero for large  $\ell$ .

*Proof.* Let  $h : \mathbb{N}_0 \rightarrow \mathbb{R}$  be such that  $h(0) = 0$ . Then,

$$\begin{aligned}\mathbb{E}[h(X_{n+1})] &= \mathbb{E}[\mathbb{E}[h(X_{n+1}) | X_n]] \\ &= \mathbb{E}\left[h(X_n) \frac{n - f(X_n)}{n+1} + h(X_n + 1) \frac{f(X_n)}{n+1}\right] \\ &= \frac{n}{n+1} \mathbb{E}[h(X_n)] + \frac{1}{n+1} \mathbb{E}[f(X_n)\Delta h(X_n)].\end{aligned}$$

By iteration and using that  $X_1 = 0$ , we get

$$\mathbb{E}[h(X_{n+1})] = \frac{1}{n+1} \sum_{\ell=1}^n \mathbb{E}[f(X_\ell)\Delta h(X_\ell)].$$

Define

$$h(k) = \mathcal{A}g_A(k) - v_A(0) = v_A(k) + (g_A(0) - g_A(k)) - v_A(0).$$

Thus, we obtain

$$\begin{aligned}
\mathbb{E}[\mathcal{A}g_A(X_{n+1})] &= \mathbb{E}[h(X_{n+1})] + v_A(0) \\
&= \frac{1}{n+1} \sum_{\ell=1}^n \left( \mathbb{E}[f(X_\ell)\Delta v_A(X_\ell)] - \mathbb{E}[f(X_\ell)\Delta g_A(X_\ell)] \right) + v_A(0) \\
&= \frac{1}{n+1} \sum_{\ell=1}^n \sum_{k=0}^{\ell-1} f(k)\Delta v_A(k)\mathbb{P}(X_\ell = k) \\
&\quad - \frac{1}{n+1} \sum_{\ell=1}^n \sum_{k=0}^{\ell-1} (v_A(k) - v_A(0))\mathbb{P}(X_\ell = k) + \frac{v_A(0)}{n+1}. \tag{24}
\end{aligned}$$

For the second sum, we write

$$\begin{aligned}
\sum_{k=0}^{\ell-1} (v_A(k) - v_A(0))\mathbb{P}(X_\ell = k) &= \sum_{k=0}^{\ell-1} \sum_{i=0}^{k-1} \Delta v_A(i)\mathbb{P}(X_\ell = k) \\
&= \sum_{i=0}^{\ell-1} \Delta v_A(i) \sum_{k=i+1}^{\ell-1} \mathbb{P}(X_\ell = k) = \sum_{i=0}^{\ell-1} \Delta v_A(i)\mathbb{P}(X_\ell \geq i+1).
\end{aligned}$$

Combining the latter with (24) yields the statement of the lemma.  $\square$

## 2.4 Bounding the difference $f(k)\mathbb{P}(X_\ell = k) - \mathbb{P}(X_\ell \geq k+1)$

**Proposition 2.7.** *Suppose  $f$  satisfies  $f(k) \leq k+1$ . Define*

$$h(k, \ell) := f(k)\mathbb{P}(X_\ell = k) - \mathbb{P}(X_\ell \geq k+1),$$

where  $(X_\ell)_{\ell \geq 1}$  is the Markov chain from Lemma 2.5.

(i) *Then, for any  $k \in \mathbb{N}_0, \ell \in \mathbb{N}$  we have*

$$h(k, \ell) \geq 0,$$

and for  $k \geq \ell$ , we have  $h(k, \ell) = 0$ .

(ii) *Suppose there exists  $K$  such that  $k \leq f(k) \leq k+1$  for all  $0 \leq k \leq K$ , then for all  $\ell \leq K+1$ , we have*

$$h(k+1, \ell) - h(k, \ell) \geq 0 \quad \text{for all } k \leq \ell - 2. \tag{25}$$

(iii) *Assume that there exist  $k_*$  such that  $f(k) \leq k$  for all  $k \geq k_*$  and  $f(k) > k$  for  $k < k_*$ . Then, for all  $\ell \in \mathbb{N}$  there exists  $I(\ell) \in \{0, \dots, \ell-1\}$  such that*

$$h(k+1, \ell) - h(k, \ell) \begin{cases} \geq 0 & \text{if } k < I(\ell), \\ \leq 0 & \text{if } k \geq I(\ell). \end{cases} \tag{26}$$

Moreover,  $I(\ell+1) \in \{I(\ell), I(\ell)+1\}$ .

(iv) Assume there exists  $k_* \in \mathbb{N}_0$  such that  $f(k) \leq k$  for all  $k \geq k_*$ . Then, there exists a constant  $C > 0$  such that for all  $k \in \mathbb{N}_0, \ell \in \mathbb{N}$ ,

$$h(k, \ell) \leq \frac{C}{\ell}.$$

(v) If there exists  $\gamma \in (0, 1)$  such that  $f(k) \in [k, k + \gamma]$  for all  $k \in \mathbb{N}_0$ , then

$$\sup_{k \in \mathbb{N}_0} h(k, \ell) \leq C\ell^{-(1-\gamma)}.$$

Before we start with the proof of the proposition, we derive a recursion for the coefficients  $h$  and also for its increments.

**Lemma 2.8.** *Let  $h$  be defined as in Proposition 2.7, then  $h(k, \ell) = 0$  for all  $k \geq \ell$  and for all  $\ell \in \mathbb{N}, k \in \mathbb{N}_0$ , we have*

$$h(k, \ell + 1) = \left( \frac{\ell}{\ell + 1} - \frac{f(k)}{\ell + 1} \right) h(k, \ell) + \frac{f(k)}{\ell + 1} h(k - 1, \ell), \quad (27)$$

where we define  $h(-1, \ell) = 0$ . Moreover, if we define  $\Delta^{(1)}h(k, \ell) := h(k + 1, \ell) - h(k, \ell)$ , we have that for all  $\ell \in \mathbb{N}$  and  $k \leq \ell - 1$

$$\Delta^{(1)}h(k, \ell + 1) = \left( \frac{\ell}{\ell + 1} - \frac{f(k + 1)}{\ell + 1} \right) \Delta^{(1)}h(k, \ell) + \frac{f(k)}{\ell + 1} \Delta^{(1)}h(k - 1, \ell). \quad (28)$$

*Proof.* Note that since  $X_\ell \leq \ell - 1$   $\mathbb{P}$ -a.s, we have  $h(k, \ell) = 0$  for any  $k \geq \ell$ . Moreover, by the definition of the Markov chain  $(X_n)$ , for  $k \geq 1$ ,

$$\begin{aligned} h(k, \ell + 1) &= f(k)\mathbb{P}(X_{\ell+1} = k) - \mathbb{P}(X_{\ell+1} \geq k + 1) \\ &= f(k) \left( \left(1 - \frac{f(k) + 1}{\ell + 1}\right) \mathbb{P}(X_\ell = k) + \frac{f(k - 1)}{\ell + 1} \mathbb{P}(X_\ell = k - 1) \right) \\ &\quad - \left( \frac{\ell}{\ell + 1} \mathbb{P}(X_\ell \geq k + 1) + \frac{f(k)}{\ell + 1} \mathbb{P}(X_\ell = k) \right) \\ &= \left( \frac{\ell}{\ell + 1} - \frac{f(k)}{\ell + 1} \right) h(k, \ell) - \frac{f(k)}{\ell + 1} \mathbb{P}(X_\ell \geq k + 1) \\ &\quad + \frac{f(k)}{\ell + 1} (f(k - 1)\mathbb{P}(X_\ell = k - 1) - \mathbb{P}(X_\ell = k)) \\ &= \left( \frac{\ell}{\ell + 1} - \frac{f(k)}{\ell + 1} \right) h(k, \ell) + \frac{f(k)}{\ell + 1} h(k - 1, \ell). \end{aligned}$$

Note for  $k = 0$ , we have

$$\begin{aligned} h(0, \ell + 1) &= f(0)\mathbb{P}(X_{\ell+1} = 0) - \mathbb{P}(X_{\ell+1} \geq 1) \\ &= f(0) \left( \left(1 - \frac{f(0)}{\ell + 1}\right) \mathbb{P}(X_\ell = 0) + \frac{1}{\ell + 1} \mathbb{P}(X_\ell \geq 1) \right) \\ &\quad - \frac{f(0)}{\ell + 1} \mathbb{P}(X_\ell = 0) - \left(1 - \frac{1}{\ell + 1}\right) \mathbb{P}(X_\ell \geq 1) \\ &= \left(1 - \frac{f(0) + 1}{\ell + 1}\right) h(0, \ell) = \left(\frac{\ell}{\ell + 1} - \frac{f(0)}{\ell + 1}\right) h(0, \ell). \end{aligned}$$

Therefore, the identity (27) also holds for  $k = 0$  since we defined  $h(-1, \ell) = 0$  for all  $\ell \in \mathbb{N}$ .

Note that by (27),

$$\begin{aligned} & h(k+1, \ell+1) - h(k, \ell+1) \\ &= \left( \frac{\ell}{\ell+1} - \frac{f(k+1)}{\ell+1} \right) h(k+1, \ell) + \frac{f(k+1)}{\ell+1} h(k, \ell) \\ &\quad - \left( \frac{\ell}{\ell+1} - \frac{f(k)}{\ell+1} \right) h(k, \ell) - \frac{f(k)}{\ell+1} h(k-1, \ell) \\ &= \left( \frac{\ell}{\ell+1} - \frac{f(k+1)}{\ell+1} \right) (h(k+1, \ell) - h(k, \ell)) + \frac{f(k)}{\ell+1} (h(k, \ell) - h(k-1, \ell)), \end{aligned}$$

which shows (28).  $\square$

*Proof of Proposition 2.7.* Before we start with the proof, note that for  $\ell = 2$ , we have

$$\begin{aligned} h(0, 2) &= f(0)\mathbb{P}(X_2 = 0) - \mathbb{P}(X_2 \geq 1) \\ &= f(0) \left( 1 - \frac{f(0)}{2} \right) - \frac{f(0)}{2} = \frac{f(0)(1-f(0))}{2}, \end{aligned} \quad (29)$$

and moreover,

$$h(1, 2) = f(1)\mathbb{P}(X_2 = 1) = \frac{f(1)f(0)}{2}. \quad (30)$$

(i) We now show that for any  $\ell \in \mathbb{N}$ :  $h(k, \ell) \geq 0$  for any  $k \in \mathbb{N}_0$  by induction on  $\ell$ . The induction hypothesis follows from (29) and (30) since  $f(0) \leq 1$ . We now assume that the statement holds for some  $\ell$ , then from (27) and for  $k \leq \ell - 1$  we have

$$h(k, \ell+1) = \left( \frac{\ell}{\ell+1} - \frac{f(k)}{\ell+1} \right) h(k, \ell) + \frac{f(k)}{\ell+1} h(k-1, \ell),$$

which is nonnegative due to the induction hypothesis and the condition that  $f(k) \leq k+1 \leq \ell$ . For  $k = \ell$  we have  $h(\ell, \ell+1) = f(\ell)\mathbb{P}(X_{\ell+1}) \geq 0$ . This implies the induction step since all other terms are 0.

(ii) Let now  $f$  be such that there exists  $K$  such that  $k \leq f(k) \leq k+1$  for all  $k \leq K$ . As before we will use induction on  $\ell$  to show the stated result. For  $\ell = 2$  we get from (29) and (30)

$$h(1, 2) - h(0, 2) = \frac{f(0)}{2} (f(0) + f(1) - 1) \geq 0,$$

as  $f(1) \geq 1$  by assumption.

Suppose that the statement (25) is true for some  $\ell \leq K$ . By (28) we obtain

$$\Delta^{(1)} h(k, \ell+1) = \left( \frac{\ell}{\ell+1} - \frac{f(k+1)}{\ell+1} \right) \Delta^{(1)} h(k, \ell) + \frac{f(k)}{\ell+1} \Delta^{(1)} h(k-1, \ell),$$

which is nonnegative for  $k \leq \ell - 2$  due to the induction hypothesis and since  $f(k+1) \leq f(\ell-1) \leq \ell$ . It remains to show that  $\Delta^{(1)}h(\ell-1, \ell+1) \geq 0$ . Again by (28) and using that  $\Delta^{(1)}h(\ell-1, \ell) = -h(\ell-1, \ell)$ , we get that

$$\begin{aligned}\Delta^{(1)}h(\ell-1, \ell+1) &= \left( \frac{\ell}{\ell+1} - \frac{f(\ell)}{\ell+1} \right) (-h(\ell-1, \ell)) + \frac{f(\ell-1)}{\ell+1} \Delta^{(1)}h(\ell-2, \ell) \\ &= \frac{f(\ell) - \ell}{\ell+1} h(\ell-1, \ell) + \frac{f(\ell-1)}{\ell+1} \Delta^{(1)}h(\ell-2, \ell),\end{aligned}$$

which is nonnegative by induction hypothesis and since  $f(\ell) \geq \ell$ .

(iii) Again we show by induction on  $\ell$  that there exists  $I(\ell) \in \{0, \dots, \ell-1\}$  such that (26) is valid. Note for  $\ell = 2$  the statement holds trivially. Moreover, for  $\ell \leq k_*$  the statement holds by (ii) with  $I(\ell) = \ell-1$ .

Suppose the statement (26) is true for some  $\ell \geq k_*$ . By (28) we obtain

$$\Delta^{(1)}h(k, \ell+1) = \left( \frac{\ell}{\ell+1} - \frac{f(k+1)}{\ell+1} \right) \Delta^{(1)}h(k, \ell) + \frac{f(k)}{\ell+1} \Delta^{(1)}h(k-1, \ell).$$

From this we can deduce that if  $k < I(\ell)$ , then since  $I(\ell) \leq \ell-1$ , we have  $f(k+1) \leq k+2 \leq \ell$ , so that  $\Delta^{(1)}h(k, \ell+1) \geq 0$ . Conversely, if  $k > I(\ell)$  and  $k \leq \ell-2$ , we can deduce similarly that  $\Delta^{(1)}h(k, \ell+1) \leq 0$ . It remains to show that  $\Delta^{(1)}h(k, \ell+1) \leq 0$  for  $k = \ell, \ell-1$  if  $\ell-1 > I(\ell)$ . Note the case  $k = \ell$  holds since  $\Delta^{(1)}h(\ell, \ell+1) = -h(\ell, \ell+1)$ . By the recursion (27) together with (30), we get

$$h(\ell, \ell+1) = \frac{f(\ell)}{\ell+1} h(\ell-1, \ell) = f(\ell) \prod_{i=1}^{\ell} \frac{f(i-1)}{i+1}. \quad (31)$$

Now, since  $\Delta^{(1)}h(\ell-2, \ell) \leq 0$ , we have again by (27),

$$\begin{aligned}\Delta^{(1)}h(\ell-1, \ell+1) &= h(\ell, \ell+1) - \left( \frac{\ell}{\ell+1} - \frac{f(\ell-1)}{\ell+1} \right) h(\ell-1, \ell) - \frac{f(\ell-1)}{\ell+1} h(\ell-2, \ell) \\ &\leq h(\ell, \ell+1) - \frac{\ell}{\ell+1} h(\ell-1, \ell) \\ &= f(\ell) \prod_{j=1}^{\ell} \frac{f(j-1)}{j+1} - \frac{\ell}{\ell+1} f(\ell-1) \prod_{j=1}^{\ell-1} \frac{f(j-1)}{j+1} \\ &= \frac{f(\ell-1)}{\ell+1} (f(\ell) - \ell) \prod_{j=1}^{\ell-1} \frac{f(j-1)}{j+1},\end{aligned}$$

which is negative as  $f(\ell) \leq \ell$  since  $\ell \geq k_*$ . In particular, we have seen that  $I(\ell+1) \in \{I(\ell), I(\ell)+1\}$ .

(iv) Define

$$C := f(0) \left( 1 \vee \max_{1 \leq k \leq k_*} \prod_{i=1}^k \frac{f(i)}{i} \right).$$

Then, we will show inductively that for all  $\ell \in \mathbb{N}$ :

$$h(k, \ell) \leq \frac{C}{\ell}, \quad \text{for all } k \leq \ell - 1. \quad (32)$$

For  $\ell = 1$ , we have

$$h(0, 1) = f(0)\mathbb{P}(X_1 = 0) = f(0) \leq C.$$

Now, assume that (32) holds for some  $\ell \in \mathbb{N}$ . Using the identity (27) we obtain for  $k \leq \ell - 1$

$$\begin{aligned} h(k, \ell + 1) &= \left( \frac{\ell}{\ell + 1} - \frac{f(k)}{\ell + 1} \right) h(k, \ell) + \frac{f(k)}{\ell + 1} h(k - 1, \ell) \\ &\leq \left( \frac{\ell}{\ell + 1} - \frac{f(k)}{\ell + 1} \right) \frac{C}{\ell} + \frac{f(k)}{\ell + 1} \frac{C}{\ell} = \frac{C}{\ell + 1}, \end{aligned}$$

where we used that  $f(k) \leq k + 1 \leq \ell$  in the second step. For  $k = \ell$ , we have by (31)

$$h(\ell, \ell + 1) = f(\ell) \prod_{i=1}^{\ell} \frac{f(i-1)}{i+1} = \frac{f(0)}{\ell + 1} \prod_{i=1}^{\ell} \frac{f(i)}{i}.$$

Then, if  $\ell \leq k^*$ , this is trivially bounded by  $C/(\ell + 1)$ . Furthermore, if  $\ell > k^*$ , then

$$h(\ell, \ell + 1) = \frac{f(0)}{\ell + 1} \prod_{i=1}^{\ell} \frac{f(i)}{i} \leq \frac{C}{\ell + 1} \prod_{i=k^*+1}^{\ell} \frac{f(i)}{i} \leq \frac{C}{\ell + 1},$$

since  $f(i) \leq i$  for all  $i \geq k^*$ . This completes the induction step.

(v) Note that by (ii),  $k \mapsto h(k, \ell)$  is increasing for  $k \leq \ell - 1$ . In particular, we have by (i)

$$\sup_{k \in \mathbb{N}_0} h(k, \ell) = \sup_{k \leq \ell - 1} h(k, \ell) = h(\ell - 1, \ell).$$

By (31), we get that

$$h(\ell - 1, \ell) = f(\ell - 1) \prod_{i=1}^{\ell-1} \frac{f(i-1)}{i+1} \leq \frac{\prod_{i=0}^{\ell-1} (i + \gamma)}{\ell!} = \frac{1}{\Gamma(\gamma)} \frac{\Gamma(\ell + \gamma)}{\Gamma(\ell + 1)} \sim \frac{1}{\Gamma(\gamma)} \ell^{\gamma-1},$$

using the asymptotics for the Gamma function. This immediately gives statement (v).  $\square$

## 2.5 Proofs of Theorems 1.4 and 1.5

We can combine our previous estimates to prove the two main theorems simultaneously.

*Proofs of Theorems 1.4 and 1.5.* We first consider the case that the preferential attachment model satisfies Assumption (A) with  $d_0 = 0$  so that we can generate the indegree of a uniform vertex using the Markov chain  $(X_\ell)_{\ell \geq 1}$  defined in Lemma 2.5. Using the notation  $h(k, \ell) = f(k)\mathbb{P}(X_\ell = k) - \mathbb{P}(X_\ell \geq k+1)$ , we have from Lemma 2.6, for any  $A \subset \mathbb{N}_0$ ,

$$\mathbb{E}[\mathcal{A}g_A(X_{n+1})] = \frac{1}{n+1} \left( \left( \sum_{\ell=1}^n \sum_{k=0}^{\ell-1} \Delta v_A(k) h(k, \ell) \right) + v_A(0) \right).$$

Using a discrete integration by parts formula and the fact that  $h(\ell, \ell) = 0$ , we can rewrite the inner sum as

$$\sum_{k=0}^{\ell-1} \Delta v_A(k) h(k, \ell) = -v_A(0)h(0, \ell) - \sum_{k=0}^{\ell-1} v_A(k+1) \Delta^{(1)} h(k, \ell).$$

Under the assumptions of Theorem 1.4 we can apply Proposition 2.7(iii) and for Theorem 1.5 part (ii) of Proposition 2.7 to deduce that in both cases there exist  $I(\ell)$  such that  $\Delta^{(1)} h(k, \ell) \geq 0$  for  $k < I(\ell)$  and  $\Delta^{(1)} h(k, \ell) \leq 0$  for  $k \geq I(\ell)$ . In particular, we have that

$$\begin{aligned} & \left| \sum_{k=0}^{\ell-1} v_A(k+1) \Delta^{(1)} h(k, \ell) \right| \\ & \leq \sup_k |v_A(k)| \left( h(I(\ell), \ell) - h(0, \ell) + h(I(\ell), \ell) - h(\ell, \ell) \right) \\ & \leq 2 \sup_k |v_A(k)| \sup_{k \leq \ell-1} h(k, \ell). \end{aligned}$$

Therefore, we get

$$\begin{aligned} & |\mathbb{E}[\mathcal{A}g_A(X_{n+1})]| \\ & \leq \frac{|v_A(0)|}{n+1} + \left| \frac{1}{n+1} \sum_{\ell=1}^n \left( \left( \sum_{k=0}^{\ell-1} v_A(k+1) \Delta^{(1)} h(k, \ell) \right) + v_A(0)h(0, \ell) \right) \right| \\ & \leq \frac{|v_A(0)|}{n+1} + \frac{2}{n+1} \sup_k |v_A(k)| \sum_{\ell=1}^n \sup_{k \leq \ell-1} h(k, \ell). \end{aligned}$$

Hence, if we combine this estimate with Proposition 2.4 we obtain that

$$d_{\text{TV}}(W_{n+1}, W) = \sup_{A \subset \mathbb{N}_0} |\mathbb{E}[\mathcal{A}g_A(X_{n+1})]| \leq \frac{1}{n+1} + \frac{2}{n+1} \sum_{\ell=1}^n \sup_{k \leq \ell-1} h(k, \ell).$$

Finally, we note that in the case of Theorem 1.4, we can apply Proposition 2.7 (iv) to deduce that in this case, there exists a constant  $C > 0$  such that

$$d_{\text{TV}}(W_{n+1}, W) \leq \frac{1}{n+1} + \frac{2C}{n+1} \sum_{\ell=1}^n \frac{1}{\ell},$$

which immediately gives the required bound. In the case of Theorem 1.5, we can instead apply Proposition 2.7 (v) to get a constant  $C > 0$  such that

$$d_{\text{TV}}(W_{n+1}, W) \leq \frac{1}{n+1} + \frac{2C}{n+1} \sum_{\ell=1}^n \ell^{-(1-\gamma)},$$

which again yields the statement of the theorem.

Finally, we consider the case when the model satisfies Assumptions (A) with  $d_0 > 0$ . In this case, by the same argument as in Lemma 2.5, the indegree of a uniformly chosen vertex has the same distribution as a Markov chain  $(\tilde{X}_n)_{n \geq 1}$  with  $\tilde{X}_1 = d_0$ , but the same transition probabilities as  $(X_n)_{n \geq 1}$ . Let  $\tau = \inf\{k \geq 2 : \tilde{X}_k = 0\}$ . We can couple  $(X_n), (\tilde{X}_n)$  by first letting  $\tilde{X}_n$  evolve and then letting  $X_n$  evolve independently until time  $\tau$ . Further, we set  $X_k := \tilde{X}_k$  for all  $k \geq \tau$ . By the characterization of  $d_{\text{TV}}$  in terms of couplings, we thus have

$$d_{\text{TV}}(X_n, \tilde{X}_n) \leq \mathbb{P}(\tau > n) = \prod_{i=1}^n \left(1 - \frac{1}{i+1}\right) = \frac{1}{n+1}.$$

By the first part of the proof, this completes the proof also for  $d_0 > 0$ .  $\square$

### 3 Rate of convergence for the outdegree

We now consider the model introduced in Example 1.1, where connections to old vertices are made independently, and prove Theorem 1.6, i.e. show the rate of convergence for the outdegree  $D_n = \deg_n^+(n)$  of vertex  $n$ .

We will need the following moment bound.

**Lemma 3.1.** *For the preferential attachment model as in Example 1.1 with  $f(k) \leq \gamma k + 1$  for all  $k$ , and some  $\gamma \in (0, 1)$ , we have for all  $n \in \mathbb{N}$ ,*

$$\mathbb{E}[f(\deg_n^-(i))] \leq \left(\frac{n}{i}\right)^\gamma, \quad \text{for all } i \in [n].$$

*Proof.* For the linear attachment rule  $f^{(\gamma)} := \gamma \cdot +1$ , the statement of the lemma is proved in [DM13, Lemma 2.7]. Denote the degrees in the corresponding preferential attachment model by  $\deg_n^{-,\gamma}(i)$  and consider general  $f$  and associated degrees  $\deg_n^-(i)$ . Then, since  $f \leq f^{(\gamma)}$  we can couple the models so that  $\deg_n^-(i) \leq \deg_n^{-,\gamma}(i)$  for all  $i \in [n], n \in \mathbb{N}$ . In particular, we have that

$$\mathbb{E}[f(\deg_n^-(i))] \leq \mathbb{E}[f^{(\gamma)}(\deg_n^-(i))] \leq \mathbb{E}[f^{(\gamma)}(\deg_n^{-,\gamma}(i))] \leq \left(\frac{n}{i}\right)^\gamma,$$

as required.  $\square$

Using a result of [BH84] for Poisson approximation (again based on the Chen-Stein method), we can now prove Theorem 1.6.



*Proof of Theorem 1.6.* By the independence assumption for incoming edges, it follows that the indegree evolutions  $(\deg_k^-(i))_{k \geq i}$  and  $(\deg_k^-(j))_{k \geq j}$  are independent if  $i \neq j$ . In particular, if we write  $X_{i,n} = \mathbb{1}\{\text{there is an edge from } n \text{ to } i\} = \deg_n^-(i) - \deg_{n-1}^-(i)$ , then we can write the outdegree  $D_n$  of vertex  $n$  as

$$D_n = \sum_{i=1}^{n-1} X_{i,n},$$

i.e. as the sum of independent Bernoulli variables. Note that

$$p_{i,n} := \mathbb{P}(X_{i,n} = 1) = \mathbb{E} \left[ \mathbb{E} \left[ \deg_n^-(i) - \deg_{n-1}^-(i) \mid \mathcal{G}_{n-1} \right] \right] = \mathbb{E} \left[ \frac{f(\deg_{n-1}^-(i))}{n-1} \right].$$

Therefore,

$$\lambda_n := \mathbb{E}[D_n] = \mathbb{E} \left[ \frac{1}{n-1} \sum_{i=1}^{n-1} f(\deg_{n-1}^-(i)) \right] = \mathbb{E}[f(W_{n-1})],$$

where  $W_{n-1}$  denotes the indegree of a uniformly chosen vertex after the insertion of vertex  $n-1$ . From the proof of Theorem 1.1 (b) in [DM09] we know that  $\lambda_n \rightarrow \mathbb{E}[f(W)]$  if  $W \sim \mu$ . Applying [BH84, Thm. 1.1] we obtain that

$$d_{\text{TV}}(D_n, \text{Po}(\lambda_n)) \leq \frac{1 - e^{-\lambda_n}}{\lambda_n} \sum_{i=1}^{n-1} p_{i,n}^2 \leq \min\left\{1, \frac{1}{\lambda_n}\right\} \sum_{i=1}^{n-1} p_{i,n}^2. \quad (33)$$

It remains to control the sum on the right hand side. By Lemma 3.1, we have that

$$\sum_{i=1}^{n-1} p_{i,n}^2 = \frac{1}{(n-1)^2} \sum_{i=1}^{n-1} \mathbb{E}[f(\deg_{n-1}^-(i))]^2 \leq \frac{1}{(n-1)^2} \sum_{i=1}^{n-1} \left(\frac{n}{i}\right)^{2\gamma}$$

Since  $\lambda_n \rightarrow \lambda := \mathbb{E}[f(W)]$  we can deduce from (33) that

$$d_{\text{TV}}(D_n, \text{Po}(\lambda_n)) \leq C \begin{cases} \frac{1}{n+1}, & \text{for } 0 < \gamma < \frac{1}{2}, \\ \frac{\log(n)}{n}, & \text{for } \gamma = \frac{1}{2}, \\ n^{-2(1-\gamma)}, & \text{for } \frac{1}{2} < \gamma < 1, \end{cases},$$

for a suitable constant  $C > 0$ , which proves the first part of Theorem 1.6.

For the final part of Theorem 1.6, we assume that  $f(k) = \gamma k + \beta$ , for  $\gamma \in (0, 1)$ ,  $\beta \in [0, 1]$ . First note that in this case by (14)

$$\begin{aligned} \lambda &= \mathbb{E}[f(W)] = \gamma \mathbb{E}[W] + \beta = \gamma \sum_{k \geq 1} \mu([k, \infty)) + \beta \\ &= \gamma \sum_{k \geq 1} f(k-1) \mu_{k-1} + \beta = \gamma \lambda + \beta. \end{aligned}$$

In particular,  $\lambda = \frac{\beta}{1-\gamma}$ . Following a similar argument as in the proof of Theorem 1.1 (b) in [DM09], we have that

$$\begin{aligned}
\mathbb{E}[f(W_{n+1})] &= \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbb{E} \left[ \mathbb{E}[f(\deg_{n+1}^-(i)) \mid \mathcal{G}_n] \right] \\
&= \frac{1}{n+1} \sum_{i=1}^n \mathbb{E} \left[ \mathbb{E}[f(\deg_{n+1}^-(i)) - f(\deg_n^-(i)) \mid \mathcal{G}_n] \right] \\
&\quad + \frac{f(0)}{n+1} + \frac{1}{n+1} \sum_{i=1}^n \mathbb{E} \left[ \mathbb{E}[f(\deg_n^-(i)) \mid \mathcal{G}_n] \right] \\
&= \frac{1}{n+1} \left( \sum_{i=1}^n \mathbb{E} \left[ \mathbb{E}[\gamma(\deg_{n+1}^-(i) - \deg_n^-(i)) \mid \mathcal{G}_n] \right] + \beta + \sum_{i=1}^n \mathbb{E}[f(\deg_n^-(i))] \right) \\
&= \frac{1}{n+1} \sum_{i=1}^n \gamma \mathbb{E} \left[ \frac{f(\deg_n^-(i))}{n} \right] + \frac{\beta}{n+1} + \frac{1}{n+1} \sum_{i=1}^n \mathbb{E}[f(\deg_n^-(i))] \\
&= \left(1 - \frac{1-\gamma}{n+1}\right) \mathbb{E}[f(W_n)] + \frac{\beta}{n+1}.
\end{aligned}$$

Using that  $\lambda = \frac{\beta}{1-\gamma}$ , we obtain that for  $\bar{\lambda}_{n+1} := \mathbb{E}[f(W_n)] - \lambda$ ,

$$\bar{\lambda}_{n+1} = \left(1 - \frac{1-\gamma}{n}\right) \bar{\lambda}_n.$$

Hence,

$$|\bar{\lambda}_{n+1}| = \prod_{i=1}^n \left(1 - \frac{1-\gamma}{i}\right) |\bar{\lambda}_1| \leq C n^{-(1-\gamma)},$$

for a suitable constant  $C > 0$ , as claimed.  $\square$

**Acknowledgements:** We would like to thank Christian Döbler for helpful discussions. The authors gratefully acknowledge support from the DFG in form of the Research Training Group 1916 *Combinatorial Structures in Geometry*.

## References

- [ACJP08] A. Aiello, W. Bonato, C. Cooper, J. Janssen, and P. Prałat. A spatial web graph model with local influence regions. *Internet Math.*, 5(1-2):175–196, 2008.
- [BA99] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [Bar88] A. D. Barbour. Stein’s method and Poisson process convergence. *Journal of Applied Probability*, Special Vol. 25A:175–184, 1988.

- [BH84] A. D. Barbour and P. Hall. On the rate of poisson convergence. *Mathematical Proceedings of the Cambridge Philosophical Society*, 95(3):473480, 1984.
- [Bha07] S. Bhamidi. Universal techniques to analyze preferential attachment trees: Global and local analysis. *Preprint available at <http://www.unc.edu/~bhamidi/>*, 2007.
- [BL12] G. Brightwell and M. Luczak. Vertices of high degree in the preferential attachment tree. *Electronic Journal of Probability*, 17:no. 14, 2012.
- [BRST01] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, 18(3):279–290, 2001.
- [BX01] T. C. Brown and A. Xia. Stein’s method and birth-death processes. *The Annals of Probability*, 29(3):1373–1403, 2001.
- [Che75] L. H. Y. Chen. Poisson approximation for dependent trials. *Ann. Probab.*, 3(3):534–545, 1975.
- [DM09] S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: degree evolutions. *Electron. J. Probab.*, 14:no. 43, 1222–1267, 2009.
- [DM13] S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: the giant component. *Ann. Probab.*, 41(1):329–384, 2013.
- [For09] E. Ford. *Barabási-Albert random graphs, scale-free distributions and bounds for approximation through Stein’s method*. PhD thesis, Oxford, 2009.
- [Hof17] R. van der Hofstad. *Random graphs and complex networks. Vol. 1*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017.
- [Nor98] J. R. Norris. *Markov chains*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, 1998.
- [PRR13a] E. A. Peköz, A. Röllin, and N. Ross. Degree asymptotics with rates for preferential attachment random graphs. *Ann. Appl. Probab.*, 23(3):1188–1218, 06 2013.
- [PRR13b] E. A. Peköz, A. Röllin, and N. Ross. Total variation error bounds for geometric approximation. *Bernoulli*, 19(2):610–632, 05 2013.

- [PRR17] E. Peköz, A. Röllin, and N. Ross. Joint degree distributions of preferential attachment random graphs. *Advances in Applied Probability*, 49(2):368387, 2017.
- [Ros11] N. Ross. Fundamentals of Stein’s method. *Probability Surveys*, 8:210–293, 2011.
- [Ros13] N. Ross. Power laws in preferential attachment graphs and Stein’s method for the negative binomial distribution. *Advances in Applied Probability*, 45(3):876–893, 2013.
- [RTV07] A. Rudas, B. Tóth, and B. Valkó. Random trees and general branching processes. *Random Structures & Algorithms*, 31(2):186–202, 2007.
- [Ste72] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. II: Probability theory*, pages 583–602, 1972.