Consistency of Hill estimators in a linear preferential attachment model

Tiandong Wang · Sidney I. Resnick

Received: 16 November 2017 / Revised: 27 August 2018 / Accepted: 28 August 2018 / Published online: 8 September 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract
Preferential attachment is widely used to model power-law behavior of degree distributions in both directed and undirected networks. Statistical estimates of the tail exponent of the power-law degree distribution often use the Hill estimator as one of the key summary statistics. The consistency of the Hill estimator for network data has not been explored and the major goal in this paper is to prove consistency in certain models. To do this, we first derive the asymptotic behavior of the degree sequence via embedding the degree growth of a fixed node into a birth immigration process and then show the convergence of the tail empirical measure. From these steps, the consistency of the Hill estimator is obtained. Simulations are provided as an illustration for the asymptotic distribution of the Hill estimator.

Keywords Hill estimators · Power laws · Preferential attachment · Continuous time branching processes

AMS 2000 Subject Classifications 60G70 · 60B10 · 60G55 · 60G57 · 05C80 · 62E20

1 Introduction

The preferential attachment model gives a growing sequence of random graphs in which nodes and edges are added to the network based on probabilistic rules, and is used to mimic the evolution of social networks, collaborator and citation networks, as well as recommender networks. The probabilistic rule depends on the node degree

1 School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853, USA
and captures the feature that nodes with larger degrees tend to attract more edges. Empirical analysis of social network data shows that degree distributions follow power laws and theoretically, this is true for linear preferential attachment models. This agreement makes preferential attachment a popular choice for network modeling (Bollobás et al. 2003; Durrett 2010; Krapivsky et al. 2001; Krapivsky and Redner 2001; van der Hofstad 2017). This paper only focuses on the undirected case but the preferential attachment mechanism has been applied to both directed and undirected graphs. Limit theory for degree counts can be found in Resnick and Samorodnitsky (2016), Bhamidi (2007), Krapivsky and Redner (2001) for the undirected case and Wang and Resnick (2017), Samorodnitsky et al. (2016), Resnick and Samorodnitsky (2015), Wang and Resnick (2016) and Krapivsky et al. (2001) for the directed case.

Estimating the power-law index of the degree distribution is an important way to characterize a network. The indices of these distributions control the likelihood that nodes with large degrees appear in the data. Data repositories of large network datasets such as KONECT (http://konect.uni-koblenz.de/, Kunegis 2013) provide a Hill estimate as one of the key summary statistics for almost all listed networks. Figure 1 displays part of the statistical summaries for the Flickr friendship data given on KONECT (http://konect.uni-koblenz.de/networks/flickr-links), and “power law exponent” corresponds to the power-law index Hill estimate of the degree frequencies.

One possible way to obtain the power-law index estimate is to hypothesize a generative model and within that model estimate parameters using, say, maximum likelihood (MLE). This is done in Gao and van der Vaart (2017) who then plug in the estimated parameter into the theoretical formula for the power-law tail index and by MLE invariance they obtain the tail index estimate. For the directed case, see Wan et al. (2017a). Although the MLE method gives consistent and asymptotically efficient estimates, it requires parametric model correctness and no data corruption; this is hard to guarantee outside of a simulated network. Wan et al. (2017b) show that the MLE approach is less robust against modeling error and data corruption, compared to an estimation method that starts with Hill estimation (Hill 1975) of tail indices coupled with minimum distance threshold selection (Clauset et al. 2009). The methodology used by KONECT to estimate degree frequency indices uses the Hill estimator for each dataset despite the fact that heretofore there has been no theoretical justification for the technique. Hill estimator consistency has been proved only for data from a stationary sequence of random variables, which is assumed to be either iid (Mason 1982) or satisfy structural or mixing assumptions, e.g. Resnick and
Stărică (1995), Resnick and Stărică (1998), Rootzén et al. (1990), and Hsing (1991). Therefore, proving the consistency of the Hill estimator for network data is the focus of this paper.

The Hill estimator and other tail descriptors are often analyzed using the tail empirical estimator. Using standard point measure notation, let

\[
\epsilon_x(A) = \begin{cases} 
1, & \text{if } x \in A, \\
0, & \text{if } x \not\in A.
\end{cases}
\]

For positive iid random variables \( \{X_i : i \geq 1\} \) whose distribution has a regularly varying tail with index \(-\alpha < 0\), we have the following convergence in the space of Radon measures on \((0, \infty)\) of the sequence of empirical measures

\[
\sum_{i=1}^{n} \epsilon_{X_i/b(n)}(\cdot) \Rightarrow \text{PRM}(\nu_\alpha(\cdot)), \quad \text{with} \quad \nu_\alpha(y, \infty] = y^{-\alpha}, y > 0, \quad (1.1)
\]

to the limit Poisson random measure with mean measure \( \nu_\alpha(\cdot) \). Here \( b(n) \) satisfies \( P[X_1 > b(n)] \sim 1/n \). From Eq. 1.1 other extremal properties of \( \{X_n\} \) follow Resnick (1987, Chapter 4.4). See for example the application given in this paper after Theorem 6. Further, for any intermediate sequence \( k_n \to \infty, k_n/n \to 0 \) as \( n \to \infty \), the sequence of tail empirical measures also converges to a deterministic limit,

\[
\hat{\nu}_n(\cdot) := \frac{1}{k_n} \sum_{i=1}^{n} \epsilon_{X_i/b(n/k_n)}(\cdot) \Rightarrow \nu_\alpha(\cdot), \quad (1.2)
\]

which is one way to prove consistency of the Hill estimator for iid data (Resnick 2007, Chapter 4.4). We seek a similar dual pair as Eqs. 1.1 and 1.2 for network models that facilitates the study of the Hill estimator and extremal properties of node degrees.

With this goal in mind, we first find the limiting distribution for the degree sequence in a linear preferential attachment model, from which a similar convergence result to Eq. 1.1 follows. Embedding the network growth model into a continuous time branching process (cf. Athreya 2007; Athreya et al. 2008; Bhamidi 2007) is a useful tool in this case. We model the growth of the degree of each single node as a birth process with immigration. Whenever a new node is added to the network, a new birth immigration process is initiated. In this embedding, the total number of nodes in the network growth model also forms a birth immigration process. Using results from the limit theory of continuous time branching processes (cf. Resnick 1992, Chapter 5.11; Tavaré 1987), we give the limiting distribution of the degree of a fixed node as well as the maximal degree growth. Another way to embed the preferential attachment model is exploited in Gao et al. (2017), using the tree model originated in Rudas et al. (2007). However, our birth immigration process framework gives a more direct way to model the degree growth of each individual node.

Simulation evidence suggests that the Hill estimator applied to undirected network data is consistent, but formally proving the analogue of Eq. 1.2 is challenging and requires showing concentration inequalities for degree counts. We establish the concentration results using the embedding techniques assuming a particular linear preferential attachment model requiring each new node attaches to an existing...
node in the graph. For a more sophisticated model allowing for unconnected graph components, extra techniques, e.g. Stein’s method, must be employed to prove the consistency. Also, the asymptotic distribution of the Hill estimator in the preferential attachment model is not clear and simulation evidence (see Drees et al.) does not rule out non-normality. This non-normality is possibly due to the threshold selection method which chooses the “optimal” portion of the data to compute the Hill estimate (Clauset et al. 2009) but further investigation is left for the future.

Structure of the paper: We review background on the tail empirical measure and Hill estimator in the rest of this section. Section 2 gives the linear preferential attachment model. Section 3 summarizes facts about pure birth and birth-immigration processes. We analyze network degree growth in Section 4 using a sequence of birth-immigration processes and give the limiting empirical measures of normalized degrees in the style of Eq. 1.1. We prove consistency of the Hill estimator for the model in Section 5 and adapt the proof for a more complicated model allowing unconnected components in Section 6. Simulation results are given in Section 7 and they provide insights on the asymptotic behavior of Hill estimators in the preferential attachment model.

Parameter estimation based on maximum likelihood or approximate MLE for directed preferential attachment models is studied in Wan et al. (2017a). A comparison between MLE model based methods and asymptotic extreme value methods is included in Wan et al. (2017b).

1.1 Background

We consider the Hill estimator as a functional of the tail empirical measure so we start with necessary background (cf. Resnick 2007, Chapter 3.3.5).

For $\mathbb{E} = (0, \infty]$, let $M_+(\mathbb{E})$ be the set of non-negative Radon measures on $\mathbb{E}$. A point measure $m$ is an element of $M_+(\mathbb{E})$ of the form $m = \sum_i \epsilon_{x_i}$. The set $M_p(\mathbb{E})$ is the set of all Radon point measures of this form and $M_p(\mathbb{E})$ is a closed subset of $M_+(\mathbb{E})$ in the vague metric.

For $\{X_n, n \geq 1\}$ iid and non-negative with common regularly varying distribution tail $F \in RV_{-\alpha}, \alpha > 0$, there exists a sequence $\{b(n)\}$ such that for a limiting Poisson random measure with mean measure $\nu_{\alpha}$ and $\nu_{\alpha}(y, \infty] = y^{-\alpha}$ for $y > 0$, written as $\text{PRM}(\nu_{\alpha})$, we have (1.1) in $M_p(0, \infty]$, and for some $k_n \to \infty$, $k_n/n \to 0$, we have (1.2) in $M_+(0, \infty]$. Note the limit in Eq. 1.1 is random while that in Eq. 1.2 is deterministic. Define the Hill estimator $H_{k,n}$ based on $k$ upper order statistics of $\{X_1, \ldots, X_n\}$ as in Hill (1975)

$$H_{k,n} := \frac{1}{k} \sum_{i=1}^{k} \log \frac{X(i)}{X(k+1)},$$

where $X(1) \geq X(2) \geq \ldots \geq X(n)$ are order statistics of $\{X_i : 1 \leq i \leq n\}$. In the iid case there are many proofs of consistency (cf. Csörgő et al. 1991a; de Haan and Resnick 1998; Hall 1982; Mason 1982; Mason and Turova 1994): For $k = k_n \to \infty$, $k_n/n \to 0$, we have

$$H_{k_n,n} \overset{p}{\to} 1/\alpha \quad \text{as } n \to \infty. \quad (1.3)$$
We consider $H_{kn,n}$ as a functional of $\hat{\nu}_n$ as in Resnick (2007, Theorem 4.2) who shows (1.3) follows from Eq. 1.2 and we follow this approach for the network context. The next section constructs an undirected preferential attachment model and gives behavior of $D_i(n)$, the degree of node $i$ at the $n$th stage of construction. Theorem 6 shows that for $\delta$, a parameter in the model construction, the degree sequence has empirical measure

$$\sum_{i=1}^{n} \epsilon D_i(n)/n^{1/(2+\delta)}$$

that converges weakly to some random limit point measure in $M_p(0, \infty]$. We prove the analogy to Eq. 1.2 in the network case,

$$\frac{1}{k_n} \sum_{i=1}^{n} \epsilon D_i(n)/b(n/k_n) \Rightarrow \nu_{2+\delta}, \quad \text{in} \ M_+(0, \infty],$$

with the function $b(n) = cn^{1/(2+\delta)}$ and intermediate sequence $k_n$. This leads to consistency for $1/(2+\delta)$ of the Hill estimator $H_{k,n}$ applied to the data $D_1(n), \ldots, D_n(n)$.

2 Preferential attachment models

2.1 Model setup

We consider an undirected preferential attachment model initiated from the initial graph $G(1)$, which consists of one node 1 and a self loop. Node 1 then has degree 2 at stage $n = 1$. For $n \geq 1$, we obtain a new graph $G(n+1)$ by appending a new node $n + 1$ to the existing graph $G(n)$. The graph $G(n)$ consists of $n$ edges and $n$ nodes. Denote the set of nodes in $G(n)$ by $[n] := \{1, 2, \ldots, n\}$. For $i \in [n]$, $D_i(n)$ is the degree of node $i$ in $G(n)$. Given $G(n)$, for a parameter $\delta > -f(1)$, the new node $n + 1$ is connected to one of the existing nodes $i \in [n]$ with probability

$$f(D_i(n)) + \delta \sum_{i \in [n]} (f(D_i(n)) + \delta),$$

where the preferential attachment function $f(j), j \geq 1$ is deterministic and non-decreasing. In this case, the new node $n + 1$ for $n \geq 1$, is always born with degree 1.

Consider three choices of preferential attachment function:

1. **Linear case**: If the preferential attachment function is $f(j) = j$ for $j = 1, 2, \ldots$, then the model is called the linear preferential attachment model. Since every time we add a node and an edge the degree of 2 nodes is increased by 1, we have $\sum_{i=1}^{n} D_i(n) = 2n, n \geq 1$. Therefore, the attachment probability in Eq. 2.1 becomes

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

where $\delta > -1$ is a constant. Degree frequencies are power laws.
2. **Super-linear case:** If \( f \) grows faster than linearly, i.e. \( f(j) = j^\beta \) for \( \beta > 1 \), then there is one node connecting to infinitely many nodes. See Oliveira and Spencer (2005) for a comprehensive study.

3. **Sub-linear case:** If \( f \) grows sub-linearly, i.e. \( f(j) = j^\beta \) for \( 0 < \beta < 1 \), then the degree distribution is much lighter-tailed compared to the linear case. See Bhamidi (2007), Rudas et al. (2007), and Gao et al. (2017) for references.

Due to our interest in power laws, we only consider the linear case.

### 2.2 Power-law tails

Continuing with \( f(j) = j \), suppose \( G(n) \) is a random graph generated after \( n \) steps. Let \( N_k(n) \) be the number of nodes in \( G(n) \) with degree equal to \( k \), i.e.

\[
N_k(n) := \sum_{i=1}^{n} 1_{\{D_i(n) = k\}},
\]

then \( N_{>k}(n) := \sum_{j>k} N_j(n) \), \( k \geq 1 \), is the number of nodes in \( G(n) \) with degree strictly greater than \( k \). For \( k = 0 \), we set \( N_{>0}(n) = n \).

Following the proof technique in van der Hofstad (2017, Theorem 8.3), which uses concentration inequalities and martingale methods, we have for fixed \( k \geq 1 \), as \( n \to \infty \),

\[
\frac{N_k(n)}{n} \xrightarrow{p} p_k = \frac{(2+\delta)\Gamma(k+\delta)\Gamma(3+2\delta)}{(k+3+2\delta)\Gamma(1+\delta)} \sim \frac{(2+\delta)\Gamma(3+2\delta)}{\Gamma(1+\delta)} k^{-(3+\delta)}; \quad (2.3)
\]

\((p_k)_{k \geq 0}\) is a probability mass function (pmf) and the asymptotic form, as \( k \to \infty \), follows from Stirling’s formula. Let \( p_{>k} = \sum_{j>k} p_j \) be the complementary cumulative distribution function (cdf). By Scheffé’s lemma as well as van der Hofstad (2017, Equation (8.4.6)), we have

\[
\frac{N_{>k}(n)}{n} \xrightarrow{p} p_{>k} := \frac{\Gamma(k+1+\delta)\Gamma(3+2\delta)}{\Gamma(k+3+2\delta)\Gamma(1+\delta)}, \quad (2.4)
\]

and again by Stirling’s formula we get from Eq. 2.4 as \( k \to \infty \),

\[
p_{>k} \sim c \cdot k^{-(2+\delta)}, \quad c = \frac{\Gamma(3+2\delta)}{\Gamma(1+\delta)}.
\]

In other words, the tail distribution of the asymptotic degree sequence in a linear preferential attachment model is asymptotic to a power law with tail index \( 2+\delta \). Since the Hill estimator is widely used to estimate this tail index, we prove consistency of this estimator.

### 3 Preliminaries: continuous time Markov branching processes

We now review two continuous time Markov branching processes. We will embed \( D_i(n), i \in [n] \) in a constructed process and obtain the network asymptotics from the constructed process.

\( \copyright \) Springer
3.1 Linear birth processes

A linear birth process \( \{ \zeta(t) : t \geq 0 \} \) is a continuous time Markov process taking values in the set \( \mathbb{N}^+ = \{1, 2, 3, \ldots \} \) and having a transition rate

\[
q_{i,i+1} = \lambda i, \quad i \in \mathbb{N}^+, \quad \lambda > 0.
\]

The process \( \{ \zeta(t) : t \geq 0 \} \) is a mixed Poisson process; see Resnick (1992, Theorem 5.11.4), Kendall (1966) and Waugh (1970) among other sources. If \( \zeta(0) = 1 \) then the representation is

\[
\zeta(t) = 1 + N_0(W(e^{\lambda t} - 1), t \geq 0), \quad \text{where} \quad N_0(t) = \sum_{i \geq 1} 1_{\{t \geq \tau_i\}}
\]

(3.1)

where \( \{N_0(t) : t \geq 0\} \) is a unit rate homogeneous Poisson on \( \mathbb{R}_+ \) with \( N_0(0) = 0 \) and \( W \perp N_0(\cdot) \) is a unit exponential random variable independent of \( N_0 \). Since \( N_0(t)/t \to 1 \) almost surely as \( t \to \infty \), it follows immediately that

\[
\frac{\zeta(t)}{e^{\lambda t}} \xrightarrow{a.s.} W, \quad \text{as} \quad t \to \infty.
\]

(3.2)

We use these facts in Section 4.2 to analyze the asymptotic behavior of the degree growth in a preferential attachment network.

3.2 Birth processes with immigration

Apart from individuals within the population giving birth to new individuals, population size can also increase due to immigration which is assumed independent of births. The linear birth process with immigration (B.I. process), \( \{BI(t) : t \geq 0\} \), having lifetime parameter \( \lambda > 0 \) and immigration parameter \( \theta \geq 0 \) is a continuous time Markov process with state space \( \mathbb{N} = \{0, 1, 2, 3, \ldots \} \) and transition rate

\[
q_{i,i+1} = \lambda i + \theta.
\]

When \( \theta = 0 \) there is no immigration and the B.I. process becomes a pure birth process.

For \( \theta > 0 \), the B.I. process starting from 0 can be constructed from a Poisson process and an independent family of iid linear birth processes (Tavaré 1987). Suppose that \( N_\theta(t) \) is the counting function of homogeneous Poisson points \( 0 < \tau_1 < \tau_2 < \ldots \) with rate \( \theta \) and independent of this Poisson process we have independent copies of a linear birth process \( \{\zeta_i(t) : t \geq 0\}_{i \geq 1} \) with parameter \( \lambda > 0 \) and \( \zeta_i(0) = 1 \) for \( i \geq 1 \). Let \( BI(0) = 0 \), then the B.I. process is a shot noise process with form

\[
BI(t) := \sum_{i=1}^{\infty} \zeta_i(t-\tau_i)1_{\{t \geq \tau_i\}} = \sum_{i=1}^{N_\theta(t)} \zeta_i(t-\tau_i).
\]

(3.3)

Theorem 1 modifies slightly the statement of Tavaré (1987, Theorem 5) summarizing the asymptotic behavior of the B.I. process.
Theorem 1: For \( BI(t) : t \geq 0 \) as in Eq. 3.3, we have as \( t \to \infty \),
\[
e^{-\lambda t} BI(t) \xrightarrow{a.s.} \sum_{i=1}^{\infty} W_i e^{-\lambda \tau_i} =: \sigma
\] (3.4)
where \( \{W_i : i \geq 1\} \) are independent unit exponential random variables which for each \( i \geq 1 \) are almost sure limits,
\[
W_i = \lim_{t \to \infty} e^{-t} \zeta_i(t).
\]
The random variable \( \sigma \) in Eq. 3.4 is a.s. finite and has a Gamma density given by
\[
f(x) = \frac{1}{\Gamma(\theta/\lambda)} x^{\theta/\lambda - 1} e^{-x}, \quad x > 0.
\]
The form of \( \sigma \) in Eq. 3.4 and its Gamma density is justified in Tavaré (1987). It can be guessed from Eq. 3.3 and some cavalier interchange of limits and infinite sums. Transforming Poisson points \( \{(W_i, \tau_i), i \geq 1\} \), summing and recognizing a Gamma Lévy process at \( t = 1 \) gives the density of \( \sigma \).

Remark 2: For a B.I. process \( \{BI(t)\}_{t \geq 0} \) with \( BI(0) = j \geq 1 \), modifying the representation in Eq. 3.3 gives
\[
BI(t) = \sum_{i=1}^{j} \zeta_i(t) + \sum_{i=j+1}^{\infty} \zeta_i(t - \tau_i) 1_{\{t \geq \tau_i\}}.
\]
Therefore, \( e^{-\lambda t} BI(t) \xrightarrow{a.s.} \sigma' \) where \( \sigma' \) has a Gamma density given by \( g(x) = x^{j + \theta/\lambda - 1} e^{-x} / \Gamma(j + \theta/\lambda), x > 0 \).

4 Embedding process

Our approach to the weak convergence of the sequence of empirical measures in Eq. 1.4 embeds the degree sequences \( \{D_i(n), 1 \leq i \leq n, n \geq 1\} \) into a process constructed from B.I. processes. The embedding idea is proposed in Athreya et al. (2008) and we tailor it for our setup finding it flexible enough to accommodate the linear preferential attachment model introduced in Section 2.1.

4.1 Embedding

Here is how we embed the network growth model using a sequence of independent B.I. processes.

4.1.1 Model construction

Let \( \{BI_i(t) : t \geq 0\}_{i \geq 1} \) be independent B.I. processes such that
\[
BI_1(0) = 2, \quad BI_i(0) = 1, \quad \forall i \geq 2.
\] (4.1)
Each has transition rate is $q_{j,j+1} = j + \delta, \delta > -1$. For $i \geq 1$, let $\{\tau_k^{(i)} : k \geq 1\}$ be the jump times of $\{BI_i(t) : t \geq 0\}$ and set $\tau_0^{(i)} := 0$ for all $i \geq 1$. For $k \geq 1$,

$$BI_1(\tau_k^{(1)}) = k + 2, \quad BI_i(\tau_k^{(i)}) = k + 1, \quad i \geq 2.$$  

Therefore,

$$\tau_k^{(1)} - \tau_{k-1}^{(1)} \sim \text{Exp}(k + 1 + \delta), \quad \text{and} \quad \tau_k^{(i)} - \tau_{k-1}^{(i)} \sim \text{Exp}(k + \delta), \quad i \geq 2.$$  

and $\{\tau_k^{(i)} - \tau_{k-1}^{(i)} : i \geq 1, k \geq 1\}$ are independent.

Set $T_1 = 0$ and relative to $BI_1(\cdot)$ define

$$T_2 := \tau_1^{(1)}, \quad (4.2)$$

i.e. the first time that $BI_1(\cdot)$ jumps. Start the new B.I. process $\{BI_2(t - T_2) : t \geq T_2\}$ at $T_2$ and let $T_3$ be the first time after $T_2$ that either $BI_1(\cdot)$ or $BI_2, (\cdot)$ jumps so that,

$$T_3 = \min\{T_i + \tau_k^{(i)} : k \geq 1, T_i + \tau_k^{(i)} > T_2, i = 1, 2\}.$$  

Start a new, independent B.I. process $\{BI_3(t - T_3)\}_{t \geq T_3}$ at $T_3$. See Fig. 2, which assumes $\tau_1^{(2)} + T_2 < \tau_2^{(1)}$. Continue in this way. When $n$ lines have been created, define $T_{n+1}$ to be the first time after $T_n$ that one of the processes $\{BI_i(t - T_i) : t \geq T_i\}_{1 \leq i \leq n}$ jumps, i.e.

$$T_{n+1} := \min\{T_i + \tau_k^{(i)} : k \geq 1, T_i + \tau_k^{(i)} > T_n, 1 \leq i \leq n\}. \quad (4.3)$$  

At $T_{n+1}$, start a new, independent B.I. process $\{BI_{n+1}(t - T_{n+1})\}_{t \geq T_{n+1}}$.

### 4.1.2 Embedding

The following embedding theorem is similar to the one proved in Athreya et al. (2008) and summarizes how to embed in the B.I. construction.

**Theorem 3** Fix $n \geq 1$. Suppose

$$D(n) := (D_1(n), \ldots, D_n(n))$$

| $BI_1(0)$ | $BI_1(T_2^A)$ | $BI_2(0)$ | $BI_2(T_3^A)$ |
|-----------|----------------|-----------|----------------|
| $= 2$     | $= 3$          | $= 1$     | $= 3$          |

Fig. 2 Embedding procedure for Model A assuming $\tau_1^{(2)} + T_2^A < \tau_2^{(1)}$
is the degree sequence of nodes in the graph $G(n)$ and \{${T_n}$\}$_{n \geq 1}$ is defined as in Eq. 4.3. For each fixed $n$, define

$$\tilde{D}(n) := (BI_1(T_n), BI_2(T_n - T_2), \ldots, BI_{n-1}(T_n - T_{n-1}), BI_n(0)),$$

and then $D(n)$ and $\tilde{D}(n)$ have the same distribution in $\mathbb{R}^n$.

**Proof** By the model construction, at each $T_n$, $n \geq 2$, we start a new B.I. process $BI_n(\cdot)$ with initial value equal to 1 and one of $BI_i$, $1 \leq i \leq n - 1$ also increases by 1. This makes the sum of the values of $BI_i$, $1 \leq i \leq n$, increase by 2 so that

$$\sum_{i=1}^n (BI_i(T_n - T_i) + \delta) = (2 + \delta)n.$$ 

The rest is essentially the proof of Athreya et al. (2008, Theorem 2.1) which we now outline.

Both $\{D(n), n \geq 1\}$ and $\{\tilde{D}(n), n \geq 1\}$ are Markov on the state space $\cup_{n \geq 1} \mathbb{R}^n_+$ since

$$D(n+1) = (D(n), 1) + (e_{j+1}^{(n)}, 0),$$

$$\tilde{D}(n+1) = (\tilde{D}(n), 1) + (e_{j+1}^{(n)}, 0),$$

where for $n \geq 1$, $e_j^{(n)}$ is a vector of length $n$ of 0’s except for a 1 in the $j$-th entry and

$$P[J_{n+1} = j | D(n)] = \frac{D_j(n) + \delta}{(2 + \delta)n}, \quad 1 \leq j \leq n,$$

and $L_{n+1}$ records which B.I. process in $\{BI_i(t - T_i) : t \geq T_i\}_{1 \leq i \leq n}$ is the first to have a new birth after $T_n$.

When $n = 1$,

$$\tilde{D}(1) = BI_1(0) = 2 = D_1(1) = D(1),$$

so to prove equality in distribution for any $n$, it suffices to verify that the transition probability from $\tilde{D}(n)$ to $\tilde{D}(n + 1)$ is the same as that from $D(n)$ to $D(n + 1)$.

According to the preferential attachment setup, we have

$$P(D(n + 1) = (d_1, d_2, \ldots, d_i + 1, d_{i+1}, \ldots, d_n, 1) | D(n) = (d_1, d_2, \ldots, d_n)) = \frac{d_i + \delta}{(2 + \delta)n}, \quad 1 \leq i \leq n.$$  \hspace{1cm} (4.4)

At time $T_n$, there are $n$ B.I. processes and each of them has a population size of $BI_i(T_n - T_i)$, $1 \leq i \leq n$. Therefore, $T_{n+1} - T_n$ is the minimum of $n$ independent exponential random variables, $\{E_n^{(i)}\}_{1 \leq i \leq n}$, with means

$$(BI_i(T_n - T_i) + \delta)^{-1}, \quad 1 \leq i \leq n,$$
which gives for any $1 \leq i \leq n$,
\[
P \left( L_{n+1} = i \mid \tilde{D}(n) = (d_1, d_2, \ldots, d_n) \right)
= P \left( \tilde{D}(n + 1) = (d_1, d_2, \ldots, d_i + 1, d_{i+1}, \ldots, d_n, 1) \mid \tilde{D}(n) = (d_1, \ldots, d_n) \right)
= P \left( E_n^{(i)} < \bigwedge_{j=1, j \neq i}^{n} E_n^{(j)} \mid \tilde{D}(n) = (d_1, d_2, \ldots, d_n) \right)
= \frac{B I_i (T_n - T_i) + \delta}{\sum_{i=1}^{n} (B I_i (T_n - T_i) + \delta)} = \frac{d_i + \delta}{(2 + \delta)n}.
\]
This agrees with the transition probability in Eq. 4.4, thus completing the proof. □

Remark 4 This B.I. process construction can also be generalized for other choices of the preferential attachment functions $f$. For example, its applications to the super- and sub-linear preferential attachment models are studied in Athreya (2007).

4.2 Asymptotic properties

One important reason to use the embedding technique specified in Section 4.1 is that asymptotic behavior of the degree growth in a preferential attachment model can be characterized explicitly. These asymptotic properties then help us derive weak convergence of the empirical measure, which is analogous to Eq. 1.1 in the iid case.

4.2.1 Branching times

We first consider the asymptotic behavior of the branching times $\{T_n\}_{n \geq 1}$, which are defined in Section 4.1.

**Proposition 5** For $\{T_n\}_{n \geq 1}$ defined in Eq. 4.3, we have
\[
\frac{n}{e^{(2+\delta)T_n}} \overset{a.s.}{\longrightarrow} W, \quad W \sim \text{Exp} (1) . \tag{4.5}
\]

**Proof** Define a counting processes
\[
N(t) := \frac{1}{2} \sum_{i=1}^{\infty} B I_i (t - T_i) 1_{\{t \geq T_i\}}.
\]
Then we have
\[
N(t) 1_{\{t \in [T_n, T_{n+1})\}} = n.
\]
In other words, $\{T_n\}_{n \geq 1}$ are the jump times of the counting process $N(\cdot)$, with the following structure
\[
\{T_{n+1} - T_n : n \geq 1\} \overset{d}{=} \left\{ \frac{E_i}{(2 + \delta)i}, i \geq 1 \right\}, \tag{4.6}
\]
where $\{E_i : i \geq 1\}$ are iid unit exponential random variables.
From Eq. 4.6, we see that $N(\cdot)$ is a pure birth process with $N(0) = 1$ and transition rate

$$q_{i,i+1} = (2 + \delta)i, \quad i \geq 1.$$  

Replacing $t$ with $T_n$ in Eq. 3.2 gives (4.5).

4.2.2 Convergence of the measure

Using embedding techniques, Theorem 6 gives convergence of empirical measures of scaled node degrees, which is the analogue of Eq. 1.1 for the iid case.

**Theorem 6** Suppose that

1. $\{T_i : i \geq 1\}$ are distributed as in Eq. 4.6.
2. $W$ is the limit random variable as given in Eq. 4.5.
3. $\{\sigma_i\}_{i \geq 1}$ is a sequence of independent Gamma random variables specified in Eqs. 4.12 and 4.10 below.

Then in $M_p((0, \infty))$, we have for $\delta \geq 0$,

$$\sum_{i=1}^{n} \varepsilon_{D_i(n)/n^{1/(2+\delta)}}(\cdot) \Rightarrow \sum_{i=1}^{\infty} \varepsilon_{\sigma_i e^{-T_i}/W^{1/(2+\delta)}}(\cdot).$$  

(4.7)

**Remark 7** From Eq. 4.7 we get for any fixed $k \geq 1$, that in $\mathbb{R}_+^k$,

$$\left(\frac{D_1(n)}{n^{1/(2+\delta)}}, \ldots, \frac{D_k(n)}{n^{1/(2+\delta)}}\right) \Rightarrow W^{-1/(2+\delta)}\left((\sigma_1 e^{-T_1})_{(1)}, \ldots, (\sigma_k e^{-T_k})_{(k)}\right),$$  

(4.8)

where a subscript inside parentheses indicates ordering so that $D_{(1)}(n) \geq \cdots \geq D_{(k)}$ and the limit on the right side of Eq. 4.8 represents the ordered $k$ largest points from the right side of Eq. 4.7.

To prove Theorem 6, we first show the following lemma, which gives the asymptotic limit of the degree sequence.

**Lemma 8** Suppose that

1. $\{T_i : i \geq 1\}$ are distributed as in Eq. 4.6.
2. $W$ is the limit random variable as given in Eq. 4.5.

Then the degree sequence $\{D_i(n) : 1 \leq i \leq n\}$ satisfies:

(i) For each $i \geq 1$,

$$\frac{D_i(n)}{n^{1/(2+\delta)}} \Rightarrow \frac{\sigma_i e^{-T_i}}{W^{1/(2+\delta)}},$$  

(4.9)

where $\{\sigma_i\}_{i \geq 1}$ are a sequence of independent Gamma random variables with

$$\sigma_1 \sim \text{Gamma}(2 + \delta, 1), \quad \text{and} \quad \sigma_i \sim \text{Gamma}(1 + \delta, 1), \quad i \geq 2.$$  

(4.10)

Furthermore, for $i \geq 1$, $\sigma_i$ is independent from $e^{-T_i}$.  

© Springer
(ii) For $\delta > -1,$
\[
\max_{i \geq 1} \frac{D_i(n)}{n^{1/(2+\delta)}} \Rightarrow W^{-1/(2+\delta)} \max_{i \geq 1} \sigma_i e^{-T_i},
\]
where we set $D_i(n) := 0$ for all $i \geq n + 1.$

Proof (i) Applying the results in Remark 2 gives that as $t \to \infty,$
\[
\frac{BI_i(t - T_i)}{e^{t - T_i}} \xrightarrow{a.s.} \sigma_i, \quad i \geq 1,
\]
where $\{\sigma_i\}_{i \geq 1}$ are independent Gamma random variables with
\[
\sigma_1 \sim \text{Gamma}(2 + \delta, 1) \quad \text{and} \quad \sigma_i \sim \text{Gamma}(1 + \delta, 1), \quad i \geq 2.
\]
Thus as $n \to \infty,$
\[
\frac{BI_i(T_n - T_i)}{e^{T_n - T_i}} \xrightarrow{a.s.} \sigma_i, \quad i \geq 1.
\]
Combining (4.12) with (4.5), we have for fixed $1 \leq i \leq n,$
\[
\frac{BI_i(T_n - T_i)}{n^{1/(2+\delta)}} \xrightarrow{a.s.} \frac{\sigma_i e^{-T_i}}{W^{1/(2+\delta)}},
\]
Then (4.9) follows from Theorem 3. For $i \geq 2,$ the independence of $\sigma_i$ and $T_i$ follows from the construction and this completes the proof of (i).

(ii) For $i \geq n + 1,$ $BI_i(T_n - T_i) = 0$ so from Theorem 3, it suffices to show
\[
\max_{i \geq 1} \frac{BI_i(T_n - T_i)}{n^{1/(2+\delta)}} \xrightarrow{a.s.} \max_{i \geq 1} \frac{\sigma_i e^{-T_i}}{W^{1/(2+\delta)}},
\]
which is proved in Athreya et al. (2008, Theorem 1.1(iii)).

Using Lemma 8, we prove Theorem 6.

Proof of Theorem 6 Note that the limit random variables
\[
\sigma_i e^{-T_i} W^{-1/(2+\delta)}, \quad i \geq 1,
\]
have continuous distributions, so for any $y > 0,$
\[
P \left( \sum_{i=1}^{\infty} \epsilon_{\sigma_i e^{-T_i} W^{1/(2+\delta)}}(\{y\}) = 0 \right) = 1.
\]
Hence, by Kallenberg’s theorem for weak convergence to a point process on an interval (see Kallenberg 2017, Theorem 4.18 and Resnick 1987, Proposition 3.22), proving (4.7) requires checking
(a) For $y > 0,$ as $n \to \infty,$
\[
E \left( \sum_{i=1}^{n} \epsilon_{\frac{D_i(n)}{n^{1/(2+\delta)}}(y, \infty)} \right) \to E \left( \sum_{i=1}^{\infty} \epsilon_{\sigma_i e^{-T_i} W^{1/(2+\delta)}}(y, \infty) \right).
\]
(b) For $y > 0$, as $n \to \infty$,

$$
P \left( \sum_{i=1}^{n} \epsilon D_i(n)/n^{1/(2+\delta)} (y, \infty] = 0 \right)
\longrightarrow P \left( \sum_{i=1}^{\infty} \epsilon \sigma_i e^{-T_i/W^{1/(2+\delta)}} (y, \infty] = 0 \right).
$$

To show (4.13), first note that for any $M > 0$,

$$
E \left( \sum_{i=1}^{M} \epsilon D_i(n)/n^{1/(2+\delta)} (y, \infty] \right) = \sum_{i=1}^{M} P \left( D_i(n)/n^{1/(2+\delta)} > y \right)
\longrightarrow \sum_{i=1}^{M} P \left( \sigma_i e^{-T_i/W^{1/(2+\delta)}} > y \right)
= E \left( \sum_{i=1}^{M} \epsilon \sigma_i e^{-T_i/W^{1/(2+\delta)}} (y, \infty] \right),
$$

as $n \to \infty$. By Chebyshev’s inequality we have for any $k > 2 + \delta$,

$$
E \left( \sum_{i=M+1}^{n} \epsilon D_i(n)/n^{1/(2+\delta)} (y, \infty] \right) = \sum_{i=M+1}^{n} P \left( D_i(n)/n^{1/(2+\delta)} > y \right)
\leq y^{-k} \sum_{i=M+1}^{n} E \left[ \left( \frac{D_i(n)}{n^{1/(2+\delta)}} \right)^k \right].
$$

Also, we have for $\delta \geq 0$,

$$
E \left[ \left( \frac{D_i(n)}{n^{1/(2+\delta)}} \right)^k \right] \leq E \left[ \left( \frac{D_i(n) + \delta}{n^{1/(2+\delta)}} \right)^k \right].
$$

Following a similar argument that leads to van der Hofstad (2017, Equation (8.7.26)), we have

$$
E \left[ \left( \frac{D_i(n) + \delta}{n^{1/(2+\delta)}} \right)^k \right] \leq E \left[ \left( \frac{\sigma_i e^{-T_i/W^{1/(2+\delta)}}}{W^{1/(2+\delta)}} \right)^k \right] \sim C_{k,\delta} i^{-\frac{k}{2+\delta}},
$$

for $i$ large and $C_{k,\delta} > 0$. Hence, continuing from Eq. 4.15, we have

$$
E \left( \sum_{i=M+1}^{n} \epsilon D_i(n)/n^{1/(2+\delta)} (y, \infty] \right) \leq y^{-k} \sum_{i=M+1}^{n} E \left[ \left( \frac{D_i(n)}{n^{1/(2+\delta)}} \right)^k \right]
\leq y^{-k} \sum_{i=M+1}^{\infty} E \left[ \left( \frac{\sigma_i e^{-T_i}}{W^{1/(2+\delta)}} \right)^k \right],
$$

since $k/(2 + \delta) > 1$. This verifies Condition (a).
To see Eq. 4.14, we have
\[
\left\{ \sum_{i=1}^{n} \epsilon_{D_i(n)/n^{1/(2+\delta)}}(y, \infty) = 0 \right\} = \left\{ \frac{D_i(n)}{n^{1/(2+\delta)}} \leq y, 1 \leq i \leq n \right\} \\
= \left\{ \max_{1 \leq i \leq n} \frac{D_i(n)}{n^{1/(2+\delta)}} \leq y \right\}.
\]

Since we set \( D_i(n) = 0 \) for all \( i \geq n + 1 \), then
\[
\left\{ \max_{1 \leq i \leq n} \frac{D_i(n)}{n^{1/(2+\delta)}} \leq y \right\} = \left\{ \max_{i \geq 1} \frac{D_i(n)}{n^{1/(2+\delta)}} \leq y \right\}.
\]

Similarly,
\[
\left\{ \sum_{i=1}^{\infty} \epsilon_{\sigma_i e^{-T_i/w^{1/(2+\delta)}}}(y, \infty) = 0 \right\} = \left\{ \max_{i \geq 1} \frac{\sigma_i e^{-T_i}}{w^{1/(2+\delta)}} \leq y \right\}.
\]

By Eq. 4.11, we have for \( y > 0 \),
\[
P\left( \max_{i \geq 1} \frac{D_i(n)}{n^{1/(2+\delta)}} \leq y \right) \to P\left( \max_{i \geq 1} \frac{\sigma_i e^{-T_i}}{w^{1/(2+\delta)}} \leq y \right), \quad \text{as } n \to \infty,
\]
which gives (4.14) and completes the proof of (iv).

5 Consistency of the Hill estimator

We now turn to Eq. 1.5 as preparation for considering consistency of the Hill estimator. We first give a plausibility argument based on the form of the limit point measure in Eq. 4.7. However, proving (1.5) requires showing \( N_{>k}(n)/n \) concentrates on \( p_{>k} \), for all \( k \geq 1 \), which in other words means controlling the bias for \( N_{>k}(n)/n \) and the discrepancy between \( E(N_{>k}(n)/n) \) and \( p_{>k} \).

5.1 Heuristics

Before starting formalities, here is a heuristic explanation for the consistency of the Hill estimator when applied to preferential attachment data. Since the Gamma random variables \( \sigma_i \) have light tailed distributions, one may expect that \( \{\sigma_i : i \geq 1\} \) will not distort the consistency result and so we pretend the \( \sigma_i \)'s are absent; then what remains in the limit points is monotone in \( i \). Set \( Y_i := e^{-T_i}/w^{1/(2+\delta)} \) and apply the Hill estimator to the \( Y_i \)’s to get
\[
H_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \log \left( \frac{Y_i}{Y_{k+1}} \right) = \frac{1}{k} \sum_{i=1}^{k} (T_{k+1} - T_i).
\]

Recall from Eq. 4.6 that
\[
T_{n+1} - T_n \overset{d}{=} E_n/(n(2 + \delta)),
\]
where $E_n$, $n \geq 1$ are iid unit exponential random variables. Then
\[
H_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \sum_{l=i}^{k} (T_{l+1} - T_l) = \frac{1}{k} \sum_{l=1}^{k} l(T_{l+1} - T_l) = \frac{1}{k} \sum_{l=1}^{k} E_l 2 + \delta \xrightarrow{a.s.} \frac{1}{2 + \delta},
\]
by strong law of large numbers, provided that $k \to \infty$.

There are clear shortcomings to this approach, the most obvious being that we only dealt with the points at asymptopia rather than $\{D_i(n), 1 \leq i \leq n\}$. Furthermore we simplified the limit points by neglecting the $\sigma_i$'s. We have not found an effective way to analyze order statistics of $\{\sigma_i e^{-T_i/W^{1/(2+\delta)}} : i \geq 1\}$.

Concentration results for degree counts provide a traditional tool to prove (1.5) and we pursue this in the next subsection.

\section*{5.2 Convergence of the tail empirical measure}

We now analyze the convergence of the tail empirical measure. First consider the degree of each node in $G(n)$,
\[
(D_1(n), D_2(n), \ldots, D_n(n)),
\]
and let
\[
D_{(1)}(n) \geq D_{(2)}(n) \geq \cdots \geq D_{(n)}(n)
\]
be the corresponding order statistics. Then the tail empirical measure becomes
\[
\hat{\nu}_n(\cdot) := \frac{1}{k_n} \sum_{i=1}^{n} \epsilon_{D_i(n)/D_{(k_n)}(n)}(\cdot),
\]
for some intermediate sequence $\{k_n\}$, i.e. $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.

**Theorem 9** Suppose that $\{k_n\}$ is some intermediate sequence satisfying
\[
\liminf_{n \to \infty} k_n/(n \log n)^{1/2} > 0 \quad \text{and} \quad k_n/n \to 0 \quad \text{as} \quad n \to \infty,
\]
then
\[
\hat{\nu}_n \Rightarrow \nu_{2+\delta},
\]
in $M_+((0, \infty])$, where $\nu_{2+\delta}(x, \infty] = x^{-(2+\delta)}$, $x > 0$.

**Proof** We proceed in a series of steps.

**Step 1.** We first show that with
\[
b(n/k_n) = \left( \frac{\Gamma(3 + 2\delta)}{\Gamma(1 + \delta)} \right)^{1/2 + \delta} (n/k_n)^{1/2 + \delta},
\]
we have in $M_+((0, \infty])$,
\[
\frac{1}{k_n} \sum_{i=1}^{n} \epsilon_{D_i(n)/b(n/k_n)} \Rightarrow \nu_{2+\delta},
\]
(5.3)
and it suffices to justify for any \( y > 0 \),

\[
\left| \frac{1}{k_n} N_{> [b(n/k_n)y]}(n) - y^{-(2+\delta)} \right| \xrightarrow{P} 0, \quad n \to \infty. \tag{5.4}
\]

The left side of Eq. 5.4 is bounded by

\[
\begin{aligned}
&\left| \frac{1}{k_n} N_{> [b(n/k_n)y]}(n) - y^{-(2+\delta)} \right| \\
\leq & \left| \frac{1}{k_n} N_{> [b(n/k_n)y]}(n) - \mathbb{E}(N_{> [b(n/k_n)y]}(n)) \right| \\
&+ \left| \frac{1}{k_n} \mathbb{E}(N_{> [b(n/k_n)y]}(n)) - np_{> [b(n/k_n)y]} \right| \\
&+ \left| \frac{n}{k_n} p_{> [b(n/k_n)y]} - y^{-(2+\delta)} \right| \\
= &: I + II + III. \tag{5.5}
\end{aligned}
\]

Using Stirling’s formula, van der Hofstad (2017, Equation 8.3.9) gives

\[
\frac{\Gamma(t + a)}{\Gamma(t)} = t^a (1 + O(1/t)).
\]

Recall the definition of \( p_{> k} \) in Eq. 2.4 for fixed \( k \), then we have

\[
\begin{aligned}
n \frac{k_n}{n} p_{> [b(n/k_n)y]} &= \frac{n}{k_n} \Gamma(3 + 2\delta) \Gamma([b(n/k_n)y] + 1 + \delta) \\
&\quad \times \frac{1}{\Gamma(3 + 2\delta) k_n} (b(n/k_n)y)^{-(2+\delta)} \left( 1 + O \left( \frac{1}{b(n/k_n)} \right) \right) \\
&= y^{-(2+\delta)} \left( 1 + O \left( \frac{1}{b(n/k_n)} \right) \right). \tag{5.6}
\end{aligned}
\]

Hence, \( III \to 0 \) as \( n \to \infty \).

Consider \( I \) and we have for \( \epsilon > 0 \),

\[
P \left( \left| \frac{1}{k_n} N_{> [b(n/k_n)y]}(n) - \mathbb{E}(N_{> [b(n/k_n)y]}(n)) \right| > \epsilon \right)
= P \left( \left| N_{> [b(n/k_n)y]}(n) - \mathbb{E}(N_{> [b(n/k_n)y]}(n)) \right| > \epsilon k_n \right).
\]

Following the proof in van der Hofstad (2017, Proposition 8.4), we have for any \( C > 2\sqrt{2} \),

\[
P \left( \left| N_{> k}(n) - \mathbb{E}(N_{> k}(n)) \right| \geq C \sqrt{n \log n} \right) = o(1/n).
\]

Since \( N_{> k}(n) = 0 \) a.s. for all \( k > n \), then

\[
P \left( \max_k \left| N_{> k}(n) - \mathbb{E}(N_{> k}(n)) \right| \geq C \sqrt{n \log n} \right)
= P \left( \max_{0 \leq k \leq n} \left| N_{> k}(n) - \mathbb{E}(N_{> k}(n)) \right| \geq C \sqrt{n \log n} \right)
\leq \sum_{k=1}^{n} P \left( \left| N_{> k}(n) - \mathbb{E}(N_{> k}(n)) \right| \geq C \sqrt{n \log n} \right) = o(1). \tag{5.7}
\]
Therefore, for \( \{k_n\} \) satisfying (5.1), we have
\[
\mathbf{P}\left( |N_{>\lfloor b(n/k_n)y \rfloor}(n) - \mathbb{E}(N_{>\lfloor b(n/k_n)y \rfloor}(n))| > \epsilon k_n \right) \\
\leq \mathbf{P}\left( \max_k |N_{>k}(n) - \mathbb{E}(N_{>k}(n))| \geq \epsilon k_n \right) = o(1),
\]
which gives \( I \to P \).

Then we are left to show \( II \to 0 \) as \( n \to \infty \). Let \( U_n \) be a uniform random variable on \( \{1, 2, \ldots, n\} \), then
\[
\frac{1}{k_n} \mathbb{E}(N_{>\lfloor b(n/k_n)y \rfloor}(n)) = \frac{1}{k_n} \sum_{i=1}^{n} \mathbf{P}(D_i(n) > \lfloor b(n/k_n)y \rfloor) = \frac{n}{k_n} \mathbb{E}\left( \mathbf{P}(D_{U_n}(n) > \lfloor b(n/k_n)y \rfloor) \right).
\]
Let \( B_a(p) \) be a negative binomial integer valued random variable with parameters \( a > 0 \) and \( p \in (0, 1) \) (abbreviated as \( \text{NB}(a, p) \)), and the generating function of \( B_a(p) \) is
\[
\mathbb{E}\left( s^{B_a(p)} \right) = \left( s + \frac{(1-s)}{p} \right)^{-a}, \quad 0 \leq s \leq 1.
\]
Suppose that \( \{B^{(i)}_{1+\delta}(p) : i \geq 1\} \) is a sequence of iid \( \text{NB}(1+\delta, p) \) random variables and \( B_{2+\delta}(p) \) is another \( \text{NB}(2+\delta, p) \) random variable independent from \( \{B^{(i)}_{1+\delta}(p) : i \geq 1\} \). Then by the B.I. process construction, we have for \( k, t \geq 0 \),
\[
\mathbf{P}(B(1)(t) > k) = \mathbf{P}\left[ 2 + B_{2+\delta}(e^{-t}) > k \right] \\
\mathbf{P}(B(1)(t) > k) = \mathbf{P}\left[ 1 + B^{(i)}_{1+\delta}(e^{-t}) > k \right], \quad i \geq 2.
\]
Therefore, applying the embedding technique gives
\[
\frac{n}{k_n} \mathbb{E}(\mathbf{P}(D_{U_n}(n) > \lfloor b(n/k_n)y \rfloor)) = \frac{n}{k_n} \sum_{i=1}^{n} \mathbf{P}(B(1)(T_n - T_i) > \lfloor b(n/k_n)y \rfloor) \\
= \frac{n}{k_n} \sum_{i=1}^{n} \mathbf{P}\left[ 1 + B^{(i)}_{1+\delta}(e^{-(T_n - T_i)}) > \lfloor b(n/k_n)y \rfloor \right] \\
+ \frac{1}{k_n} \left( \mathbf{P}(B(1)(T_n) > \lfloor b(n/k_n)y \rfloor) - \mathbb{P}\left[ 1 + B^{(i)}_{1+\delta}(e^{-T_n}) > \lfloor b(n/k_n)y \rfloor \right] \right) \\
= \frac{n}{k_n} \mathbb{E}\left( \mathbf{P}\left[ 1 + B^{(U_n)}_{1+\delta}(e^{-(T_n - T_{U_n})}) > \lfloor b(n/k_n)y \rfloor \right] \right) \\
+ \frac{1}{k_n} \left( \mathbf{P}(B(1)(T_n) > \lfloor b(n/k_n)y \rfloor) - \mathbb{P}\left[ 1 + B^{(1)}_{1+\delta}(e^{-T_n}) > \lfloor b(n/k_n)y \rfloor \right] \right).
\]
The distribution of \( \{T_{n+1} - T_n : n \geq 1\} \) in Eq. 4.6 implies
\[
\{T_n - T_i : i = 1, 2, \ldots, n-1\} \overset{d}{=} \frac{1}{2+\delta} \left\{ \sum_{j=i+1}^{n} \frac{E_j}{n-j} : i = 1, 2, \ldots, n-1 \right\}.
\]
Note that \( \{T_n - T_i : i \geq 1\} \) is a sequence of non-decreasing random variables, Eq. 5.9 coincides with the Renyi’s representation for order statistics of iid exponential random variables. Let \( T \) be a unit exponential random variable. Then we have

\[
T_n - T_{U_n} \overset{d}{=} \frac{T}{2 + \delta}.
\]

(5.10)

Also because \( \{B_{1+\delta}(p) : i \geq 1\} \) are iid, we have

\[
\frac{n}{k_n} \mathbb{E}\left( P\left[ 1 + B^{(1)}_{1+\delta} \left( e^{-T/(2+\delta)} \right) > [b(n/k_n)y] \right] \right) = \frac{n}{k_n} \mathbb{P}_{>\left[b(n/k_n)y\right]}(Y_{1+\delta} = k) = p_k.
\]

(5.11)

where the last equality follows by noting that

\[
\mathbb{E}\left( P\left[ 1 + B^{(1)}_{1+\delta} \left( e^{-T/(2+\delta)} \right) = k \right] \right) = \frac{(2 + \delta)\Gamma(k + \delta)\Gamma(3 + 2\delta)}{(2 + \delta)\Gamma(k + 3 + 2\delta)\Gamma(1 + \delta)} = p_k.
\]

Therefore, combining (5.11) with (5.8) gives

\[
II \leq \frac{1}{k_n} \left| \mathbb{P}(BI_1(t) > [b(n/k_n)y]) - \mathbb{P}\left[ 1 + B^{(1)}_{1+\delta} \left( e^{-T_n} \right) > [b(n/k_n)y] \right] \right| \leq \frac{2}{k_n} \to 0, \quad \text{as } n \to \infty.
\]

This completes the proof of Eq. 5.3.

**Step 2.** Using (5.3) and inversion (cf. Resnick 2007, Proposition 3.2), we have for \( y > 0, \)

\[
\frac{D([k_ny])(n)}{b(n/k_n)} \to y^{-\frac{1}{2+\delta}}, \quad \text{in } D(0, \infty), \quad (5.12)
\]

Moreover,

\[
\left( \frac{1}{k_n} \sum_{i=1}^{n} \epsilon_{D_i(n)/b(n/k_n)}, \frac{D([k_ny])(n)}{b(n/k_n)} \right) \Rightarrow (\nu_{2+\delta}, 1) \quad (5.13)
\]

in \( M_+((0, \infty)) \times (0, \infty) \).

**Step 3.** With (5.13), we use a scaling argument to prove (5.2). Define the operator

\[
S : M_+((0, \infty)) \times (0, \infty) \mapsto M_+((0, \infty))
\]

by

\[
S(\nu, c)(A) = \nu(cA).
\]

By the proof in Resnick (2007, Theorem 4.2), the mapping \( S \) is continuous at \((\nu_{2+\delta}, 1)\). Therefore, applying the continuous mapping \( S \) to the joint weak convergence in Eqs. 5.13 gives 5.2. □
5.3 Consistency of the Hill estimator

We are now able to prove the consistency of the Hill estimator applied to \( \{D_i(n) : 1 \leq i \leq n\} \), i.e.

\[
H_{k_n,n} = \frac{1}{k_n} \sum_{i=1}^{k_n} \log \frac{D(i)(n)}{D(k_n+1)(n)}.
\]

Theorem 10 Let \( \{k_n\} \) be an intermediate sequence satisfying (5.1), then

\[
H_{k_n,n} \xrightarrow{p} \frac{1}{2 + \delta}.
\]

Proof First observe

\[
H_{k_n,n} = \int_1^\infty \hat{v}_n(y, \infty) \frac{dy}{y}.
\]

Fix \( M > 0 \) large and define a mapping \( f \mapsto \int_1^M f(y) \frac{dy}{y} \) from \( D(0, \infty] \mapsto \mathbb{R}_+ \). This map is a.s. continuous so

\[
\int_1^M \hat{v}_n(y, \infty) \frac{dy}{y} \xrightarrow{p} \int_1^M v_{2+\delta}(y, \infty) \frac{dy}{y},
\]

and it remains to show by the second converging together theorem (Resnick 2007, Theorem 3.5) that

\[
\lim_{M \to \infty} \limsup_{n \to \infty} P \left( \int_M^\infty \hat{v}_n(y, \infty) \frac{dy}{y} > \varepsilon \right) = 0. \tag{5.14}
\]

The probability in Eq. 5.14 is

\[
P \left( \int_M^\infty \hat{v}_n(y, \infty) \frac{dy}{y} > \varepsilon \right)
\leq P \left( \int_M^\infty \hat{v}_n(y, \infty) \frac{dy}{y} > \varepsilon, \left| \frac{D(k_n)(n)}{b(n/k_n)} - 1 \right| < \eta \right)
+ P \left( \int_M^\infty \hat{v}_n(y, \infty) \frac{dy}{y} > \varepsilon, \left| \frac{D(k_n)(n)}{b(n/k_n)} - 1 \right| \geq \eta \right)
\leq P \left( \int_M^\infty \frac{1}{k_n} \sum_{i=1}^{k_n} \varepsilon D_i(n)/b(n/k_n) \left( (1 - \eta)y, \infty \right) \frac{dy}{y} > \varepsilon \right)
+ P \left( \left| \frac{D(k_n)(n)}{b(n/k_n)} - 1 \right| \geq \eta \right) =: A + B.
\]
By Eq. 5.4, $B \to 0$ as $n \to \infty$, and using the Markov inequality, $A$ is bounded by

$$\frac{1}{\varepsilon} \mathbb{E} \left( \int_{M}^{\infty} \frac{1}{k_n} \sum_{i=1}^{n} \varepsilon_{D_i(n)/b(n/k_n)}((1-\eta)y, \infty) \frac{dy}{y} \right)$$

$$= \frac{1}{\varepsilon} \mathbb{E} \left( \int_{M(1-\eta)}^{\infty} \frac{1}{k_n} \sum_{i=1}^{n} \varepsilon_{D_i(n)/b(n/k_n)}(y, \infty) \frac{dy}{y} \right)$$

$$\leq \frac{1}{\varepsilon} \int_{M(1-\eta)}^{\infty} \frac{1}{k_n} \mathbb{E} \left( N_{> [b(n/k_n)y]}(n) \right) \frac{dy}{y}.$$ 

Recall the first step in the proof of Theorem 9. Both $II$ and $III$ converging to $0$ as $n \to \infty$ gives that for $y > 0$,

$$\frac{1}{k_n} \mathbb{E} \left( N_{> [b(n/k_n)y]}(n) \right) \to y^{-(2+\delta)}.$$  \hspace{1cm} (5.15)

Let $U(t) := \mathbb{E}(N_{> [t]}(n))$ and Eq. 5.15 becomes: for $y > 0$,

$$\frac{1}{k_n} U(b(n/k_n)y) \to y^{-(2+\delta)}, \quad \text{as } n \to \infty.$$ 

Since $U(\cdot)$ is a non-increasing function, $U \in RV_{-(2+\delta)}$ by Resnick (2007, Proposition 2.3(ii)). Therefore, Karamata’s theorem gives

$$A \leq \frac{1}{\varepsilon} \int_{M(1-\eta)}^{\infty} \frac{1}{k_n} \mathbb{E} \left( N_{> [b(n/k_n)y]}(n) \right) \frac{dy}{y} \sim C(\delta, \eta)M^{-(2+\delta)},$$

with some positive constant $C(\delta, \eta) > 0$. Also, $M^{-2-\delta} \to 0$ as $M \to \infty$, and Eq. 5.14 follows. \hfill \square

### 6 Extension to a non-connected preferential attachment model

In this section, we extend the use of the embedding technique to a variant of the model in Section 2.1. In the sequel, we refer to the model introduced in Section 2.1 as Model A and the variant studied in this section is called Model B. For $i \in [n]$, $D_i^B(n)$ is the degree of node $i$ in $G^B(n)$. Given graph $G^B(n)$, the graph $G^B(n+1)$ is obtained by either:

- Adding a new node $n + 1$ and a new edge connecting to an existing node $i \in [n]$ with probability

$$\frac{D_i^B(n) + \delta}{\sum_{i=1}^{n} (D_i^B(n) + \delta) + 1 + \delta} = \frac{D_i^B(n) + \delta}{(2+\delta)n + 1 + \delta};$$

or

- Adding a new node $n + 1$ with a self loop with probability

$$\frac{1 + \delta}{\sum_{i=1}^{n} (D_i^B(n) + \delta) + 1 + \delta} = \frac{1 + \delta}{(2+\delta)n + 1 + \delta}.$$
This is the model studied in van der Hofstad (2017), Chapter 8. Most proofs are omitted in this section (except for Theorem 13), since they are similar to those for Model A.

### 6.1 Convergence results for Model B

The B.I. process framework can still be used for Model B. We keep the independent sequence of \{BI_i(t) : t ≥ 0\}_{i \geq 1} initialized as in Eq. 4.1, as well as the definition of \{\tau_k^{(i)} : k ≥ 1\} for \(i ≥ 1\).

Set \(T_0^B = T_1^B = 0\) and start two B.I. processes \(BI_1(\cdot)\) and \(BI_2(\cdot)\) at \(T_1^B\). At time \(T_n^B\) with \(n ≥ 1\), there exist \(n + 1\) B.I. processes. We define \(T_{n+1}^B\) as the first time after \(T_n^B\) that one of the processes \(\{BI_i(t - T_{i-1}^B) : t ≥ T_{i-1}^B\}_{1 ≤ i ≤ n+1}\) jumps, i.e.

\[
T_{n+1}^B := \min\{T_{i-1}^B + \tau_k^{(i)} : k ≥ 1, T_{i-1}^B + \tau_k^{(i)} > T_n^B, 1 ≤ i ≤ n+1\}, \quad (6.1)
\]

and start a new, independent B.I. process \(\{BI_{n+2}(t - T_{n+1}^B)\}_{t ≥ T_{n+1}^B}\) at \(T_{n+1}^B\).

Under similar arguments as in Theorem 3, we have the following embedding results.

**Corollary 11** Let \(D_B(n) := (D_1^B(n), \ldots, D_n^B(n))\) be the degree sequence in Model B and define

\[
\tilde{D}_B(n) := (BI_1(T_n^B), BI_2(T_n^B), \ldots, BI_{n-1}(T_n^B - T_{n-2}^B), BI_n(T_n^B - T_{n-1}^B)).
\]

Then \(D_B(n)\) and \(\tilde{D}_B(n)\) have the same distribution in \(\mathbb{R}^n\).

The following corollary summarizes the convergence results in Model B, which is a slight variant of Proposition 5 and Theorem 6.

**Corollary 12** (i) For a sequence of iid standard exponential random variables \(\{B_i : i ≥ 1\}\), the branching times \(\{T_i^B : i ≥ 1\}\) satisfies

\[
\{T_i^B : i ≥ 2\} \overset{d}{=} \left\{ \sum_{j=2}^i \frac{B_j}{(2 + \delta)j - 1} : i ≥ 2 \right\},
\]

and \(\frac{n}{e^{(2+\delta)T_n^B}} \xrightarrow{a.s.} W_B\), with \(W_B \sim \text{Gamma}\left(\frac{3+2\delta}{1+\delta}, 1\right)\).

(ii) In \(M_p(0, \infty]\), we have for \(\delta ≥ 0\),

\[
\sum_{i=1}^n \epsilon D_i^B(n)^{1/(2+\delta)} \Rightarrow \sum_{i=1}^\infty \sum_{\sigma_j} e^{-T_{i-1}^B/W_B^{1/(2+\delta)}}.
\]
6.2 Consistency of Hill estimator in Model B

Similar to Model A, in order to examine the consistency of the Hill estimator in Model B, we first analyze the convergence of the tail empirical measure. Let

\[ D^B_{(1)}(n) \geq D^B_{(2)}(n) \geq \cdots \geq D^B_{(n)}(n) \]

be the order statistics for \( D^B(n) \).

**Theorem 13** Suppose that \( \{k_n\} \) is some intermediate sequence satisfying (5.1), then

\[
\frac{1}{k_n} \sum_{i=1}^{n} \epsilon_{D^B(n)/D^B_{(k_n)}(n)}(i) \Rightarrow \nu_{2+\delta},
\]

in \( M_+((0, \infty]) \), where \( \nu_{2+\delta}(x, \infty] = x^{-(2+\delta)}, x > 0 \).

**Proof** The proof consists of three steps, similar to Theorem 6. In particular, we only need to check the first step and once that has been established, the rest follows exactly as in the proof of Theorem 6. Hence, we show (5.4) holds also for Model B.

Recall the three parts in Eq. 5.5. Using van der Hofstad (2017, Proposition 8.4) and Stirling’s formula, we know that \( I \overset{P}{\to} 0 \) and \( III \overset{P}{\to} 0 \), respectively, for \( \{k_n\} \) satisfying (5.1). So we are left with showing

\[
\frac{1}{k_n} \left| \mathbb{E}(N_{> [b(n/k_n)y]}(n)) - np_{> [b(n/k_n)y]} \right| \to 0, \quad n \to \infty.
\]

Note that in Model B, we have

\[
\{T_i^B : i \geq 2\} \overset{d}{=} \left\{ \sum_{j=2}^{i} \frac{B_j}{(2+\delta)j-1} : i \geq 2 \right\},
\]

where \( \{B_i : i \geq 1\} \) is a sequence of iid standard exponential random variables. This makes the order statistic argument used to prove (5.10) not applicable to Model B. However, we instead use the results in Ross (2013, Theorem 1.1) to conclude that for some constant \( C_\delta > 0 \),

\[
\frac{n}{k_n} \left| \mathbb{E}\left( \frac{1}{n} N_{> [b(n/k_n)y]}(n) \right) - p_{> [b(n/k_n)y]} \right| \leq \frac{n}{k_n} \sup_{x \geq 0} \left| \mathbb{E}\left( \frac{1}{n} N_{> x}(n) \right) - p_x \right| = \frac{n}{k_n} \sup_{x \geq 0} \left| \mathbb{E}\left( P(D^B_{Un}(n) > x) \right) - p_x \right|
\]

\[
\leq \frac{n}{k_n} C_\delta \log \frac{n}{k_n} = C_\delta \frac{\log n}{k_n},
\]

and \( \log n/k_n \to 0 \) as \( n \to \infty \), for \( \{k_n\} \) satisfying (5.1). This completes the proof of the theorem.

\[\square\]
Now define the Hill estimator for Model B as
\[ H_{kn,n}^B = \frac{1}{kn} \sum_{i=1}^{kn} \log \frac{D_{(i)}(n)}{D_{(kn+1)}(n)}. \]

The consistency of \( H_{kn,n}^B \) is a direct result of Theorem 13, after applying the proof machinery for Theorem 6. We omit the proof and only give the statement.

**Corollary 14** Let \( \{k_n\} \) be an intermediate sequence satisfying (5.1), then
\[ H_{kn,n}^B \overset{p}{\to} \frac{1}{2+\delta} \quad \text{as } n \to \infty. \]

### 7 Simulation studies

In this section, we use simulations to analyze some unresolved issues with regard to the asymptotic distribution of the Hill estimator in the preferential attachment model. Further research is needed on these issues. Here we only consider Model A since simulation results suggest that the numerical differences between the two models are minor.

Threshold selection or choosing \( k_n \) is an important problem when computing the Hill estimator. We adopt the threshold selection method proposed in Clauset et al. (2009), which is widely used in computer science and network analyses; see the KONECT website (Kunegis 2013). This method is encoded in the \texttt{plfit} script, which can be found at [http://tuvalu.santafe.edu/~aaronc/powerlaws/plfit.r](http://tuvalu.santafe.edu/~aaronc/powerlaws/plfit.r). Here is a summary of this method that we refer to as the “minimum distance method”. Given a sample of \( n \) iid observations, \( Z_1, \ldots, Z_n \) from a power law distribution with tail index \( \hat{\alpha} \), the minimum distance method suggests using the thresholded data consisting of the \( k \) upper-order statistics, \( Z(1) \geq \ldots \geq Z(k) \), for estimating \( \alpha \). The tail index is estimated by
\[ H_{k,n}^{-1} = \hat{\alpha}(k) := \left( \frac{1}{k} \sum_{i=1}^{k} \log \frac{Z(i)}{Z(k+1)} \right)^{-1}, \quad k \geq 1. \]

To select \( k \), we first compute the Kolmogorov-Smirnov (KS) distance between the empirical tail distribution of the upper \( k \) observations and the power-law tail with index \( \hat{\alpha}(k) \):
\[ d_k := \sup_{y \geq 1} \left| \frac{1}{k} \sum_{i=1}^{n} \epsilon Z_i / Z(k+1)(y, \infty] - y^{-\hat{\alpha}(k)} \right|, \quad 1 \leq k \leq n. \]

Then the chosen \( k^* \) is the one that minimizes the KS distance, i.e.
\[ k^* := \arg\min_{1 \leq k \leq n} d_k, \]

© Springer
and we estimate the tail index and threshold by \( \hat{\alpha}(k^*) \) and \( Z_{(k^*+1)} \) respectively. This estimator performs reasonably well if the thresholded portion comes from a Pareto tail and also seems effective for social network data (Drees et al.).

In the iid case, under some second order condition, we know that for some intermediate sequence \( \{k_n\} \),

\[
\sqrt{k_n} (H_{k_n,n} - \alpha^{-1}) \Rightarrow N(0, \alpha^{-2}), \quad \text{as } n \to \infty,
\]

de Haan and Ferreira (2006) and Resnick (2007, Chapter 9.1). However, it remains unclear whether this is also true in the preferential attachment setup.

We start with the limit distribution of the minimum distance estimator \( \hat{\alpha}(k^*) \). In other words, we analyze the distribution of

\[
\sqrt{k^*} (\hat{\alpha}(k^*) - \alpha),
\]

and examine whether it is close to some normal distribution, provided that we have a large preferential attachment model. To do this, we chose \( \alpha = 1.5, 2, 2.5, 3, 3.5 \) and \( n = 10^5 \), generated 500 replications of Model A for each value of \( \alpha \) and computed \( \sqrt{k^*} (\hat{\alpha}(k^*) - \alpha) \) for each replication. QQ plots corresponding to different values of \( \alpha \) are given in Fig. 3. We see that for small values of \( \alpha \), the distribution of \( \sqrt{k^*} (\hat{\alpha}(k^*) - \alpha) \) is close to normal, but it becomes more right-skewed as tails become lighter. However, at this point, it is not clear whether this non-normality is due to the non-normality of the Hill estimator or the minimum distance method.

To investigate this further, we chose

\[
k_n = (n \log n)^{1/2}, n^{0.55}, n^{0.58}, n^{0.6}, n^{0.65}, n^{0.68}, n^{0.7}.
\]

![QQ plots of \( \sqrt{k^*} (\hat{\alpha}(k^*) - \alpha) \) with \( n = 10^5 \) and \( \alpha = 1.5, 2, 2.5, 3, 3.5, 4 \), based on 500 replications of Model A for each value of \( \alpha \). The red dashed lines are the traditional qq-lines used to check normality of the estimates](image)

**Fig. 3** QQ plots of \( \sqrt{k^*} (\hat{\alpha}(k^*) - \alpha) \) with \( n = 10^5 \) and \( \alpha = 1.5, 2, 2.5, 3, 3.5, 4 \), based on 500 replications of Model A for each value of \( \alpha \). The red dashed lines are the traditional qq-lines used to check normality of the estimates
Using the 500 replications obtained under different values of $\alpha$, we computed $\hat{\alpha}(k_n)$ with $k_n$ varied as above. Also, for each value of $\alpha$, we recorded the MSE of $\hat{\alpha}(k_n)$ under different choices of $k_n$, picked $k_n$ giving the smallest MSE and plotted the distribution of the corresponding $\sqrt{k_n}(\hat{\alpha}(k_n) - \alpha)$ in Fig. 4. Different from Fig. 3, most of the QQ plots in Fig. 4 confirm the normality of $\sqrt{k_n}(\hat{\alpha}(k_n) - \alpha)$. This suggests that the non-normality displayed in Fig. 3 is possibly the result of the minimum distance method. Further work on the minimum distance method is ongoing in Drees et al. At the moment, neither the normality of $\sqrt{k_n}(\hat{\alpha}(k_n) - \alpha)$ nor that of $\sqrt{k^*}(\hat{\alpha}(k^*) - \alpha)$ has been properly analyzed and this requires more investigation.

Acknowledgments This work was supported by Army MURI grant W911NF-12-1-0385 to Cornell University. The authors would like to thank the editor and referees for their helpful comments.

References

Athreya, K.B.: Preferential attachment random graphs with general weight function. Internet Math. 4(4), 401–418 (2007). https://doi.org/10.1080/15427951.2007.10129150

Athreya, K.B., Ghosh, A.P., Sethuraman, S.: Growth of preferential attachment random graphs via continuous-time branching processes. Proc. Math. Sci. 118(3), 473–494 (2008)

Bhamidi, S.: Universal techniques to analyze preferential attachment trees: Global and local analysis. Available: http://www.unc.edu/bhamidi/preferent.pdf (2007)

Bollobás, B., Borgs, C., Chayes, J., Riordan, O.: Directed scale-free graphs. In: Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, 2003), pp. 132–139. ACM, New York (2003)

Clauset, A., Shalizi, C.R., Newman, M.E.J.: Power-law distributions in empirical data. SIAM Rev. 51(4), 661–703 (2009). https://doi.org/10.1137/070710111. ISSN 0036-1445
Samorodnitsky, G., Resnick, S., Towsley, D., Davis, R., Willis, A., Wan, P.: Nonstandard regular variation of in-degree and out-degree in the preferential attachment model. J. Appl. Probab. 53(1), 146–161 (2016)

Tavaré, S.: The birth process with immigration, and the genealogical structure of large populations. J. Math. Biol. 25(2), 161–168 (1987)

van der Hofstad, R.: Random Graphs and Complex Networks, vol. 1. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge (2017). ISBN 978-1-107-17287-6

Wan, P., Wang, T., Davis, R.A., Resnick, S.I.: Fitting the linear preferential attachment model. Electron. J. Statist. 11(2), 3738–3780 (2017). https://doi.org/10.1214/17-EJS1327. ISSN 1935-7524

Wan, P., Wang, T., Davis, R.A., Resnick, S.I.: Are extreme value estimation methods useful for network data? Submitted (2017)

Wang, T., Resnick, S.I.: Multivariate regular variation of discrete mass functions with applications to preferential attachment networks. Methodol. Comput. Appl. Probab., 1–14. ISSN 1573-7713. https://doi.org/10.1007/s11009-016-9503-x (2016)

Wang, T., Resnick, S.I.: Asymptotic normality of in- and out-degree counts in a preferential attachment model. Stoch. Model. 33(2), 229–255 (2017). https://doi.org/10.1080/15326349.2016.1256219

Waugh, W.A.O’N.: Transformation of a birth process into a Poisson process. J. Roy. Statist. Soc. Ser. B 32, 418–431 (1970). ISSN 0035-9246. http://links.jstor.org/sici?sici=0035-9246(1970)32:3(418:TOABPI)2.0.CO;2-A&origin=MSN