POWER LAWS IN PREFERENTIAL ATTACHMENT GRAPHS AND STEIN'S METHOD FOR THE NEGATIVE BINOMIAL DISTRIBUTION

NATHAN ROSS,* *University of California, Berkeley*

Abstract

For a family of linear preferential attachment graphs, we provide rates of convergence for the total variation distance between the degree of a randomly chosen vertex and an appropriate power law distribution as the number of vertices tends to ∞. Our proof uses a new formulation of Stein's method for the negative binomial distribution, which stems from a distributional transformation that has the negative binomial distributions as the only fixed points.

Keywords: Stein's method; negative binomial distribution; random graph; preferential attachment; distributional transformation; power law

2010 Mathematics Subject Classification: Primary 05C80
Secondary 60F05; 60C05

1. Introduction

Preferential attachment random graphs were introduced in [2] as a stochastic mechanism to explain power law degree distributions empirically observed in real-world networks such as the World Wide Web. These graphs evolve by sequentially adding vertices and edges in a random way so that connections to vertices with high degree are favored. There has been much interest in properties of these models and their many embellishments; the text [21] is probably the best survey of this vast literature. Like the seminal work [2] (and the mathematically precise formulation [4]), much of this research is devoted to showing that if the number of vertices of the graph is large, then the proportion of vertices having degree k approximately decays as $c_\gamma k^{-\gamma}$ for some constant c_γ and $\gamma > 1$; the so-called power law behavior.

Our main result in this vein is Theorem 1.1 below, which, for a family of linear preferential attachment graphs, provides rates of convergence for the total variation distance between the degree of a randomly chosen vertex and an appropriate power law distribution as the number of vertices tends to ∞. The result is new and the method of proof is also of interest since it differs substantially from proofs of similar results (see, e.g. Section 8.5 of [21]). Our proof of Theorem 1.1 uses a new formulation of Stein's method for the negative binomial distribution, Theorem 1.2 below (see [19] and the references therein for a basic introduction to Stein's method). The result stems from a distributional transformation that has negative binomial distributions as the only fixed points (we will shortly see the relationship between the negative binomial distribution and power laws). Similar strategies have recently found success in analyzing degree distributions in preferential attachment models; see [17] and Section 6 of [16]—the latter is a special case of our results and is the template for our proofs. The remainder of the introduction is devoted to stating our results in greater detail.

First we define the family of preferential attachment models we study; these are the same models studied in Chapter 8 of [21], which are a generalization of the models first defined in [4], which in turn are a formalization of the heuristic models described in [2]. The family of models is parameterized by $m \in \mathbb{N}$ and $\delta > -m$. For $m = 1$ and given δ, the model starts with one vertex with a single loop where one end of the loop contributes to the 'in-degree' and the other to the 'out-degree'. Now, for $2 \leq k \leq n$, given the graph with $k-1$ vertices, add vertex k along with an edge emanating 'out' from k 'in' to a random vertex chosen from the set $\{1, \ldots, k\}$ with probability proportional to the total degree of that vertex plus δ, where initially vertex k has degree 1. That is, at step k, the chance that vertex k connects to itself is $(\delta + 1)/(k(2 + \delta) - 1)$. After n steps of this process, we denote the resulting random graph by $G_n^{1,\delta}$.

For $m > 1$, we define $G_n^{m,\delta}$ by first generating $G_{nm}^{1,\delta/m}$, and then 'collapsing' consecutive vertices into groups of size m, starting from the first vertex, and retaining all edges. Note that with this setup, it is possible for a vertex to connect to itself or other vertices more than once and as many as m times (in fact, the first vertex always consists of m loops) and all of these connections contribute to the in- and out-degrees of a vertex (e.g. the first vertex has both in- and out-degrees m).

Here and below, we think of δ and m as fixed and let W_n be the in-degree of a randomly chosen vertex from $G_n^{m,\delta}$. We provide a bound on the total variation distance between W_n and a limiting distribution which is a mixture of negative binomial distributions. For $r > 0$ and $0 < p \leq 1$, we write $X \sim \text{NB}(r, p)$ if

$$\mathbb{P}(X = k) = \frac{\Gamma(r + k)}{k! \, \Gamma(r)} (1 - p)^k p^r, \qquad k = 0, 1, \ldots.$$

Definition 1.1. For $m \in \mathbb{N}$, $\delta > -m$, and U uniform on $(0, 1)$, denote the mixture distribution $\text{NB}(m + \delta, U^{1/(2+\delta/m)})$ by $K(m, \delta)$.

For our main result, we define the total variation distance between two nonnegative integer-valued random variables X and Y as

$$d_{\text{TV}}(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{A \subseteq \mathbb{Z}_+} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \tag{1.1}$$

$$= \frac{1}{2} \sum_{k \in \mathbb{Z}_+} |\mathbb{P}(X = k) - \mathbb{P}(Y = k)|, \tag{1.2}$$

where here and below $\mathbb{Z}_+ = \{0, 1, \ldots\}$.

Theorem 1.1. *If W_n is the in-degree of a randomly chosen vertex from the preferential attachment graph $G_n^{m,\delta}$ and $K(m, \delta)$ is the mixed negative binomial distribution of Definition 1.1, then, for some constant $C_{m,\delta}$,*

$$d_{\text{TV}}(\mathcal{L}(W_n), K(m, \delta)) \leq C_{m,\delta} \frac{\log(n)}{n}.$$

To see the power law behavior of $K(m, \delta)$, we record the following easy result which is a more standard representation of $K(m, \delta)$ through its point probabilities. The proof follows from direct computation and then Stirling's formula (or Lemma 3.1 below). These formulae with additional discussion are also found in Section 8.3 of [21], specifically Equations (8.3.2) and (8.3.9-10), and were first discovered in [7] and [11]. The representation of $K(m, \delta)$ as a mixture of negative binomial distributions does not seem to be well known.

Lemma 1.1. *If $m \in \mathbb{N}$, $\delta > -m$, and $Z \sim K(m, \delta)$, then, for $l = 0, 1, \ldots,$*

$$\mathbb{P}(Z = l) = \left(2 + \frac{\delta}{m}\right) \frac{\Gamma(l + m + \delta)\Gamma(m + 2 + \delta + \delta/m)}{\Gamma(m + \delta)\Gamma(l + m + 3 + \delta + \delta/m)},$$

and, for $c_{m,\delta} = (2 + \delta/m)\Gamma(m + 2 + \delta + \delta/m)/\Gamma(m + \delta)$,

$$\mathbb{P}(Z = k) \asymp \frac{c_{m,\delta}}{k^{3+\delta/m}} \quad as\ k \to \infty.$$

Before discussing our Stein's method result, we make a few final remarks. The usual mathematical statement implying power law behavior of the degrees of a random graph in this setting is that the empirical degree distribution converges to $K(m, \delta)$ in probability (see Theorem 8.2 of [21]). Such a result implies that the total variation distance in Theorem 1.1 tends to 0 (see Exercise 8.14 of [21]), but does not provide a rate. Another result similar to Theorem 1.1 is Proposition 8.4 of [21] which states that, for $Z \sim K(m, \delta)$,

$$|\mathbb{P}(W_n = k) - \mathbb{P}(Z = k)| \leq \frac{C}{n},$$

which according to (1.2) neither implies nor is implied by Theorem 1.1. Finally, regarding other preferential attachment models, our results can likely be extended to some other models where the limiting distribution is $K(m, \delta)$, for example, where the update rule is that we consider here, but the starting graph is not. For other preferential attachment graphs where the limiting degree distribution is not $K(m, \delta)$ (such as those of [20]), it may be possible to prove analogs of Theorem 1.1 using methods similar to ours, but we do not pursue this here.

To state our general result which we use to prove Theorem 1.1, we first define a distributional transformation. For $r > 0$ and $n \geq 1$, let $U_{r,n}$ be a random variable having the distribution of the number of white balls drawn in $n - 1$ draws in a standard Pólya urn scheme starting with r white balls and 1 black ball. That is, for fixed r, we construct $U_{r,n}$ sequentially by setting $U_{r,1} = 0$, and, for $k \geq 1$,

$$\mathbb{P}(U_{r,k+1} = U_{r,k} + 1 \mid U_{r,k}) = 1 - \mathbb{P}(U_{r,k+1} = U_{r,k} \mid U_{r,k}) = \frac{r + U_{r,k}}{r + k}. \qquad (1.3)$$

Also, for a nonnegative integer-valued random variable X with finite mean, we say that X^s has the *size bias* distribution of X if

$$\mathbb{P}(X^s = k) = \frac{k\mathbb{P}(X = k)}{\mathbb{E}X}, \qquad k = 1, 2, \ldots.$$

Definition 1.2. Let X be a nonnegative integer-valued random variable with finite mean, and let X^s denote a random variable having the size bias distribution of X. We say that the random variable X^{*r} has the *r-equilibrium* transformation if

$$X^{*r} \stackrel{\mathrm{D}}{=} U_{r,X^s},$$

where we understand U_{r,X^s} to mean $\mathcal{L}(U_{r,X^s} \mid X^s = k) = \mathcal{L}(U_{r,k})$.

As we will see below in Corollary 2.1, $X^{*r} \stackrel{\mathrm{D}}{=} X$ if and only if $X \sim \mathrm{NB}(r, p)$ for some $0 < p < 1$. Thus, if some nonnegative integer-valued random variable W has approximately the same distribution as W^{*r}, it is plausible that W is approximately distributed as a negative binomial distribution. The next result makes this heuristic precise. Here and below we denote the indicator of an event B by $\mathbb{I}_B$ or $\mathbb{I}[B]$.

Theorem 1.2. *Let W be a nonnegative integer-valued random variable with $\mathbb{E}W = \mu$. Also, let $r > 0$ and W^{*r} be coupled to W and have the r-equilibrium transformation of Definition 1.2. If $p = r/(r + \mu)$ and $c_{r,p} = \min\{(r + 2)(1 - p), 2 - p\} \le 2$, then, for an event B,*

$$d_{\text{TV}}(\mathcal{L}(W), \text{NB}(r, p)) \le c_{r,p}\mathbb{E}[\mathbb{I}_B|W^{*r} - W|] + 2(\text{e}\max\{1, r\} + 1)\mathbb{P}(B^c)$$
$$\le 2(\text{e}\max\{1, r\} + 1)\mathbb{P}(W^{*r} \ne W).$$

Remark 1.1. Analogs of Theorem 1.2 for other distributions which use fixed points of distributional transformations are now well established in the Stein method literature. For example, the book [3] develops Stein's method for the Poisson approximation using the fact that a nonnegative integer-valued random variable X with finite mean has the Poisson distribution if and only if $X \overset{\text{D}}{=} X^s - 1$. Also, there is the zero bias transformation for the normal distribution [9], the equilibrium transformation for the exponential distribution [15], a less standard distribution [17], and the special case where $r = 1$ above, the discrete equilibrium transformation for the geometric distribution [16] (see also [14] for an unrelated transformation used for the geometric approximation).

Remark 1.2. The fact that negative binomial distributions are the fixed points of the r-equilibrium transformation is the discrete analog of the fact, perhaps more familiar, that a nonnegative random variable X has the gamma distribution with shape parameter α if and only if

$$X \overset{\text{D}}{=} B_{\alpha,1}X^s,$$

where $B_{\alpha,1}$ is a beta variable with density $\alpha x^{\alpha-1}$ for $0 < x < 1$ independent of X^s; see [18].

The layout of the remainder of the article is as follows. In Section 2 we develop Stein's method for the negative binomial distribution using the r-equilibrium transformation and prove Theorem 1.2. In Section 3 we use Theorem 1.2 to prove Theorem 1.1.

2. Negative binomial approximation

The proof of Theorem 1.2 roughly follows the usual development of Stein's method of distributional approximation using fixed points of distributional transformations (see the references of Remark 1.1). Specifically, if W is a nonnegative integer-valued random variable of interest and Y has the negative binomial distribution, then, using definition (1.1), we want to bound $|\mathbb{P}(W \in A) - \mathbb{P}(Y \in A)|$ uniformly for $A \subseteq \mathbb{Z}_+$. Typically, this program has three components.

1. Define a *characterizing operator* $\mathcal{A}$ for the negative binomial distribution which has the property that

$$\mathbb{E}\mathcal{A}g(Y) = 0$$

   for all g in a large enough class of functions if and only if $\mathcal{L}(Y) \sim \text{NB}(r, p)$.

2. For $A \subseteq \mathbb{Z}_+$, define g_A to solve

$$\mathcal{A}g_A(k) = \mathbb{I}[k \in A] - \mathbb{P}(Y \in A). \tag{2.1}$$

3. Using (2.1), note that

$$|\mathbb{P}(W \in A) - \mathbb{P}(Y \in A)| = |\mathbb{E}\mathcal{A}g_A(W)|.$$

   Now use properties of the solutions g_A and the distributional transformation to bound the right-hand side of this equation.

Obviously, there must be some relationship between the characterizing operator of item 1 and the distributional transformation of item 3; this is typically the subtle part of the program above. For item 1, we use the characterizing operator for the negative binomial distribution as defined in [5].

Theorem 2.1. ([5].) *If $W \geq 0$ has a finite mean then $W \sim \mathrm{NB}(r, p)$ if and only if*

$$\mathbb{E}[(1 - p)(r + W)g(W + 1) - Wg(W)] = 0 \tag{2.2}$$

for all bounded functions g.

We need to develop the connection between the characterizing operator of Theorem 2.1 and the r-equilibrium transformation. To this end, for a function g, define

$$D^{(r)}g(k) = \left(\frac{k}{r} + 1\right)g(k + 1) - \frac{k}{r}g(k),$$

and note that the negative binomial characterizing operator of (2.2) can be written as

$$r(1 - p)D^{(r)}g(W) - pWg(W). \tag{2.3}$$

The key relationship is the following.

Lemma 2.1. *If the integer-valued random variable $X \geq 0$ has finite mean $\mu > 0$, X^{*r} has the r-equilibrium distribution of X, and g is a function such that the expectations below are well defined, then*

$$\mu \mathbb{E}D^{(r)}g(X^{*r}) = \mathbb{E}Xg(X).$$

Proof. We show that

$$\mathbb{E}D^{(r)}g(U_{r,n}) = g(n), \tag{2.4}$$

which, using the definition of the size bias distribution implies that

$$\mu \mathbb{E}D^{(r)}g(X^{*r}) = \mu \mathbb{E}g(X^s) = \mathbb{E}Xg(X),$$

as desired. To show (2.4), we use induction on n. The equality is obvious for $n = 1$ since $U_{r,1} = 0$. We assume that (2.4) holds for n and show that it holds for $n + 1$. By conditioning on the previous step in the urn process defining $U_{r,n+1}$ and using (1.3), we find that, for a function f such that the expectations below are well defined,

$$\mathbb{E}f(U_{r,n+1}) = \frac{1}{r + n}\mathbb{E}(U_{r,n} + r)f(U_{r,n} + 1) + \mathbb{E}\left(1 - \frac{U_{r,n} + r}{r + n}\right)f(U_{r,n}).$$

Combining this equality with the induction hypothesis in the form

$$\mathbb{E}(U_{r,n} + r)f(U_{r,n} + 1) = rf(n) + \mathbb{E}U_{r,n}f(U_{r,n})$$

yields

$$\mathbb{E}f(U_{r,n+1}) = \frac{r}{r + n}f(n) + \frac{n}{r + n}\mathbb{E}f(U_{r,n}).$$

Now taking $f = D^{(r)}g$ and using the induction hypothesis again yields (2.4). ∎

We now record the following result which, while not necessary for the proof of Theorem 1.2, underlies our whole approach for the negative binomial approximation.

Corollary 2.1. *If the integer-valued random variable $X \geq 0$ is such that $\mathbb{E}X = r(1 - p)/p$ for some $0 < p < 1$, then $X \sim \mathrm{NB}(r, p)$ if and only if*

$$X \overset{\mathrm{D}}{=} X^{*r}.$$

Proof. If $X \overset{\mathrm{D}}{=} X^{*r}$ then combining Theorem 2.1 and Lemma 2.1, noting representation (2.3), we easily see that $X \sim \mathrm{NB}(r, p)$.

Conversely, we assume that $Y \sim \mathrm{NB}(r, p)$, and show that $Y^{*r} \overset{\mathrm{D}}{=} Y$ using the method of moments. According to (4.3) of [10, p. 178],

$$\mathbb{E}[U_{r,n}(U_{r,n} - 1) \cdots (U_{r,n} - k + 1)] = \frac{r(n - 1) \cdots (n - k)}{r + k},$$

which implies that, for X with finite $k + 1$ moments and $\mathbb{E}X = r(1 - p)/p$,

$$\mathbb{E}[X^{*r} \cdots (X^{*r} - k + 1)] = \frac{r\mathbb{E}[(X^s - 1) \cdots (X^s - k)]}{r + k} = \frac{p}{1 - p} \frac{\mathbb{E}[X(X - 1) \cdots (X - k)]}{r + k}.$$

Now, from Equation (2.29) of [10, p. 84], if $Y \sim \mathrm{NB}(r, p)$ then

$$\mathbb{E}[Y \cdots (Y - k + 1)] = r \cdots (r + k - 1)\left(\frac{1 - p}{p}\right)^k.$$

Combining this with the calculation above, we find that, for all $k \geq 1$,

$$\mathbb{E}[Y^{*r} \cdots (Y^{*r} - k + 1)] = \mathbb{E}[Y \cdots (Y - k + 1)].$$

Since Y has a well-behaved moment generating function (i.e. exists in a neighborhood around 0), the moment sequence determines the distribution and so $Y \overset{\mathrm{D}}{=} Y^{*r}$, as desired.

The next two lemmas take care of item 2 in the program outlined above, and obtain the properties of the solution for item 3. We prove Theorem 1.2 immediately after the lemmas. For a function $g : \mathbb{Z}_+ \to \mathbb{R}$, define $\Delta g(k) = g(k + 1) - g(k)$.

Lemma 2.2. *If $Y \sim \mathrm{NB}(r, p)$ and, for $A \subseteq \mathbb{Z}_+$, $g := g_A$ satisfies the Stein equation*

$$(1 - p)(r + k)g(k + 1) - kg(k) = \mathbb{I}[k \in A] - \mathbb{P}(Y \in A), \tag{2.5}$$

then, for $k = 0, 1 \ldots,$

$$|(k + 1)g(k + 1)| \leq \frac{\max\{1, r\}\mathrm{e}}{p} \quad and \quad |\Delta g(k)| \leq \min\left\{\frac{1}{(1 - p)(r + k)}, \frac{1}{k}\right\}.$$

Proof. The second assertion bounding $|\Delta g(k)|$ is Theorem 2.10 applied to Example 2.9 of [6]. For the first assertion, from a standard representation of the solution to the Stein equation for discrete distributions [5, Equation (2)],

$$(k + 1)g(k + 1) = \frac{[\mathbb{P}(Y \in A, Y \leq k) - \mathbb{P}(Y \in A)\mathbb{P}(Y \leq k)]}{\mathbb{P}(Y = k + 1)}$$

$$= \frac{[\mathbb{P}(Y \in A, Y \leq k)\mathbb{P}(Y \geq k + 1) - \mathbb{P}(Y \in A, Y \geq k + 1)\mathbb{P}(Y \leq k)]}{\mathbb{P}(Y = k + 1)},$$

so we find that

$$|(k+1)g(k+1)| \leq \frac{\mathbb{P}(Y \geq k+1)\mathbb{P}(Y \leq k)}{\mathbb{P}(Y = k+1)}, \tag{2.6}$$

and the bound also holds with either term alone in the numerator.

If $r = 1$ (the geometric distribution) then we can compute (2.6) exactly as $(1 - (1 - p)^{k+1})/p \leq 1/p$, as desired. If $0 < r < 1$ then Proposition 1(b) of [12] implies that $\mathbb{P}(Y \geq k+1)/\mathbb{P}(Y = k+1) \leq 1/p$, which implies the result in this case.

If $r > 1$ then we bound (2.6) in three cases: $k+1 \geq r(1-p)/p$, $k+1 \leq (r-1)(1-p)/p$, and $(r-1)(1-p)/p + 1 \leq k+1 \leq r(1-p)/p - 1$. For the first case, Proposition 1(b) of [12] implies that, for $k+1 \geq r(1-p)/p$,

$$\frac{\mathbb{P}(Y \geq k+1)}{\mathbb{P}(Y = k+1)} \leq \left(1 - (1-p)\frac{k+1+r}{k+2}\right)^{-1}. \tag{2.7}$$

The right-hand side is decreasing in k, so setting $k+1 = r(1-p)/p$ and simplifying, we find that, for $k+1 \geq r(1-p)/p$, (2.7) is bounded by $r/p - r + 1 \leq r/p$, as desired. For the other two cases, we use the representation (see, e.g. Equation (2.27) of [1])

$$\mathbb{P}(Y \leq k) = \frac{\Gamma(r+k+1)}{\Gamma(r)\Gamma(k+1)} \int_0^p u^{r-1}(1-u)^k \, du,$$

which implies that (2.6) is bounded by

$$\frac{(k+1)\int_0^p u^{r-1}(1-u)^k \, du}{p^r(1-p)^{k+1}}. \tag{2.8}$$

The maximum of the integrand is achieved at $p_* = (r-1)/(r+k-1)$ and if $k+1 \leq (r-1)(1-p)/p$ then $p_* \geq p$, which implies that

$$\int_0^p u^{r-1}(1-u)^k \, du \leq p^r(1-p)^k$$

and, thus, that (2.8) is bounded by $(k+1)/(1-p) \leq (r-1)/p \leq r/p$ due to the restriction on the value of k.

Finally, assume that $(r-1)(1-p)/p + 1 \leq k+1 \leq r(1-p)/p - 1$ and note that in order for such k to exist, $0 < p \leq \frac{1}{3}$, which we assume holds for the remainder of the proof. With p_* as above,

$$\int_0^p u^{r-1}(1-u)^k \, du \leq p p_*^{r-1}(1-p_*)^k,$$

and the lower bound on the range of k implies that $p_* \leq p$ and so we find that (2.8) is bounded above by

$$\frac{(k+1)(1-p_*)^k}{(1-p)^{k+1}} \leq \frac{r}{p}\left(\frac{1-p_*}{1-p}\right)^k. \tag{2.9}$$

Recalling that $1 - p \leq 1 - p_* = k/(r+k-1)$, it is easy to see that the right-hand side of (2.9) is increasing in k. Substituting the maximum value of k for this case, $r(1-p)/p - 2$, into p_* and then this into (2.9) and simplifying, we find that (2.8) is bounded above by

$$\frac{r}{p}\left(\frac{r/p - 2/(1-p)}{r/p - 3}\right)^{r/p-r-2} \leq \frac{r}{p}e^{3-2/(1-p)} \leq \frac{r}{p}e,$$

where the first inequality follows since $r/p - r - 2 \leq r/p - 3$ and, for $a, x > 0$, $(1+a/x)^x \leq e^a$.

We need the following easy corollary of Lemma 2.2.

Lemma 2.3. *If, for $A \subseteq \mathbb{Z}_+$, $g := g_A$ satisfies the Stein equation (2.5) then*

$$\sup_{k \in \mathbb{Z}_+} |D^{(r)}g(k)| \leq \frac{\max\{r, 1\}e + 1}{r(1 - p)},$$

$$\sup_{k \in \mathbb{Z}_+} |\Delta(D^{(r)}g(k))| \leq \min\left\{1 + \frac{2}{r}, \frac{2 - p}{r(1 - p)}\right\}.$$

Proof. For the first assertion, since g solves the Stein equation (2.5),

$$|r(1 - p)D^{(r)}g(k)| \leq |pkg(k) + \mathbb{I}[k \in A] - \mathbb{P}(Y \in A)|$$
$$\leq |pkg(k)| + |\mathbb{I}[k \in A] - \mathbb{P}(Y \in A)|$$
$$\leq \max\{r, 1\}e + 1,$$

where we have used Lemma 2.2.

For the second assertion, it is easy to see that

$$\Delta(D^{(r)}g(k)) = \frac{r + k + 1}{r}\Delta g(k + 1) - \frac{k}{r}\Delta g(k),$$

and the lemma follows after taking the absolute value, applying the triangle inequality, and judiciously using Lemma 2.2.

Proof of Theorem 1.2. Following the usual Stein method machinery, for $g := g_A$ solving (2.5) for $A \subseteq \mathbb{Z}_+$ and $Y \sim \mathrm{NB}(r, p)$, we have

$$d_{\mathrm{TV}}(\mathcal{L}(W), \mathrm{NB}(r, p)) = \sup_{A \subseteq \mathbb{Z}_+} |\mathbb{E}[\mathbb{I}[W \in A] - \mathbb{P}(Y \in A)]|$$
$$= \sup_{A \subseteq \mathbb{Z}_+} |\mathbb{E}[(1 - p)(r + W)g_A(W + 1) - Wg_A(W)]|,$$
$$= p \sup_{A \subseteq \mathbb{Z}_+} |\mathbb{E}[\mu D^{(r)}g_A(W) - Wg_A(W)]|.$$

Lemma 2.1 implies that, for $g := g_A$,

$$p\mathbb{E}[\mu D^{(r)}g(W) - Wg(W)]$$
$$= p\mu\mathbb{E}[D^{(r)}g(W) - D^{(r)}g(W^{*_r})]$$
$$= p\mu\mathbb{E}[(D^{(r)}g(W) - D^{(r)}g(W^{*_r}))\mathbb{I}_B] + p\mu\mathbb{E}[(D^{(r)}g(W) - D^{(r)}g(W^{*_r}))\mathbb{I}_{B^c}]$$
$$=: R_1 + R_2.$$

Using the fact that $\mu = r(1 - p)/p$, we have

$$p\mu|D^{(r)}g(W) - D^{(r)}g(W^{*_r})| \leq 2r(1 - p) \sup_{k \in \mathbb{Z}_+} |D^{(r)}g(k)|,$$

and so Lemma 2.3 implies that $|R_2| \leq 2(e \max\{1, r\} + 1)\mathbb{P}(B^c)$.

To bound $|R_1|$, we write

$$|D^{(r)}g(W) - D^{(r)}g(W^{*r})|$$

$$= \left| \mathbb{I}[W > W^{*r}] \sum_{k=0}^{W-W^{*r}-1} \Delta D^{(r)}g(W^{*r}+k) - \mathbb{I}[W^{*r} > W] \sum_{k=0}^{W^{*r}-W-1} \Delta D^{(r)}g(W+k) \right|$$

$$\leq \sup_{k\in\mathbb{Z}_+} |\Delta(D^{(r)}g(k))||W^{*r} - W|.$$

Combining this with the bound of Lemma 2.3, we find that

$$|R_1| \leq \min\{(r+2)(1-p), 2-p\}\mathbb{E}|W^{*r} - W|\mathbb{I}_B,$$

which, upon adding to the bound on $|R_2|$, yields the first bound in the theorem. The second bound is obtained from the first by choosing $B = \{W = W^{*r}\}$.

3. Preferential attachment proof

In this section we prove Theorem 1.1 following the strategy of proof of the main result of Section 6 of [16], which is a special case of our results (it will likely help the reader to first understand the proof there). We use $C_{m,\delta}$ to denote a constant depending only on m and δ which may change from line to line.

Theorem 1.1 easily follows by the triangle inequality applied to the following three claims. If I is uniform on $\{1,\ldots,n\}$, independent of $W_{n,i}$ defined to be the in-degree of vertex i in $G_n^{m,\delta}$, and $\mu_{n,i} := \mathbb{E}W_{n,i}$, then

1. $d_{\mathrm{TV}}\left(\mathcal{L}(W_{n,I}), \mathrm{NB}\left(m+\delta, \dfrac{m+\delta}{\mu_{n,I}+m+\delta}\right)\right) \leq C_{m,\delta}\dfrac{\log(n)}{n}$,

2. $d_{\mathrm{TV}}\left(\mathrm{NB}\left(m+\delta, \dfrac{m+\delta}{\mu_{n,I}+m+\delta}\right), \mathrm{NB}\left(m+\delta, \left(\dfrac{I}{n}\right)^{1/(2+\delta/m)}\right)\right) \leq C_{m,\delta}\dfrac{\log(n)}{n}$,

3. $d_{\mathrm{TV}}\left(\mathrm{NB}\left(m+\delta, \left(\dfrac{I}{n}\right)^{1/(2+\delta/m)}\right), K(m,\delta)\right) \leq C_{m,\delta}\dfrac{\log(n)}{n}$.

The proofs of items 2 and 3 are relatively straightforward, while the proof of item 1 uses the following result which we show using Stein's method (i.e. Theorem 1.2).

Theorem 3.1. *Retaining the notation and definitions above, we have*

$$d_{\mathrm{TV}}\left(\mathcal{L}(W_{n,i}), \mathrm{NB}\left(m+\delta, \dfrac{m+\delta}{\mu_{n,i}+m+\delta}\right)\right) \leq \dfrac{C_{m,\delta}}{i}.$$

The layout of the remainder of this section is as follows. We first collect and prove some lemmas necessary for the proof of items 1–3 above and then prove these results. We prove Theorem 3.1 last, since it is relatively involved.

Since $G_n^{m,\delta}$ is constructed from $G_{nm}^{1,\delta/m}$, it will be helpful to denote by $W_{k,j}^{(1,\varepsilon)}$ the in-degree of vertex j in $G_k^{1,\varepsilon}$ for $k \geq j-1$, where we set $W_{j-1,1}^{(1,\varepsilon)} := 0$. The first lemma is useful for computing moment information; it is a small variation of a special case of the remarkable results of [13]; see also Proposition 8.9 of [21].

Theorem 3.2. ([13].) *If $j \geq 1$ and $\varepsilon > -1$, then the sequence of random variables*

$$\frac{\Gamma(k + (1 + \varepsilon)/(2 + \varepsilon))}{\Gamma(k + 1)} (W_{k,j}^{(1,\varepsilon)} + 1 + \varepsilon)$$

is a martingale for $k \geq j - 1$, where we take $W_{j-1,j}^{(1,\varepsilon)} := 0$. In particular, for $k \geq j - 1$,

$$\mathbb{E}W_{k,j}^{(1,\varepsilon)} + 1 + \varepsilon = (1 + \varepsilon)\frac{\Gamma(k + 1)\Gamma(j - 1 + (1 + \varepsilon)/(2 + \varepsilon))}{\Gamma(k + (1 + \varepsilon)/(2 + \varepsilon))\Gamma(j)}.$$

We also need asymptotic estimates for the ratio of gamma functions. The next result follows from Stirling's approximation.

Lemma 3.1. *For fixed $a, b > 0$, as $z \to \infty$,*

$$\frac{\Gamma(z + a)}{\Gamma(z + b)} = z^{a-b} + O(z^{a-b-1}).$$

The next lemma provides a nice asymptotic expression for expectations appearing in the proofs below.

Lemma 3.2. *If $n \geq i$ and $-\delta < m \in \mathbb{N}$, and $\mu_{n,i}^{(m,\delta)} := \mathbb{E}W_{n,i}$, then*

$$\left| \frac{\mu_{n,i}^{(m,\delta)}}{m + \delta} + 1 - \left(\frac{n}{i}\right)^{1/(2+\delta/m)} \right| \leq C_{m,\delta} \left(\frac{n}{i}\right)^{1/(2+\delta/m)} \frac{1}{i},$$

$$\left| \frac{m + \delta}{\mu_{n,i}^{(m,\delta)} + m + \delta} - \left(\frac{i}{n}\right)^{1/(2+\delta/m)} \right| \leq \frac{C_{m,\delta}(i/n)^{1/(2+\delta/m)}}{i}.$$

Proof. The second inequality follows directly from the first. For the first assertion, Theorem 3.2 implies that, for $\varepsilon > -1$ and $\mu_{k,j}^{(1,\varepsilon)} := \mathbb{E}W_{k,j}^{(1,\varepsilon)}$ for $k \geq j - 1$,

$$\mu_{k,j}^{(1,\varepsilon)} = (1 + \varepsilon)\left[\frac{\Gamma(j - 1 + (1 + \varepsilon)/(2 + \varepsilon))\Gamma(k + 1)}{\Gamma(j)\Gamma(k + (1 + \varepsilon)/(2 + \varepsilon))} - 1\right].$$

The construction of $G_n^{m,\delta}$ implies that

$$\mu_{n,i}^{(m,\delta)} = \sum_{j=1}^{m} \mu_{nm,(i-1)m+j}^{(1,\delta/m)},$$

so we find that

$$\mu_{n,i}^{(m,\delta)} + (m + \delta) = \left(1 + \frac{\delta}{m}\right)\frac{\Gamma(nm + 1)}{\Gamma(nm + (1 + \delta/m)/(2 + \delta/m))}$$

$$\times \left[\sum_{j=1}^{m} \frac{\Gamma((i - 1)m + j - 1 + (1 + \delta/m)/(2 + \delta/m))}{\Gamma((i - 1)m + j)}\right]. \qquad (3.1)$$

Now using Lemma 3.1 for the ratios of gamma functions, we find that, for $i > 1$,

$$\frac{\mu_{n,i}^{(m,\delta)}}{m + \delta} + 1 = \frac{1}{m}((nm)^{1/(2+\delta/m)} + O(n^{-(1+\delta/m)/(2+\delta/m)}))$$

$$\times \left(m((i - 1)m)^{-1/(2+\delta/m)} + \sum_{j=1}^{m} O(i^{-(3+\delta/m)/(2+\delta/m)})\right).$$

The lead term equals $(n/i)^{1/(2+\delta/m)}$ (up to the error in changing $i-1$ to i), and the second-order term is easily seen to be as desired. In the case that $i = 1$, similar arguments starting from (3.1) yield the appropriate complementary result.

To prove items 2 and 3, we have to bound the total variation distance between negative binomial distributions having different 'p' parameters. The next result is sufficient for our purposes.

Lemma 3.3. *If $r > 0$ and $0 \leq \varepsilon < p \leq 1$, then*

$$d_{\mathrm{TV}}(\mathrm{NB}(r, p), \mathrm{NB}(r, p - \varepsilon)) \leq \frac{r\varepsilon}{p - \varepsilon}. \tag{3.2}$$

Proof. Proposition 2.5 of [1] implies that, for $r > 0$ (their statement is for $r \in \mathbb{N}$, but the same proof works for all $r > 0$),

$$d_{\mathrm{TV}}(\mathrm{NB}(r, p), \mathrm{NB}(r, p - \varepsilon)) = (r + l - 1) \int_{p-\varepsilon}^{p} q(u)\, \mathrm{d}u, \tag{3.3}$$

where $0 \leq q(u) \leq 1$ and

$$l \leq \frac{r(1 - p + \varepsilon)}{p - \varepsilon} + 1.$$

Using these bounds on q and l in (3.3) implies the lemma.

Our final lemma is useful for handling the total variation distance for conditionally defined random variables.

Lemma 3.4. *Let W and V be random variables, and let X be a random element defined on the same probability space. Then*

$$d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{L}(V)) \leq \mathbb{E}\, d_{\mathrm{TV}}(\mathcal{L}(W \mid X), \mathcal{L}(V \mid X)).$$

Proof. If $f : \mathbb{R} \to [0, 1]$ then

$$|\mathbb{E}[f(W) - f(V)]| \leq \mathbb{E}|\mathbb{E}[f(W) - f(V) \mid X]| \leq \mathbb{E}\, d_{\mathrm{TV}}(\mathcal{L}(W \mid X), \mathcal{L}(V \mid X)).$$

Proof of Theorem 1.1. Using (3.2) and Lemma 3.2, we easily obtain

$$d_{\mathrm{TV}}\left(\mathrm{NB}\left(m + \delta, \frac{m + \delta}{\mu_{n,i} + m + \delta}\right), \mathrm{NB}\left(m + \delta, \left(\frac{i}{n}\right)^{1/(2+\delta/m)}\right)\right) \leq \frac{C_{m,\delta}}{i},$$

and applying Lemma 3.4 we find that

$$d_{\mathrm{TV}}\left(\mathrm{NB}\left(m + \delta, \frac{m + \delta}{\mu_{n,I} + m + \delta}\right), \mathrm{NB}\left(m + \delta, \left(\frac{I}{n}\right)^{1/(2+\delta/m)}\right)\right) \leq \frac{C_{m,\delta} \log(n)}{n},$$

which is item 2 above. Now, we couple U to I by writing $U = I/n - V$, where V is uniform on $(0, 1/n)$ and independent of I. From here, use Lemmas 3.3 and 3.4, and then the easy fact that, for $i \geq 1$ and $0 < a < 1$,

$$i^a - (i - 1)^a \leq i^{a-1},$$

to find that

$$d_{\text{TV}}\left(\text{NB}\left(m+\delta, \left(\frac{I}{n}\right)^{1/(2+\delta/m)}\right), K(m,\delta)\right)$$

$$= d_{\text{TV}}\left(\text{NB}\left(m+\delta, \left(\frac{I}{n}\right)^{1/(2+\delta/m)}\right), \text{NB}(m+\delta, U^{1/(2+\delta/m)})\right)$$

$$\leq \frac{C_{m,\delta}}{n}\sum_{i=1}^{n}\frac{(i/n)^{1/(2+\delta/m)} - ((i-1)/n)^{1/(2+\delta/m)}}{(i/n)^{1/(2+\delta/m)}}$$

$$\leq \frac{C_{m,\delta}\log(n)}{n},$$

which is item 3 above. Finally, applying Lemma 3.4 to Theorem 3.1 yields the claim in item 1 above so that Theorem 1.1 is proved.

The remainder of the section is devoted to the proof of Theorem 3.1. Since we want to apply our negative binomial approximation framework we must first construct a random variable having the $(m+\delta)$-equilibrium distribution of $W_{n,i} := W_{n,i}^{(m,\delta)}$. According to Definition 1.2, we first construct a variable having the size bias distribution of $W_{n,i}$. To facilitate this construction, we need some auxiliary variables.

We mostly work with $G_n^{m,\delta}$ through the intermediate construction of $G_{nm}^{1,\delta/m}$ discussed in the introduction. To fix notation, if, for $k \geq j$, $W_{k,j}^{(1,\delta/m)}$ is the degree of vertex j in $G_k^{1,\delta/m}$ then we write

$$W_{n,i} = \sum_{j=1}^{m} W_{nm,m(i-1)+j}^{(1,\delta/m)}. \tag{3.4}$$

Furthermore, if we let $X_{j,i}^{(\delta/m)}$ be the indicator that vertex j attaches to vertex i in $G_j^{1,\delta/m}$ (and hence also in $G_k^{1,\delta/m}$ for $j \leq k \leq mn$), then we also have

$$W_{nm,m(i-1)+j}^{(1,\delta/m)} = \sum_{k=m(i-1)+j}^{mn} X_{k,m(i-1)+j}^{(\delta/m)}. \tag{3.5}$$

The following well-known result allows us to use the decomposition of $W_{n,i}$ into a sum of indicators as per (3.4) and (3.5) to size bias $W_{n,i}$; see, e.g. Proposition 2.2 of [8] and the discussion thereafter.

Proposition 3.1. *Let $X_1, \ldots, X_n$ be zero–one random variables such that $\mathbb{P}(X_j = 1) = p_j$. For each $k = 1, \ldots, n$, let $(X_j^{(k)})_{j \neq k}$ have the distribution of $(X_j)_{j \neq k}$ conditional on $X_k = 1$. If $X = \sum_{j=1}^{n} X_j$, $\mu = \mathbb{E}[X]$, and K is chosen independent of the variables above with $\mathbb{P}(K = k) = p_k/\mu$, then $X^s = \sum_{j \neq K} X_j^{(K)} + 1$ has the size bias distribution of X.*

Roughly, Proposition 3.1 implies that in order to size bias $W_{n,i}$, we choose an indicator $X_{K,L}^{(\delta/m)}$ where, for $l = m(i-1)+1, \ldots, mi$, $k = l, \ldots, mn$, $\mathbb{P}(K = k, L = l)$ is proportional to $\mathbb{P}(X_{k,l}^{(\delta/m)} = 1)$ (and zero for other values), then attach vertex K to vertex L and sample the remaining edges conditional on this event. Note that, given $(K, L) = (k, l)$, in the graphs $G_j^{1,\delta/m}$, $1 \leq j < l$ and $k < j \leq nm$, this conditioning does not change the original rule for generating the preferential attachment graph given $G_{j-1}^{1,\delta/m}$. The following lemma implies the remarkable fact that in order to generate the graphs $G_j^{1,\delta/m}$ for $l \leq j < k$ conditional on $X_{k,l}^{(\delta/m)} = 1$ and G_{l-1}, we attach edges following the same rule as preferential attachment, but include the edge from vertex k to vertex l in the degree count.

Lemma 3.5. *Retaining the notation and definitions above, for $l, s \leq j < k$, we have*

$$\mathbb{P}(X_{j,s}^{(\delta/m)} = 1 \mid X_{k,l}^{(\delta/m)} = 1, G_{j-1}^{1,\delta/m}) = \frac{\mathbb{I}[s = l] + W_{j-1,s}^{(1,\delta/m)} + \delta/m + 1}{j(2 + \delta/m)}, \qquad (3.6)$$

where we define $W_{j-1,j}^{(1,\delta/m)} = 0$.

   Proof. By the definition of the conditional probability, we write

$$\mathbb{P}(X_{j,s}^{(\delta/m)} = 1 \mid X_{k,l}^{(\delta/m)} = 1, G_{j-1}^{1,\delta/m})$$

$$= \frac{\mathbb{P}(X_{j,s}^{(\delta/m)} = 1 \mid G_{j-1}^{1,\delta/m})\mathbb{P}(X_{k,l}^{(\delta/m)} = 1 \mid X_{j,s}^{(\delta/m)} = 1, G_{j-1}^{1,\delta/m})}{\mathbb{P}(X_{k,l}^{(\delta/m)} = 1 \mid G_{j-1}^{1,\delta/m})}, \qquad (3.7)$$

and we calculate the three probabilities appearing above. First note that

$$\mathbb{P}(X_{j,s}^{(\delta/m)} = 1 \mid G_{j-1}^{1,\delta/m}) = \frac{W_{j-1,s}^{(1,\delta/m)} + 1 + \delta/m}{(2 + \delta/m)j - 1},$$

which implies that

$$\mathbb{P}(X_{k,l}^{(\delta/m)} = 1 \mid G_{j-1}^{1,\delta/m}) = \frac{\mathbb{E}[W_{k-1,l}^{(1,\delta/m)} + 1 + \delta/m \mid G_{j-1}^{1,\delta/m}]}{(2 + \delta/m)k - 1}$$

and

$$\mathbb{P}(X_{k,l}^{(\delta/m)} = 1 \mid X_{j,s}^{(\delta/m)} = 1, G_{j-1}^{1,\delta/m})$$

$$= \frac{\mathbb{E}[W_{k-1,l}^{(1,\delta/m)} + 1 + \delta/m \mid X_{j,s}^{(\delta/m)} = 1, G_{j-1}^{1,\delta/m}]}{(2 + \delta/m)k - 1}.$$

Using Theorem 3.2, it easy to see that

$$\mathbb{E}\left[W_{k-1,l}^{(1,\delta/m)} + 1 + \frac{\delta}{m} \,\middle|\, G_{j-1}^{1,\delta/m}\right]$$

$$= \frac{\Gamma(k)\Gamma(j - 1 + (1 + \delta/m)/(2 + \delta/m))}{\Gamma(k - 1 + (1 + \delta/m)/(2 + \delta/m))\Gamma(j)}\left(W_{j-1,l}^{(1,\delta/m)} + 1 + \frac{\delta}{m}\right),$$

and also that

$$\mathbb{E}\left[W_{k-1,l}^{(1,\delta/m)} + 1 + \frac{\delta}{m} \,\middle|\, X_{j,s}^{(\delta/m)} = 1, G_{j-1}^{1,\delta/m}\right]$$

$$= \frac{\Gamma(k)\Gamma(j + (1 + \delta/m)/(2 + \delta/m))}{\Gamma(k - 1 + (1 + \delta/m)/(2 + \delta/m))\Gamma(j + 1)}\left(W_{j-1,l}^{(1,\delta/m)} + \mathbb{I}[s = l] + 1 + \frac{\delta}{m}\right).$$

Combining these calculations with (3.7) and simplifying (using in particular the fact that $\Gamma(x + 1) = x\Gamma(x)$) implies that

$$\mathbb{P}(X_{j,s}^{(\delta/m)} = 1 \mid X_{k,l}^{(\delta/m)} = 1, G_{j-1}^{1,\delta/m})$$

$$= \frac{1}{j(2 + \delta/m)}\frac{(W_{j-1,l}^{(1,\delta/m)} + \mathbb{I}[s = l] + 1 + \delta/m)(W_{j-1,s}^{(1,\delta/m)} + 1 + \delta/m)}{W_{j-1,l}^{(1,\delta/m)} + 1 + \delta/m}. \qquad (3.8)$$

By considering the cases $s = l$ and $s \neq l$ separately, it follows that (3.8) equals (3.6).

The previous lemma suggests the following (embellished) construction of a variable having distribution $(W_{n,i} \mid X_{k,l}^{(\delta/m)} = 1)$. Here and below we denote quantities related to this construction by appending (k, l). First we generate $G_{l-1}^{1,\delta/m}(k, l)$, a graph with $l - 1$ vertices, according to the usual preferential attachment model. At this point, if $l \neq k$, vertices l and k are added to the graph, along with a vertex labeled i' with an edge to it emanating from vertex k. Given $G_{l-1}^{1,\delta/m}(k, l)$ and these additional vertices and edges, we generate $G_l^{1,\delta/m}(k, l)$ by connecting vertex l to a vertex randomly chosen from the vertices $1, \ldots, l, i'$ proportional to their 'degree weight', where vertex l has degree weight $1 + \delta/m$ (from the out-edge) and i' has degree 1 (from the in-edge emanating from vertex k), and the remaining vertices have degree weight equal to their degree plus δ/m. For $l < j < k$, we generate the graphs $G_j^{1,\delta/m}(k, l)$ recursively from $G_{j-1}^{1,\delta/m}(k, l)$ by connecting vertex j to a vertex randomly chosen from the vertices $1, \ldots, j, i'$ proportional to their degree weight, where j has degree weight $1 + \delta/m$ (from the out-edge). Note that none of the vertices $1, \ldots, k - 1$ connect to vertex k. Also, define $G_k^{1,\delta/m}(k, l) = G_{k-1}^{1,\delta/m}(k, l)$. If $l = k$, we attach vertex k to i' and denote the resulting graph by $G_k^{1,\delta/m}(k, l)$. For all values (k, l), if $j = k + 1, \ldots, nm$, we generate $G_j^{1,\delta/m}(k, l)$ from $G_{j-1}^{1,\delta/m}(k, l)$ according to usual preferential attachment among the vertices $1, \ldots, j, i'$.

We have a final bit of notation before stating relevant properties of these objects. Denote the degree of vertex j in this construction by $W_{nm,j}^{(1,\delta/m)}(k, l)$, and also let

$$W_{n,i}(k, l) := \sum_{j=1}^{m} W_{nm,m(i-1)+j}^{(1,\delta/m)}(k, l).$$

Let $B_{k,l}$ be the event that in this construction all edges emanating from the vertices $m(i-1)+1$, $\ldots, mi$ attach to one of the vertices $1, \ldots, m(i - 1)$. In symbols,

$$B_{k,l} = \left\{ \begin{array}{l} X_{s,j}^{(\delta/m)}(k, l) = 0 \text{ for all } j \in \{m(i - 1) + 1, \ldots, mi, i'\} \\ s \in \{m(i - 1) + 1, \ldots, mi, i'\}/\{k\} \end{array} \right\}.$$

Finally, let W' have the r-equilibrium distribution of $W_{n,i}$, independent of all else and define

$$W_{n,i}^{*_r} = W_{n,i}(K, L)\mathbb{I}_{B_{K,L}} + W'\mathbb{I}_{B_{K,L}^c}. \tag{3.9}$$

Lemma 3.6. *Let* $l \in \{m(i - 1) + 1, \ldots, mi\}$ *and* $k \in \{l, \ldots, mn\}$, *and retain the notation and definitions above.*

1. $\mathcal{L}(W_{n,i}(k, l) + W_{nm,i'}^{(1,\delta/m)}(k, l)) = \mathcal{L}(W_{n,i} \mid X_{k,l}^{(\delta/m)} = 1)$.

2. *If* (K, L) *is a random vector such that*

$$\mathbb{P}(K = k', L = l') = \frac{\mathbb{E}X_{k',l'}^{(\delta/m)}}{\mathbb{E}W_{n,i}}, \qquad k' \geq l' \in \{m(i - 1) + 1, \ldots, mi\},$$

   then $W_{n,i}(K, L) + W_{nm,i'}^{(1,\delta/m)}(K, L)$ *has the size bias distribution of* $W_{n,i}$.

3. *Conditional on the event*

$$\{W_{n,i}(K, L) + W_{nm,i'}^{(1,\delta/m)}(K, L) = t\},$$

   $\mathcal{L}(W_{n,i}(K, L)\mathbb{I}[B_{K,L}]) = \mathcal{L}(U_{m+\delta,t}\mathbb{I}[B_{K,L}])$, *where* $U_{r,t}$ *has the Pólya urn distribution of Definition 1.2 and is independent of all else.*

4. $W_{n,i}^{*_r}$ *has the* $(m + \delta)$-*equilibrium distribution of* $W_{n,i}$.

Proof. Items 1 and 2 follow from Proposition 3.1 and Lemma 3.5. Item 3 follows since, under the conditioning, if $\mathbb{I}[B_{K,L}] = 1$ then $W_{n,i}(K, L)$ is distributed as the number of white balls *drawn* in $t - 1$ draws from a Pólya urn started with $m + \delta$ white balls and 1 black ball (it is $t - 1$ draws, rather than t, since the initial 'black ball' degree from vertex i' is included in the degree count $W_{n,i}(K, L) + W_{nm,i'}^{(1,\delta/m)}(K, L)$). Item 4 follows from items 1–3, using Definition 1.2.

Proof of Theorem 3.1. We apply Theorem 1.2 to $\mathcal{L}(W_{n,i})$ with $W_{n,i}^{*r}$ as defined by (3.9). Before constructing the coupling of $\mathcal{L}(W_{n,i})$ required in Theorem 1.2, we reduce the bound $\mathbb{P}(W_{n,i}^{*r} \neq W_{n,i})$.

First note that due to the form of $W_{n,i}^{*r}$, we have (no matter how $\mathcal{L}(W_{n,i})$ is coupled)

$$\mathbb{P}(W_{n,i}^{*r} \neq W_{n,i}) = \mathbb{P}(W_{n,i}(K, L) \neq W_{n,i}, B_{K,L}) + \mathbb{P}(W' \neq W_{n,i}, B_{K,L}^c)$$
$$\leq \mathbb{P}(W_{n,i}(K, L) \neq W_{n,i}) + \mathbb{P}(B_{K,L}^c). \tag{3.10}$$

We bound the second term of (3.10) as follows. For $l \in \{m(i - 1) + 1, \ldots, mi\}$ and $k > mi$, we directly compute

$$\mathbb{P}(B_{k,l}) = \prod_{j=m(i-1)+1}^{l-1} \frac{m(i-1)(1+\delta/m) + j - 1}{j(2+\delta/m) - 1} \prod_{j=l}^{mi} \frac{m(i-1)(1+\delta/m) + j - 1}{j(2+\delta/m)}$$

$$\geq \prod_{j=m(i-1)+1}^{mi} \frac{m(i-1)(1+\delta/m) + j - 1}{j(2+\delta/m)}$$

$$= \frac{1}{(2+\delta/m)^m} \frac{\Gamma(m(i-1)(2+\delta/m) + m)\Gamma(m(i-1) + 1)}{\Gamma(m(i-1)(2+\delta/m))\Gamma(mi + 1)} \tag{3.11}$$

$$= 1 + O\left(\frac{1}{i}\right),$$

where in the last equality we use Lemma 3.1. If $k \in \{m(i - 1) + 1, \ldots, mi\}$ then

$$\mathbb{P}(B_{k,l}) = \prod_{j=m(i-1)+1}^{l-1} \frac{m(i-1)(1+\delta/m) + j - 1}{j(2+\delta/m) - 1} \prod_{\substack{j=l \\ j \neq k}}^{mi} \frac{m(i-1)(1+\delta/m) + j - 1}{j(2+\delta/m)},$$

which is greater than or equal to (3.11) (since the omitted term is a probability), so in either case we find that

$$\mathbb{P}(B_{K,L}^c) = O\left(\frac{1}{i}\right).$$

We have only left to bound the first term of (3.10), for which we must first define the coupling of $\mathcal{L}(W_{n,i})$ to $W_{n,i}(K, L)$. For each (k, l) in the support of (K, L), we construct

$$\{(X_{s,j}^{(\delta/m)}(k, l), \widetilde{X}_{s,j}^{(\delta/m)}) : mn \geq s \geq j \in \{m(i - 1) + 1, \ldots, mi\}\}$$

to have the distribution of the indicators of the events vertex s connects to vertex j in $G_{nm}^{1,\delta/m}(k, l)$ and $G_{nm}^{1,\delta/m}$, respectively. With this fact established, define

$$W_{nm,j}^{(1,\delta/m)}(k, l) = \sum_{s=j}^{nm} X_{s,j}^{(\delta/m)}(k, l) \quad \text{and} \quad \widetilde{W}_{nm}^{(1,\delta/m)} = \sum_{s=j}^{nm} \widetilde{X}_{s,j}^{(\delta/m)},$$

which have the distribution of vertex j in the indicated graphs, and then set

$$W_{n,i}(k,l) = \sum_{j=m(i-1)+1}^{mi} W_{nm,j}^{(1,\delta/m)}(k,l) \quad \text{and} \quad W_{n,i} = \sum_{j=m(i-1)+1}^{mi} \widetilde{W}_{nm,j}^{(1,\delta/m)}.$$

From this point we bound the first term of (3.10) via

$$\mathbb{P}(W_{n,i}(k,l) \neq W_{n,i}) \leq \mathbb{P}\left(\bigcup_{j=m(i-1)+1}^{mi} \{W_{nm,j}^{(1,\delta/m)}(k,l) \neq \widetilde{W}_{nm,j}^{(1,\delta/m)}\} \right)$$

$$\leq \sum_{j=m(i-1)+1}^{mi} \mathbb{P}(W_{nm,j}^{(1,\delta/m)}(k,l) \neq \widetilde{W}_{nm,j}^{(1,\delta/m)}), \qquad (3.12)$$

and we show that each term in the sum is $O(1/i)$ (still depending on m and δ, but not on k and l), which establishes the theorem.

The constructions for different orders of j, k, l are slightly different, so assume that $j < l < k$. Let $U_{s,j}(k,l)$ be independent uniform $(0,1)$ random variables and, for the sake of brevity, let $w = 1 + \delta/m$. First define

$$X_{j,j}^{(\delta/m)}(k,l) = \mathbb{I}\left[U_{s,j}(k,l) < \frac{w}{j(2+\delta/m)-1} \right]$$

and, for $j < s < l$, given $W_{s-1,j}^{(1,\delta/m)}(k,l)$,

$$X_{s,j}^{(\delta/m)}(k,l) = \mathbb{I}\left[U_{s,j}(k,l) < \frac{W_{s-1,j}^{(1,\delta/m)}(k,l)+w}{s(2+\delta/m)-1} \right].$$

Also, let $\widetilde{X}_{s,j}^{(\delta/m)} = X_{s,j}^{(\delta/m)}(k,l)$ for $j \leq s < l$. That is, we can perfectly couple the degrees of vertex j in the two graphs up until vertex l arrives. Now, for $l \leq s < k$, given $W_{s-1,j}^{(1,\delta/m)}(k,l)$ and $\widetilde{W}_{s-1,j}^{(1,\delta/m)}$, define

$$X_{s,j}^{(\delta/m)}(k,l) = \mathbb{I}\left[U_{s,j}(k,l) < \frac{W_{s-1,j}^{(1,\delta/m)}(k,l)+w}{s(2+\delta/m)} \right], \qquad (3.13)$$

$$\widetilde{X}_{s,j}^{(\delta/m)} = \mathbb{I}\left[U_{s,j}(k,l) < \frac{\widetilde{W}_{s-1,j}^{(1,\delta/m)}+w}{s(2+\delta/m)-1} \right]. \qquad (3.14)$$

Set $X_{k,j}^{(\delta/m)}(k,l) = 0$ and $\widetilde{X}_{s,j}^{(\delta/m)}$ as in (3.14) with $s = k$ and, for $s > k$, define

$$X_{s,j}^{(\delta/m)}(k,l) = \mathbb{I}\left[U_{s,j}(k,l) < \frac{W_{s-1,j}^{(1,\delta/m)}(k,l)+w}{s(2+\delta/m)-1} \right],$$

$$\widetilde{X}_{s,j}^{(\delta/m)} = \mathbb{I}\left[U_{s,j}(k,l) < \frac{\widetilde{W}_{s-1,j}^{(1,\delta/m)}+w}{s(2+\delta/m)-1} \right].$$

For $j < l < k$, we have jointly and recursively defined the variables $X_{s,j}^{(\delta/m)}(k,l)$ and $\widetilde{X}_{s,j}^{(\delta/m)}$, and it is clear they are distributed as claimed above with $W_{nm,j}^{(1,\delta/m)}(k,l)$ and $\widetilde{W}_{nm,j}^{(1,\delta/m)}$ the required

degree counts. Also, note that $\widetilde{X}_{s,j}^{(\delta/m)} \geq X_{s,j}^{(\delta/m)}(k,l)$ and $\widetilde{W}_{s,j}^{(1,\delta/m)} \geq W_{s,j}^{(1,\delta/m)}(k,l)$, and now define the event

$$A_{s,j}(k,l) = \{\min\{j \leq t \leq nm \colon \widetilde{X}_{t,j}^{(\delta/m)} \neq X_{t,j}^{(\delta/m)}(k,l)\} = s\}.$$

Using the fact that $\widetilde{W}_{s-1,j}^{(1,\delta/m)} = W_{s-1,j}^{(1,\delta/m)}(k,l)$ under $A_{s,j}(k,l)$, we have

$$
\begin{aligned}
&\mathbb{P}(W_{nm,j}^{(1,\delta/m)}(k,l) \neq \widetilde{W}_{nm,j}^{(1,\delta/m)}) \\
&= \mathbb{P}\left(\bigcup_{s=j}^{nm} A_{s,j}(k,l)\right) \\
&\leq \sum_{s=l}^{k} \mathbb{P}\left(A_{s,j}(k,l) \cap \left\{\frac{W_{s-1,j}^{(1,\delta/m)}(k,l) + w}{s(2+\delta/m)} < U_{s,j}(k,l) < \frac{\widetilde{W}_{s-1,j}^{(1,\delta/m)} + w}{s(2+\delta/m) - 1}\right\}\right) \\
&\leq \mathbb{E}\widetilde{X}_{k,j}^{(\delta/m)} + \sum_{s=l}^{k-1} \mathbb{P}\left(\frac{W_{s-1,j}^{(1,\delta/m)} + w}{s(2+\delta/m)} < U_{s,j}(k,l) < \frac{\widetilde{W}_{s-1,j}^{(1,\delta/m)} + w}{s(2+\delta/m) - 1}\right).
\end{aligned}
$$

Now using Theorem 3.2, the estimates in Lemma 3.2, and the fact that $j, l \in \{m(i-1)+1, \ldots, mi\}$, we find that

$$
\begin{aligned}
&\mathbb{P}(W_{nm,j}^{(1,\delta/m)}(k,l) \neq \widetilde{W}_{nm,j}^{(1,\delta/m)}) \\
&\leq \frac{\mathbb{E}W_{k-1,j}^{(1,\delta/m)} + w}{k(2+\delta) - 1} + \sum_{s=l}^{k-1}(\mathbb{E}W_{s-1,j}^{(1,\delta/m)} + w)\left(\frac{1}{s(2+\delta/m) - 1} - \frac{1}{s(2+\delta/m)}\right) \\
&\leq C_{m,\delta}\left[\left(\frac{k}{j}\right)^{1/(2+\delta/m)}\frac{1}{k} + \left(\frac{k}{j}\right)^{1/(2+\delta/m)}\frac{1}{jk} \right. \\
&\qquad\qquad \left. + \sum_{s=l}^{\infty}\left(\left(\frac{s}{j}\right)^{1/(2+\delta/m)}\frac{1}{s^2} + \left(\frac{s}{j}\right)^{1/(2+\delta/m)}\frac{1}{js^2}\right)\right] \\
&\leq \frac{C_{m,\delta}}{i}.
\end{aligned}
$$

For the case $l < j < k$, the coupling is similar to that above, except it starts from (3.13) and (3.14) for $j \leq s < k$; the probability estimates are also similar. If $j > k$ then it is easy to see that the variables can be perfectly coupled. If $j = k$ or $j < l = k$ then the analog of the coupling above can only differ if the edge emanating from vertex k connects to j in $G_k^{1,\delta/m}$, which occurs with chance of order

$$\left(\frac{k}{j}\right)^{1/(2+\delta/m)}\frac{1}{k} = \mathrm{O}\left(\frac{1}{i}\right).$$

Thus, for any k, l in the support of (K, L) and $j \in \{m(i-1)+1, \ldots, mi\}$, each of the m terms in the sum (3.12) is bounded above by $C_{m,\delta}/i$, which establishes the result.

Acknowledgements

References

[1] ADELL, J. A. AND JODRÁ, P. (2006). Exact Kolmogorov and total variation distances between some familiar discrete distributions. *J. Inequal. Appl.* **2006,** 64307, 8 pp.

[2] BARABÁSI, A.-L. AND ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286,** 509–512.

[3] BARBOUR, A. D., HOLST, L. AND JANSON, S. (1992). *Poisson Approximation* (Oxford Stud. Prob. **2**). The Clarendon Press, Oxford University Press, New York.

[4] BOLLOBÁS, B., RIORDAN, O., SPENCER, J. AND TUSNÁDY, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18,** 279–290.

[5] BROWN, T. C. AND PHILLIPS, M. J. (1999). Negative binomial approximation with Stein's method. *Methodology Comput. Appl. Prob.* **1,** 407–421.

[6] BROWN, T. C. AND XIA, A. (2001). Stein's method and birth-death processes. *Ann. Prob.* **29,** 1373–1403.

[7] BUCKLEY, P. G. AND OSTHUS, D. (2004). Popularity based random graph models leading to a scale-free degree sequence. *Discrete Math.* **282,** 53–68.

[8] CHEN, L. H. Y., GOLDSTEIN, L. AND SHAO, Q.-M. (2011). *Normal Approximation by Stein's Method.* Springer, Heidelberg.

[9] GOLDSTEIN, L. AND REINERT, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Prob.* **7,** 935–952.

[10] JOHNSON, N. L. AND KOTZ, S. (1977). *Urn Models and Their Application.* John Wiley, New York.

[11] JORDAN, J. (2006). The degree sequences and spectra of scale-free random graphs. *Random Structures Algorithms* **29,** 226–242.

[12] KLAR, B. (2000). Bounds on tail probabilities of discrete distributions. *Prob. Eng. Inf. Sci.* **14,** 161–171.

[13] MÓRI, T. F. (2005). The maximum degree of the Barabási-Albert random tree. *Combinatorics Prob. Comput.* **14,** 339–348.

[14] PEKÖZ, E. A. (1996). Stein's method for geometric approximation. *J. Appl. Prob.* **33,** 707–713.

[15] PEKÖZ, E. A. AND RÖLLIN, A. (2011). New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Prob.* **39,** 587–608.

[16] PEKÖZ, E. A., RÖLLIN, A. AND ROSS, N. (2013). Total variation error bounds for geometric approximation. *Bernoulli* **19,** 610–632.

[17] PEKÖZ, E. A., RÖLLIN, A. AND ROSS, N. (2013). Degree asymptotics with rates for preferential attachment random graphs. *Ann. Appl. Prob.* **23,** 1188–1218.

[18] PITMAN, J. AND ROSS, N. (2012). Archimedes, Gauss, and Stein. *Notices AMS* **59,** 1416–1421.

[19] ROSS, N. (2011). Fundamentals of Stein's method. *Prob. Surveys* **8,** 210–293.

[20] RUDAS, A., TÓTH, B. AND VALKÓ, B. (2007). Random trees and general branching processes. *Random Structures Algorithms* **31,** 186–202.

[21] VAN DER HOFSTAD, R. (2012). Random graphs and complex networks. Lecture Notes, Eindhoven University of Technology. Available at http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf.