# Consistency of Hill Estimators in a Linear Preferential Attachment Model 

Tiandong Wang • Sidney I. Resnick

Received: November 15, 2017
Revised: June 28, 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 estimators 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

Mathematics Subject Classification (2010) 60G70 • 60B10 • 60G55 • 60G57 • 05C80 • 62E20

This work was supported by Army MURI grant W911NF-12-1-0385 to Cornell University.
Tiandong Wang
School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853
E-mail: tw398@cornell.edu
Sidney I. Resnick
School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853
E-mail: sir1@cornell.edu

## 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 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

| Square count | $428,143,604,855$ |
| :--- | :--- |
| 4-tour count | $3,483,825,597,342$ |
| Power law exponent (estimated) with $\mathrm{d}_{\text {min }}$ | $1.7310\left(\mathrm{~d}_{\text {min }}=8\right)$ |
| Gini coefficient | $88.2 \%$ |
| Relative edge distribution entropy | $82.2 \%$ |
| Assortativity | -0.015282 |

Fig. 1.1 A snapshot of summary statistics for the Flickr friendship data. The full list of key statistics is available at http://konect.uni-koblenz.de/networks/flickr-growth. the key summary statistics for almost all listed networks. Figure 1.1 displays part of the statistical summaries for the Flickr friendship data given on KONECT (http://konect.uni-koblenz.de/networks/flickr-growth), 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. (Hsing, 1991; Resnick and Stărică, 1995, 1998; Rootzén et al., 1990). 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

$$
\epsilon_{x}(A)=\left\{\begin{array}{ll}
\nsupseteq, & \text { if } x \in A \\
\emptyset, & \text { if } x \notin A
\end{array} .\right.
$$

For positive iid random variables $\left\{X_{i}: i \geq 1\right\}$ 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

$$
\begin{equation*}
\sum_{i=1}^{n} \not X_{i} / b(n)(\cdot) \Rightarrow \operatorname{PRM}\left(\nu_{\alpha}(\cdot)\right), \quad \text { with } \quad \nu_{\alpha}(y, \infty]=y^{-\alpha}, y>0 \tag{1.1}
\end{equation*}
$$

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

$$
\begin{equation*}
\hat{\nu}_{n}(\cdot):=\frac{1}{k_{n}} \sum_{i=1}^{n} \epsilon_{X_{i} / b\left(n / k_{n}\right)}(\cdot) \Rightarrow \nu_{\alpha}(\cdot), \tag{1.2}
\end{equation*}
$$

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 (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 (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 (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 attach 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 (2018)) does not rule out non-normality. This nonnormality 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 (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 $m=\sum_{i}\left(\epsilon_{x_{i}}\right.$. 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 $\left\{X_{n}, n \geq 1\right\}$ iid and non-negative with common regularly varying distribution tail $\bar{F} \in R V_{-\alpha}, \alpha>0$, 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 $\operatorname{PRM}\left(\nu_{\alpha}\right)$, we have (1.1) in $M_{p}(0, \infty]$, and for some
$k_{n} \rightarrow \infty, k_{n} / n \rightarrow 0$, we have (1.2) in $M_{+}(0, \infty]$. Note the limit in (1.1) is random while that in (1.2) is deterministic. Define the Hill estimator $H_{k, n}$ based on $k$ upper order statistics of $\left\{X_{1}, \ldots, X_{n}\right\}$ 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 $\left\{X_{i}: 1 \leq i \leq n\right\}$. In the iid case there are many proofs of consistency (cf. Csörgö et al. (1991a); de Haan and Resnick (1998); Hall (1982); Mason (1982); Mason and Turova (1994)): For $k=k_{n} \rightarrow \infty, k_{n} / n \rightarrow 0$, we have

$$
\begin{equation*}
H_{k_{n}, n} \xrightarrow{P} 1 / \alpha \quad \text { as } n \rightarrow \infty \tag{1.3}
\end{equation*}
$$

We consider $H_{k_{n}, n}$ as a functional of $\hat{\nu}_{n}$ as in Resnick (2007, Theorem 4.2) who shows (1.3) follows from (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

$$
\begin{equation*}
\sum_{i=1}^{n} \oint_{D_{i}(n) / n^{1 /(2+\delta)}} \tag{1.4}
\end{equation*}
$$

that converges weakly to some random limit point measure in $M_{p}(0, \infty]$. We prove the analogy to (1.2) in the network case,

$$
\begin{equation*}
\frac{1}{k_{n}} \sum_{i=1}^{n} \epsilon_{D_{i}(n) / b\left(n / k_{n}\right)} \Rightarrow \nu_{2+\delta}, \quad \text { in } M_{+}(0, \infty] \tag{1.5}
\end{equation*}
$$

with the function $b(n)=c n^{1 /(2+\delta)}$ 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

$$
\begin{equation*}
\frac{f\left(D_{i}(n)\right)+\delta}{\sum_{i} \in[n]}\left(f\left(D_{i}(n)\right)+\delta\right), \tag{2.1}
\end{equation*}
$$

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 $\sum_{i=1}^{n} D_{i}(n)=2 n, n \geq 1$. Therefore, the attachment probability in (2.1) become

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

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); Gao et al. (2017); Rudas et al. (2007) 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.

$$
\begin{equation*}
N_{k}(n):=\sum_{i=1}^{n}\left\{_{\left\{D_{i}(n)=k\right\}},\right. \tag{2.2}
\end{equation*}
$$

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

It is shown in vah der Hofstad (2017, Theorem 8.3) using concentration inequalities and martingale methods that for fixed $k \geq 1$, as $n \rightarrow \infty$,

$$
\begin{equation*}
\frac{N_{k}(n)}{n} \xrightarrow{P} p_{k}=(2+\delta) \frac{\Gamma(k+\delta) \Gamma(3+2 \delta)}{\Gamma(k+3+2 \delta) \Gamma(1+\delta)} \sim(2+\delta) \frac{\Gamma(3+2 \delta)}{\Gamma(1+\delta)} k^{-(3+\delta)} \tag{2.3}
\end{equation*}
$$

$\left(p_{k}\right)_{k \geq 0}$ is a pmf and the asymptotic form, as $k \rightarrow \infty$, follows from Stirling's formula. Let $p_{>k}=\sum_{j}\left(>k p_{j}\right.$ be the complementary cdf and by Scheffé's lemma as well as van der Hofftad (2017, Equation (8.4.6)), we have

$$
\begin{equation*}
\frac{N_{>k}(n)}{n} \xrightarrow{P} p_{>k}:=\frac{\Gamma(k+1+\delta) \Gamma(3+2 \delta)}{\Gamma(k+3+2 \delta) \Gamma(1+\delta)}, \tag{2.4}
\end{equation*}
$$

and again by Stirling's formula we get from (2.4) as $k \rightarrow \infty$,

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

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.

### 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

$$
\begin{equation*}
\zeta(t)=1+N_{0}\left(W\left(e^{\lambda t}-1\right), t \geq 0\right. \tag{3.1}
\end{equation*}
$$

where $\left\{N_{0}(t): t \geq 0\right\}$ 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 $N_{0}(t) / t \rightarrow 1$ almost surely as $t \rightarrow \infty$, it follows immediately that

$$
\begin{equation*}
\frac{\zeta(t)}{e^{\lambda t}} \xrightarrow{\text { a.s. }} W, \quad \text { as } t \rightarrow \infty . \tag{3.2}
\end{equation*}
$$

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), $\{B I(t): t \geq 0\}$, 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 $\left\{\zeta_{i}(t): t \geq 0\right\}_{i \geq 1}$ with parameter $\lambda>0$ and $\zeta_{i}(0)=1$ for $i \geq 1$. Let $B I(0)=0$, then the B.I. process is a shot noise process with form

$$
\begin{equation*}
B I(t):=\sum_{i=1}^{\infty} \oint_{i}\left(t-\tau_{i}\right) \mathbf{1}_{\left\{t \geq \tau_{i}\right\}}=\sum_{i=1}^{N_{\theta}(t)} \zeta_{i}\left(t-\tau_{i}\right) . \tag{3.3}
\end{equation*}
$$

Theorem 1 modifies slightly the statement of Tavaré (1987, Theorem 5) summarizing the asymptotic behavior of the B.I. process.

Theorem 1 For $\{B I(t): t \geq 0\}$ as in (3.3), we have as $t \rightarrow \infty$,

$$
\begin{equation*}
e^{-\lambda t} B I(t) \xrightarrow{\text { a.s. }} \sum_{i=1}^{\infty}\left(W_{i} e^{-\lambda \tau_{i}}=: \sigma\right. \tag{3.4}
\end{equation*}
$$

where $\left\{W_{i}: i \geq 1\right\}$ are independent unit exponential random variables which for each $i \geq 1$ are almost sure limits,

$$
W_{i}=\lim _{t \rightarrow \infty} e^{-t} \zeta_{i}(t)
$$

The random variable $\sigma$ in (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 (3.4) and its Gamma density is justified in Tavaré (1987). It can be guessed from (3.3) and some cavalier interchange of limits and infinite sums. Transforming Poisson points $\left\{\left(W_{i}, \tau_{i}\right), i \geq 1\right\}$, summing and recognizing a Gamma Lévy process at $t=1$ gives the density of $\sigma$.

Remark 2 For a B.I. process $\{B I(t)\}_{t \geq 0}$ with $B I(0)=j \geq 1$, modifying the representation in (3.3) gives

$$
B I(t)=\sum_{i=1}^{j} \zeta_{i}(t)+\sum_{i=j+1}^{\infty}\left(\zeta_{i}\left(t-\tau_{i}\right) \mathbf{1}_{\left\{t \geq \tau_{i}\right\}}\right.
$$

Therefore, $e^{-\lambda t} B I(t) \xrightarrow{\text { a.s. }} \sigma^{\prime}$ where $\sigma^{\prime}$ 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 (1.4) embeds the degree sequences $\left\{D_{i}(n), 1 \leq i \leq n, n \geq 1\right\}$ 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 $\left\{B I_{i}(t): t \geq 0\right\}_{i \geq 1}$ be independent B.I. processes such that

$$
\begin{equation*}
B I_{1}(0)=2, \quad B I_{i}(0)=1, \quad \forall i \geq 2 . \tag{4.1}
\end{equation*}
$$

Each has transition rate is $q_{j, j+1}=j+\delta, \delta>-1$. For $i \geq 1$, let $\left\{\tau_{k}^{(i)}: k \geq 1\right\}$ be the jump times of $\left\{B I_{i}(t): t \geq 0\right\}$ and set $\tau_{0}^{(i)}:=0$ for all $i \geq 1$. For $k \geq 1$,

$$
B I_{1}\left(\tau_{k}^{(1)}\right)=k+2, \quad B I_{i}\left(\tau_{k}^{(i)}\right)=k+1, i \geq 2
$$

Therefore,

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

and $\left\{\tau_{k}^{(i)}-\tau_{k-1}^{(i)}: i \geq 1, k \geq 1\right\}$ are independent.
Set $T_{1}=0$ and relative to $B I_{1}(\cdot)$ define

$$
\begin{equation*}
T_{2}:=\tau_{1}^{(1)}, \tag{4.2}
\end{equation*}
$$

i.e. the first time that $B I_{1}(\cdot)$ jumps. Start the new B.I. process $\left\{B I_{2}\left(t-T_{2}\right)\right.$ : $\left.t \geq T_{2}\right\}$ at $T_{2}$ and let $T_{3}$ be the first time after $T_{2}$ that either $B I_{1}(\cdot)$ or $B I_{2},(\cdot)$ jumps so that,

$$
T_{3}=\min \left\{T_{i}+\tau_{k}^{(i)}: k \geq 1, T_{i}+\tau_{k}^{(i)}>T_{2}, i=1,2\right\}
$$

$$
\begin{aligned}
& B I_{1}(0)=2 \quad B I_{1}\left(T_{2}^{A}\right)=3
\end{aligned}
$$

$$
\begin{aligned}
& B I_{2}\left(T_{3}^{A}-T_{2}^{A}\right)=2 \\
& B I_{3}(0)=1
\end{aligned}
$$

Fig. 4.1 Embedding procedure for Model A assuming $\tau_{1}^{(2)}+T_{2}^{A}<\tau_{2}^{(1)}$.

Start a new, independent B.I. process $\left\{B I_{3}\left(t-T_{3}\right)\right\}_{t \geq T_{3}}$ at $T_{3}$. See Figure 4.1, 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 $\left\{B I_{i}\left(t-T_{i}\right): t \geq T_{i}\right\}_{1 \leq i \leq n}$ jumps, i.e.

$$
\begin{equation*}
T_{n+1}:=\min \left\{T_{i}+\tau_{k}^{(i)}: k \geq 1, T_{i}+\tau_{k}^{(i)}>T_{n}, 1 \leq i \leq n\right\} . \tag{4.3}
\end{equation*}
$$

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

### 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

$$
\mathcal{D}(n):=\left(D_{1}(n), \ldots, D_{n}(n)\right)
$$

is the degree sequence of nodes in the graph $G(n)$ and $\left\{T_{n}\right\}_{n \geq 1}$ is defined as in (4.3). For each fixed $n$, define

$$
\widetilde{\mathcal{D}}(n):=\left(B I_{1}\left(T_{n}\right), B I_{2}\left(T_{n}-T_{2}\right), \ldots, B I_{n-1}\left(T_{n}-T_{n-1}\right), B I_{n}(0)\right),
$$

and then $\mathcal{D}(n)$ and $\widetilde{\mathcal{D}}(n)$ 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 $B I_{n}(\cdot)$ with initial value equal to 1 and one of $B I_{i}, 1 \leq i \leq n-1$ also increases by 1 . This makes the sum of the values of $B I_{i}, 1 \leq i \leq n$, increase by 2 so that

$$
\sum_{i=1}^{n}\left(B I_{i}\left(T_{n}-T_{i}\right)+\delta\right)=(2+\delta) n
$$

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

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

$$
\begin{aligned}
& \mathcal{D}(n+1)=(\not \mathscr{D}(n), 1)+\left(\oint_{\oint_{n+1}}^{(n)}, 0\right), \\
& \widetilde{\mathcal{D}}(n+1)=\left(\not \mathscr{D}^{(n)}(n), 1\right)+\left(\oint_{L_{n+1}}, 0\right),
\end{aligned}
$$

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

$$
P\left[J_{n+1}=j \mid \mathcal{D}(n)\right]=\frac{D_{j}(n)+\delta}{(2+\delta) n}, \quad 1 \leq j \leq n
$$

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

When $n=1$,

$$
\widetilde{\mathcal{D}}(1)=B I_{1}(0)=2=D_{1}(1)=\mathcal{D}(1)
$$

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

According to the preferential attachment setup, we have

$$
\begin{align*}
\mathbf{P}(\mathcal{D}(n+1) & \left.=\left(d_{1}, d_{2}, \ldots, d_{i}+1, d_{i+1}, \ldots, d_{n}, 1\right) \mathcal{D}(n)=\left(d_{1}, d_{2}, \ldots, d_{n}\right)\right) \\
& =\frac{d_{i}+\delta}{(2+\delta) n}, \quad 1 \leq i \leq n \tag{4.4}
\end{align*}
$$

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

$$
\left(B I_{i}\left(T_{n}-T_{i}\right)+\delta\right)^{-1}, \quad 1 \leq i \leq n
$$

which gives for any $1 \leq i \leq n$,

$$
\begin{aligned}
& \mathbf{P}\left(L_{n+1}=i \widetilde{\mathcal{D}}(n)=\left(d_{1}, d_{2}, \ldots, d_{n}\right)\right)( \\
& =\mathbf{P}\left(\widetilde{\mathcal{D}}(n+1)=\left(d_{1}, d_{2}, \ldots, d_{i}+1, d_{i+1}, \ldots, d_{n}, 1\right) \widetilde{\mathcal{D}}(n)=\left(d_{1}, \ldots, d_{n}\right)\right)( \\
& =\mathbf{P}\left(E_{n}^{(i)}<\bigwedge_{j=1, j \neq i}^{n} E_{n}^{(j)} \widetilde{\mathcal{D}}(n)=\left(d_{1}, d_{2}, \ldots, d_{n}\right)\right)( \\
& =\frac{B I_{i}\left(T_{n}-T_{i}\right)+\delta}{\sum_{i=1}^{n}\left(B I_{i}\left(T_{n}-T_{i}\right)+\delta\right)}=\frac{d_{i}+\delta}{(2+\delta) n} .
\end{aligned}
$$

This agqees with the transition probability in (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 (1.1) in the iid case.

### 4.2.1 Branching times.

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

Proposition 5 For $\left\{T_{n}\right\}_{n \geq 1}$ defined in (4.3), we have

$$
\begin{equation*}
\frac{n}{e^{(2+\delta) T_{n}}} \xrightarrow{\text { a.s. }} W, \quad W \sim \operatorname{Exp}(1) . \tag{4.5}
\end{equation*}
$$

Proof Define a counting processes

$$
N(t):=\frac{1}{2} \sum_{i=1}^{\infty} B I_{i}\left(t-T_{i}\right) \mathbf{1}_{\left\{t \geq T_{i}\right\}}
$$

Then we have

$$
\begin{aligned}
& N(t) \mathbf{1}_{\left\{t \in\left[T_{n}, T_{n+1}\right)\right.}=n . \\
& \text { re the junp times of the ce }
\end{aligned}
$$

In other words, $\left\{T_{n}\right\}_{n \geq 1}$ are the junp times of the counting process $N(\cdot)$, with the following structure

$$
\begin{equation*}
\left\{T_{n+1}-T_{n}: n \geq 1\right\} \stackrel{d}{=}\left\{\frac{E_{i}}{(2+\delta) i}, i \geq 1\right\}( \tag{4.6}
\end{equation*}
$$

where $\left\{E_{i}: i \geq 1\right\}$ are iid unit exponential random variables.
From (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 (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 (1.1) for the iid case.
Theorem 6 Suppose that
(1) $\left\{T_{i}: i \geq 1\right\}$, are distributed as in (4.6).
(2) $W$ is the limit random variable as given in (4.5).
(3) $\left\{\sigma_{i}\right\}_{i \geq 1}$ is a sequence of independent Gamma random variables specified in (4.12) and (4.10) below.

Then in $M_{p}((0, \infty])$, we have for $\delta \geq 0$,

$$
\begin{equation*}
\sum_{i=1}^{n} \epsilon_{D_{i}(n) / n^{1 /(2+\delta)}}(\cdot) \Rightarrow \sum_{i=1}^{\infty} \oint_{\sigma_{i} e^{-T_{i}} / W^{1 /(2+\delta)}}(\cdot) \tag{4.7}
\end{equation*}
$$

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

$$
\begin{equation*}
\left(\frac{D_{(1)}(n)}{n^{1 /(2+\delta)}}, \ldots \frac{D_{(k)}(n)}{n^{1 /(2+\delta)}}\right) \Rightarrow W^{-1 /(2+\delta)}\left(\left(\sigma \cdot e^{-T .}\right)_{(1)}, \ldots,\left(\sigma \cdot e^{-T .}\right)_{(k)}\right) \tag{4.8}
\end{equation*}
$$

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 (4.8) represents the ordered $k$ largest points from the right side of (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) $\left\{T_{i}: i \geq 1\right\}$ are distributed as in (4.6).
(2) $W$ is the limit random variable as given in (4.5).

Then the degree sequence $\left\{D_{i}(n): 1 \leq i \leq n\right\}$ satisfies:
(i) For each $i \geq 1$,

$$
\begin{equation*}
\frac{D_{i}(n)}{n^{1 /(2+\delta)}} \Rightarrow \frac{\sigma_{i} e^{-T_{i}}}{W^{1 /(2+\delta)}} \tag{4.9}
\end{equation*}
$$

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

$$
\begin{equation*}
\sigma_{1} \sim \operatorname{Gamma}(2+\delta, 1), \quad \text { and } \quad \sigma_{i} \sim \operatorname{Gamma}(1+\delta, 1), \quad i \geq 2 \tag{4.10}
\end{equation*}
$$

Furthermore, for $i \geq 1, \sigma_{i}$ is independent from $e^{-T_{i}}$.
(ii) For $\delta>-1$,

$$
\begin{equation*}
\max _{i \geq 1} \frac{D_{i}(n)}{n^{1 /(2+\delta)}} \Rightarrow W^{-1 /(2+\delta)} \max _{i \geq 1} \sigma_{i} e^{-T_{i}} \tag{4.11}
\end{equation*}
$$

where we set $D_{i}(n):=0$ for all $i \geq n+1$.
Proof (i) Applying the results in Remark 2 gives that as $t \rightarrow \infty$,

$$
\frac{B I_{i}\left(t-T_{i}\right)}{e^{t-T_{i}}} \xrightarrow{\text { a.s. }} \sigma_{i}, \quad i \geq 1,
$$

where $\left\{\sigma_{i}\right\}_{i \geq 1}$ are independent Gamma random variables with

$$
\sigma_{1} \sim \operatorname{Gamma}(2+\delta, 1) \quad \text { and } \quad \sigma_{i} \sim \operatorname{Gamma}(1+\delta, 1), \quad i \geq 2
$$

Thus as $n \rightarrow \infty$,

$$
\begin{equation*}
\frac{B I_{i}\left(T_{n}-T_{i}\right)}{e^{T_{n}-T_{i}}} \xrightarrow{\text { a.s. }} \sigma_{i}, \quad i \geq 1 . \tag{4.12}
\end{equation*}
$$

Combining (4.12) with (4.5), we have for fixed $1 \leq i \leq n$,

$$
\frac{B I_{i}\left(T_{n}-T_{i}\right)}{n^{1 /(2+\delta)}} \xrightarrow{\text { a.s. }} \frac{\sigma_{i} e^{-T_{i}}}{W^{1 /(2+\delta)}} .
$$

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 $i \geq n+1, B I_{i}\left(T_{n}-T_{i}\right)=0$ so from Theorem 3, it suffices to show

$$
\max _{i \geq 1} \frac{B I_{i}\left(T_{n}-T_{i}\right)}{n^{1 /(2+\delta)}} \xrightarrow{\text { a.s. }} \max _{i \geq 1} \frac{\sigma_{i} e^{-T_{i}}}{W^{1 /(2+\delta)}},
$$

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 $y>0$,

$$
\left.\mathbf{P} \quad \sum_{i=1}^{\infty} \oint_{\sigma_{i} e^{-T_{i}} / W^{1 /(2+\delta)}}(\{y\})=0\right)(=1 .
$$

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 $n \rightarrow \infty$,

$$
\begin{equation*}
\left.\left.\mathbf{E} \quad \sum_{i=1}^{n} \epsilon_{D_{i}(n) / n^{1 /(2+\delta)}}(y, \infty]\right) \rightarrow \mathbf{E} \quad \sum_{i=1}^{\infty} \not \sigma_{\sigma_{i} e^{-T_{i} / W^{1 /(2+\delta)}}}(y, \infty]\right)( \tag{4.13}
\end{equation*}
$$

(b) For $y>0$, as $n \rightarrow \infty$,

$$
\begin{align*}
\mathbf{P} & \left.\sum_{i=1}^{n} \not \oint_{D_{i}(n) / n^{1 /(2+\delta)}}(y, \infty]=0\right)( \\
& \left.\longrightarrow \mathbf{P} \quad \sum_{i=1}^{\infty} \oint_{\sigma_{i} e^{-T_{i} / W^{1 /(2+\delta)}}}(y, \infty]=0\right)( \tag{4.14}
\end{align*}
$$

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

$$
\begin{aligned}
\left.\mathbf{E} \sum_{i=1}^{M} \epsilon_{D_{i}(n) / n^{1 /(2+\delta)}}(y, \infty]\right) & = \\
& \sum_{i=1}^{M} \not \mathbf{p}\left(\frac{D_{i}(n)}{n^{1 /(2+\delta)}}>y\right) \\
& \longrightarrow \sum_{i=1}^{M} \not \mathbf{p}\left(\sigma_{i} e^{-T_{i}} W^{-1 /(2+\delta)}>y\right)
\end{aligned}
$$

$$
\left.=\mathbf{E} \quad \sum_{i=1}^{M} \not \oint_{\sigma_{i} e^{-T_{i}} / W^{1 /(2+\delta)}}(y, \infty]\right)(
$$

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

$$
\mathbf{E} \begin{align*}
\left.\sum_{i=M+1}^{n} \epsilon_{D_{i}(n) / n^{1 /(2+\delta)}}(y, \infty]\right) & =\sum_{i=M+}^{n}\left(\mathbf { P } \left(\frac{D_{i}(n)}{\left.n^{1 /(2+\delta)}>y\right)}\right.\right. \\
& \leq y^{-k} \sum_{i=M+}^{n}\left(\mathbf{E}\left[\left(\frac{D_{i}(n)}{n^{1 /(2+\delta)}}\right)^{k}\right]( \right. \tag{4.15}
\end{align*}
$$

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

$$
\mathbf{E}\left[\left(\frac{D_{i}(n)}{n^{1 /(2+\delta)}}\right)^{k}\right] \leq \mathbf{E}\left[\left(\frac{p_{i}(n)+\delta}{n^{1 /(2+\delta)}}\right)^{k}\right] \leq \mathbf{E}\left[\left(\left(\frac{\sigma_{i} e^{-T_{i}}}{V^{1 /(2+\delta)}}\right)^{k}\right](\right.
$$

where the last inequality follows from the result in van der Hofstad (2017, Equation (8.7.26)). From van der Hofstad (2017, Equation (8.7.22)), we have

$$
\mathbf{E}\left[( ( \frac { \sigma _ { i } e ^ { - T _ { i } } } { V ^ { 1 / ( 2 + \delta ) } } ) ^ { k } ] \left(=\frac{\Gamma\left(i-\frac{1}{2+\delta}\right)}{\Gamma\left(i+\frac{k-1}{2+\delta}\right)} \frac{\Gamma(k+1+\delta)}{\Gamma(1+\delta)} \sim C_{k, \delta} i^{-\frac{k}{2+\delta}}\right.\right.
$$

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

$$
\begin{aligned}
\left.\mathbf{E} \sum_{i=M+1}^{n} \epsilon_{D_{i}(n) / n^{1 /(2+\delta)}}(y, \infty]\right) & \left(\leq y^{-k} \sum_{i=M+}^{n}\left(\mathbf{E}\left[\left(\frac{D_{i}(n)}{n^{1 /(2+\delta)}}\right)^{k}\right]( \right.\right. \\
& \leq y^{-k} \sum_{i=M+}^{\infty}\left(\mathbf{E}\left[\left(\frac{\sigma_{i} e^{-T_{i}}}{V^{1 /(2+\delta)}}\right)^{k}\right]( \right. \\
& =y^{-k} \sum_{i=M+}^{\infty}\left(\frac{\Gamma\left(i-\frac{1}{2+\delta}\right)}{\Gamma\left(i+\frac{k-1}{2+\delta}\right)} \frac{\Gamma(k+1+\delta)}{\Gamma(1+\delta)}\right. \\
& \xrightarrow{M \rightarrow \infty} 0,
\end{aligned}
$$

since $k /(2+\delta)>1$. This verifies Condition (a).
To see (4.14), we have

$$
\begin{aligned}
\left\{\sum_{i=1}^{n} \oint_{D_{i}(n) / n^{1 /(2+\delta)}}(y, \infty]=0\right\} & =\left\{\frac{D_{i}(n)}{n^{1 /(2+\delta)}} \leq y, 1 \leq i \leq n\right\} \\
& =\left\{\max _{1 \leq i \leq n} \frac{D_{i}(n)}{n^{1 /(2+\delta)}} \leq y\right\}(
\end{aligned}
$$

Since we set $D_{i}(n)=0$ for all $i \geq n+1$, then

$$
\left\{\max _{\{\leq i \leq n} \frac{D_{i}(n)}{n^{1 /(2+\delta)}} \leq y\right\}=\left\{\max _{i \geq 1} \frac{D_{i}(n)}{n^{1 /(2+\delta)}} \leq y\right\}(
$$

Similarly,

$$
\left\{\sum_{i=1}^{\infty} \oint_{\sigma_{i} e^{-T_{i}} / W^{1 /(2+\delta)}}(y, \infty]=0\right\}=\left\{\max _{i \geq 1} \frac{\sigma_{i} e^{-T_{i}}}{W^{1 /(2+\delta)}} \leq y\right\}
$$

By (4.11), we have for $y>0$,

$$
\mathbf{P}\left(\max _{i \geq 1} \frac{D_{i}(n)}{n^{1 /(2+\delta)}} \leq y\right) \rightarrow \mathbf{P}\left(\max _{\geq 1} \frac{\sigma_{i} e^{-T_{i}}}{W^{1 /(2+\delta)}} \leq y\right)(\text { as } n \rightarrow \infty
$$

which gives (4.14) and completes the proof of (iv).

## 5 Consistency of the Hill Estimator.

We now turn to (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 (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 $E\left(N_{>k}(n) / n\right)$ 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 $\left\{\sigma_{i}: i \geq 1\right\}$ 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 $Y^{\prime} s$ to get

$$
H_{k, n}=\frac{1}{k} \sum_{i=1}^{k} \log \left(\left(\frac{Y_{i}}{k_{k+1}}\right)=\frac{1}{k} \sum_{i=1}^{k}\left(T_{k+1}-T_{i}\right)\right.
$$

Recall from (4.6) that

$$
T_{n+1}-T_{n} \stackrel{d}{=} E_{n} /(n(2+\delta)),
$$

where $E_{n}, n \geq 1$ are iid unit exponential random variables. Then

$$
H_{k, n}=\frac{1}{k} \sum_{i=1}^{k} \sum_{l=i}^{k}\left(T_{l+1}-T_{l}\right)=\frac{1}{k} \sum_{l=1}^{k} f\left(T_{l+1}-T_{l}\right)=\frac{1}{k} \sum_{l=1}^{k} \frac{E_{l}}{2+\delta} \xrightarrow{\text { a.s. }} \frac{1}{2+\delta},
$$

by strong law of large numbers, provided that $k \rightarrow \infty$.
There are clear shortcomings to this approach, the most obvious being that we only dealt with the points at asymptopia rather than $\left\{D_{i}(n), 1 \leq i \leq n\right\}$. Furthermore we simplified the limit points by neglecting the $\sigma_{i}$ 's. We have not found an effective way to analyze order statistics of $\left\{\sigma_{i} e^{-T_{i}} / W^{1 /(2+\delta)}: i \geq 1\right\}$.

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

$$
\left(D_{1}(n), D_{2}(n), \ldots, D_{n}(n)\right)
$$

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

$$
\hat{\nu}_{n}(\cdot):=\frac{1}{k_{n}} \sum_{i=1}^{n} \epsilon_{D_{i}(n) / D_{\left(k_{n}\right)}(n)}(\cdot)
$$

for some intermediate sequence $\left\{k_{n}\right\}$, i.e. $k_{n} \rightarrow \infty$ and $k_{n} / n \rightarrow 0$ as $n \rightarrow \infty$.
Theorem 9 Suppose that $\left\{k_{n}\right\}$ is some intermediate sequence satisfying

$$
\begin{equation*}
\liminf _{n \rightarrow \infty} k_{n} /(n \log n)^{1 / 2}>0 \quad \text { and } \quad k_{n} / n \rightarrow 0 \quad \text { as } \quad n \rightarrow \infty \tag{5.1}
\end{equation*}
$$

then

$$
\begin{equation*}
\hat{\nu}_{n} \Rightarrow \nu_{2+\delta} \tag{5.2}
\end{equation*}
$$

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

$$
b\left(n / k_{n}\right)=\left(\frac{\Gamma(3+2 \delta)}{\Gamma(1+\delta)}\right)^{\frac{1}{2+\delta}}\left(n / k_{n}\right)^{\frac{1}{2+\delta}}
$$

we have in $M_{+}((0, \infty])$,

$$
\begin{equation*}
\frac{1}{k_{n}} \sum_{i=1}^{n} \oint_{D_{i}(n) / b\left(n / k_{n}\right)} \Rightarrow \nu_{2+\delta} \tag{5.3}
\end{equation*}
$$

and it suffices to justify for any $y>0$,

$$
\begin{equation*}
\frac{1}{k_{n}} N_{>\left[b\left(n / k_{n}\right) y\right]}(n)-y^{-(2+\delta)} \quad \xrightarrow{P} 0, \quad n \rightarrow \infty \tag{5.4}
\end{equation*}
$$

The left side of (5.4) is bounded by

$$
\begin{aligned}
& \frac{1}{k_{n}} N_{>\left[b\left(n / k_{n}\right) y\right]}(n)-y^{-(2+\delta)} \\
\leq & \frac{1}{k_{n}} N_{>\left[b\left(n / k_{n}\right) y\right]}(n)-\mathbf{E}\left(N_{>\left[b\left(n / k_{n}\right) y\right]}(n)\right) \\
& +\frac{1}{k_{n}} \mathbf{E}\left(N_{>\left[b\left(n / k_{n}\right) y\right]}(n)\right)-n p_{>\left[b\left(n / k_{n}\right) y\right]}
\end{aligned}
$$

$$
\begin{align*}
& +\frac{n}{k_{n}} p_{>\left[b\left(n / k_{n}\right) y\right]}-y^{-(2+\delta)} \\
& =: I+I I+I I I . \tag{5.5}
\end{align*}
$$

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 (2.4) for fixed $k$, then we have

$$
\begin{align*}
\frac{n}{k_{n}} p_{>\left[b\left(n / k_{n}\right) y\right]} & =\frac{n}{k_{n}} \frac{\Gamma(3+2 \delta)}{\Gamma(1+\delta)} \frac{\Gamma\left(\left[b\left(n / k_{n}\right) y\right]+1+\delta\right)}{\Gamma\left(\left[b\left(n / k_{n}\right) y\right]+3+2 \delta\right)} \\
& =\frac{\Gamma(3+2 \delta)}{\Gamma(1+\delta)} \frac{n}{k_{n}}\left(b\left(n / k_{n}\right) y\right)^{-(2+\delta)}\left(1+O\left(\left(\frac{1}{\phi\left(n / k_{n}\right)}\right)\right)( \right. \\
& =y^{-(2+\delta)}\left(1+O\left(\frac{1}{\phi\left(n / k_{n}\right)}\right)\right)( \tag{5.6}
\end{align*}
$$

Hence, $I$ II $\rightarrow 0$ as $n \rightarrow \infty$
Consider $I$ and we have for $\epsilon>0$,

$$
\begin{aligned}
& \mathbf{P}\left(\frac{1}{k_{n}} N_{>\left[b\left(n / k_{n}\right) y\right]}(n)-\mathbf{E}\left(N_{>\left[b\left(n / k_{n}\right) y\right]}(n)\right)>\epsilon\right) \\
& \quad=\mathbf{P}\left(N_{>\left[b\left(n / k_{n}\right) y\right]}(n)-\mathbf{E}\left(N_{>\left[b\left(n / k_{n}\right) y\right]}(n)\right)>\epsilon k_{n}\right)
\end{aligned}
$$

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

$$
\begin{aligned}
& \mathbf{P}\left(\mid\left(N_{>k}(n)-\mathbf{E}\left(N_{>k}(n)\right) \mid \geq C \sqrt{n \log n}\right)(o(1 / n) .\right. \\
& \imath)=0 \text { a.s. for all } k>n, \text { then }
\end{aligned}
$$

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

$$
\begin{align*}
& \mathbf{P}\left(\max _{k}\left|N_{>k}(n)-\mathbf{E}\left(N_{>k}(n)\right)\right| \geq C \sqrt{n \log n}\right) \\
& =\mathbf{P}\left(\max _{0 \leq k \leq n}\left|N_{>k}(n)-\mathbf{E}\left(N_{>k}(n)\right)\right| \geq C \sqrt{\eta \log n}\right) \\
& \leq \sum_{k=1}^{n} \mathbf{P}\left(| | N_{>k}(n)-\mathbf{E}\left(N_{>k}(n)\right) \mid \geq C \sqrt{n \log n}\right) \mid=o(1) . \tag{5.7}
\end{align*}
$$

Therefore, for $\left\{k_{n}\right\}$ satisfying (5.1), we have

$$
\begin{aligned}
& \mathbf{P}\left(N_{>\left[b\left(n / k_{n}\right) y\right]}(n)-\mathbf{E}\left(N_{>\left[b\left(n / k_{n}\right) y\right]}(n)\right)>\epsilon k_{n}\right) \\
& \quad \leq \mathbf{P}\left(\max _{k}\left|N_{>k}(n)-\mathbf{E}\left(N_{>k}(n)\right)\right| \geq \epsilon k_{n}\right)(=o(\mathbb{1}),
\end{aligned}
$$

which gives $I \xrightarrow{P} 0$.

Then we are left to show $I I \rightarrow 0$ as $n \rightarrow \infty$. Let $U_{n}$ be a uniform random variable on $\{1,2, \ldots, n\}$, then

$$
\begin{aligned}
\frac{1}{k_{n}} \mathbf{E}\left(N_{>\left[b\left(n / k_{n}\right) y\right]}(n)\right) & =\frac{n}{k_{n}} \frac{1}{n} \sum_{i=1}^{n} \not \mathbf{p}\left(D_{i}(n)>\left[b\left(n / k_{n}\right) y\right]\right) \\
& =\frac{n}{k_{n}} \mathbf{E}\left(\mathbf{P}\left(D_{U_{n}}(n)>\left[b\left(n / k_{n}\right) y\right]\right)\right) .
\end{aligned}
$$

Let $B_{a}(p)$ be a negative binomial integer valued random variable with parameters $a>0$ and $p \in(0,1)$ (abbreviated as $N B(a, p)$ ), and the generating function of $B_{a}(p)$ is

$$
\mathbf{E}\left(s^{B_{a}(p)}\right)\left(=(s+(1-s) / p)^{-a}, \quad 0 \leq s \leq 1\right.
$$

Suppose that $\left\{B_{1+\delta}^{(i)}(p): i \geq 1\right\}$ is a sequence of iid $N B(1+\delta, p)$ random variables and $B_{2+\delta}(p)$ is another $N B(2+\delta, p)$ random variable independent from $\left\{B_{1+\delta}^{(i)}(p): i \geq 1\right\}$. Then by the B.I. process construction, we have for $k, t \geq 0$,

$$
\left.\begin{array}{l}
\mathbf{P}\left(B I_{1}(t)>k\right)=\mathbf{P}\left[2\left(+B_{2+\delta}\left(e^{-t}\right)>k\right]\right. \\
\mathbf{P}\left(B I_{i}(t)>k\right)=\mathbf{P}\left[1+B_{1+\delta}^{(i)}\left(e^{-t}\right)>k\right]
\end{array}\right\}^{2} \quad \begin{aligned}
& \\
& i \geq 2
\end{aligned}
$$

Therefore, applying the embedding technique gives

$$
\begin{align*}
\frac{n}{k_{n}} & \mathbf{E}\left(\mathbf{P}\left(D_{U_{n}}(n)>\left[b\left(n / k_{n}\right) y\right]\right)\right) \\
= & \frac{n}{k_{n}} \frac{1}{n} \sum_{i=1}^{n} \mathbf{P}\left(B I_{i}\left(T_{n}-T_{i}\right)>\left[b\left(n / k_{n}\right) y\right]\right) \\
= & \frac{n}{k_{n}} \frac{1}{n} \sum_{i=1}^{n} \mathbf{P}\left[1+B_{1+\delta}^{(i)}\left(e^{-\left(T_{n}-T_{i}\right)}\right)\left(p\left[b\left(n / k_{n}\right) y\right]\right]\right. \\
& +\frac{1}{k_{n}}\left(\not \boldsymbol{P}^{\left.\left(B I_{1}\left(T_{n}\right)>\left[b\left(n / k_{n}\right) y\right]\right)-\mathbf{P}\left[1+B_{1+\delta}^{(1)}\left(e^{-T_{n}}\right)>\left[b\left(n / k_{n}\right) y\right]\right]\right)}\right. \\
= & \frac{n}{k_{n}} \mathbf{E}\left(\mathbf{P}\left[1+B_{1+\delta}^{\left(U_{n}\right)}\left(e^{-\left(T_{n}-T_{U_{n}}\right)}\right)\left(>\left[b\left(n / k_{n}\right) y\right]\right]\right)( \right. \\
& +\frac{1}{k_{n}}\left(\boldsymbol{p}\left(B I_{1}\left(T_{n}\right)>\left[b\left(n / k_{n}\right) y\right]\right)-\mathbf{P}\left[1+B_{1+\delta}^{(1)}\left(e^{-T_{n}}\right) \not p\left[b\left(n / k_{n}\right) y\right]\right]\right) . \tag{5.8}
\end{align*}
$$

The distribution of $\left\{T_{n+1}-T_{n}: n \geq 1\right\}$ in (4.6) implies

$$
\begin{equation*}
\left\{T_{n}-T_{i}: i=1,2, \ldots, n-1\right\} \stackrel{d}{=} \frac{1}{2+\delta}\left\{\sum_{j=1}^{n-i} \frac{E_{j}}{n-j}: i=1,2, \ldots, n-1\right\}( \tag{5.9}
\end{equation*}
$$

Note that $\left\{T_{n}-T_{i}: i \geq 1\right\}$ is a sequence of non-decreasing random variables, (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

$$
\begin{equation*}
T_{n}-T_{U_{n}} \stackrel{d}{=} \frac{T}{2+\delta} . \tag{5.10}
\end{equation*}
$$

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

$$
\begin{align*}
& \frac{n}{k_{n}} \mathbf{E}\left(\mathbf{P}\left[1+B_{1+\delta}^{\left(U_{n}\right)}\left(e^{-\left(T_{n}-T_{U_{n}}\right)}\right) \beta\left[b\left(n / k_{n}\right) y\right]\right]\right)( \\
& =\frac{n}{k_{n}} \mathbf{E}\left(\mathbf { P } [ 1 + B _ { 1 + \delta } ^ { ( 1 ) } ( e ^ { - T / ( 2 + \delta ) } ) ( p [ b ( n / k _ { n } ) y ] ] ) \left(=\frac{n}{k_{n}} p_{>\left[b\left(n / k_{n}\right) y\right]},\right.\right. \tag{5.11}
\end{align*}
$$

where the last equality follows from the results in van der Hofstad (2017, Equation (8.4.10)). Therefore, combining (5.11) with (5.8) gives

$$
\begin{aligned}
I I & \leq \frac{1}{k_{n}} \mathbf{P}\left(B I_{1}(t)>\left[b\left(n / k_{n}\right) y\right]\right)-\mathbf{P}\left[1+B_{1+\delta}^{(1)}\left(\ell^{-T_{n}}\right)>\left[b\left(n / k_{n}\right) y\right]\right]( \\
& \leq \frac{2}{k_{n}} \rightarrow 0, \quad \text { as } n \rightarrow \infty .
\end{aligned}
$$

This completes the proof of (5.3).
Step 2. Using (5.3) and inversion (cf. Resnick (2007, Proposition 3.2)), we have for $y>0$,

$$
\begin{equation*}
\frac{D_{\left(\left[k_{n} y\right]\right)}(n)}{b\left(n / k_{n}\right)} \xrightarrow{P} y^{-\frac{1}{2+\delta}}, \quad \text { in } D(0, \infty] \tag{5.12}
\end{equation*}
$$

Moreover,

$$
\begin{equation*}
\left.\frac{1}{k_{n}} \sum_{i=1}^{n} \epsilon_{D_{i}(n) / b\left(n / k_{n}\right)}, \frac{D_{\left(\left[k_{n}\right]\right)}(n)}{b\left(n / k_{n}\right)}\right)\left(\Rightarrow\left(\nu_{2+\delta}, 1\right)\right. \tag{5.13}
\end{equation*}
$$

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

$$
S(\nu, c)(A)=\nu(c A) .
$$

By the proof in Resnick (2007, Theorem 4.2), the mapping $S$ is continuous at $\left(\nu_{2+\delta}, 1\right)$. Therefore, applying the continuous mapping $S$ to the joint weak convergence in (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 $\left\{D_{i}(n): 1 \leq i \leq n\right\}$, i.e.

$$
H_{k_{n}, n}=\frac{1}{k_{n}} \sum_{i=1}^{k_{n}} \log \frac{D_{(i)}(n)}{D_{\left(k_{n}+1\right)}(n)}
$$

Theorem 10 Let $\left\{k_{n}\right\}$ be an intermediate sequence satisfying (5.1), then

$$
H_{k_{n}, n} \xrightarrow{P} \frac{1}{2+\delta}
$$

Proof First observe

$$
H_{k_{n}, n}=\iint_{\hat{1}}^{\infty} \hat{\nu}_{n}(y, \infty] \frac{\mathrm{d} y}{y} .
$$

Fix $M>0$ large and define a mapping $f \mapsto \int_{1}^{M} f(y) \frac{\mathrm{d} y}{y}$ from $D(0, \infty] \mapsto \mathbb{R}_{+}$.
This map is a.s. continuous so

$$
\iint_{\mathcal{A}}^{M} \hat{\nu}_{n}(y, \infty] \frac{\mathrm{d} y}{y} \xrightarrow{P} \int_{\mathcal{N}}^{M} \int_{2+\delta}(y, \infty] \frac{\mathrm{d} y}{y}
$$

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

$$
\begin{equation*}
\lim _{M \rightarrow \infty} \limsup _{n \rightarrow \infty} \mathbf{P}\left(\iint_{I}^{\infty} \hat{\nu}_{n}(y, \infty] \frac{\mathrm{d} y}{y}>\varepsilon\right)=0 \tag{5.14}
\end{equation*}
$$

The probability in (5.14) is

$$
\begin{aligned}
& \mathbf{P}\left(\int\left(\hat{\nu}_{n}(y, \infty] \frac{\mathrm{d} y}{y}>\varepsilon\right)\left(\int_{N}^{\infty} \hat{\nu}_{n}(y, \infty] \frac{\mathrm{d} y}{y}>\varepsilon, \frac{D_{\left(k_{n}\right)}(n)}{b\left(n / k_{n}\right)}-1<\eta\right)( \right. \\
\leq & \mathbf{P}\left(\iint_{1}^{\infty} \hat{\nu}_{n}(y, \infty] \frac{\mathrm{d} y}{y}>\varepsilon, \frac{D_{\left(k_{n}\right)}(n)}{b\left(n / k_{n}\right)}-1 \geq \eta\right)( \\
& +\mathbf{P}\left(\int_{1}\right) \\
\leq & \mathbf{P} \quad \int\left(\frac{1}{k_{n}} \sum_{i=1}^{n} \epsilon_{D_{i}(n) / b\left(n / k_{n}\right)}((1-\eta) y, \infty] \frac{\mathrm{d} y}{y}>\varepsilon\right) \\
& +\mathbf{P}\left(\left(\frac{D_{\left(k_{n}\right)}(n)}{b\left(n / k_{n}\right)}-1 \geq \eta\right)(=: A+B .\right.
\end{aligned}
$$

By (5.4), $B \rightarrow 0$ as $n \rightarrow \infty$, and using the Markov inequality, $A$ is bounded by

$$
\frac{1}{\varepsilon} \mathbf{E} \quad \int\left(\int_{1}^{\infty} \frac{1}{k_{n}} \sum_{i=1}^{n} \epsilon_{D_{i}(n) / b\left(n / k_{n}\right)}((1-\eta) y, \infty] \frac{\mathrm{d} y}{y}\right)
$$

$$
\begin{aligned}
& \left.=\frac{1}{\varepsilon} \mathbf{E} \iint_{\Lambda(1-\eta)}^{\infty} \frac{1}{k_{n}} \sum_{i=1}^{n} \epsilon_{D_{i}(n) / b\left(n / k_{n}\right)}(y, \infty] \frac{\mathrm{d} y}{y}\right) \\
& \leq \frac{1}{\varepsilon} \iint_{\Lambda(1-\eta)}^{\infty} \frac{1}{k_{n}} \mathbf{E}\left(N_{>\left[b\left(n / k_{n}\right) y\right]}(n)\right) \frac{\mathrm{d} y}{y} .
\end{aligned}
$$

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

$$
\begin{equation*}
\frac{1}{k_{n}} \mathbf{E}\left(N_{>\left[b\left(n / k_{n}\right) y\right]}(n)\right) \nrightarrow y^{-(2+\delta)} . \tag{5.15}
\end{equation*}
$$

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

$$
\frac{1}{k_{n}} U\left(b\left(n / k_{n}\right) y\right) \rightarrow y^{-(2+\delta)}, \quad \text { as } n \rightarrow \infty
$$

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

$$
A \leq \frac{1}{\varepsilon} \iint_{(1-\eta)}^{\infty} \frac{1}{k_{n}} \mathbf{E}\left(N_{>\left[b\left(n / k_{n}\right) y\right]}(n)\right) \frac{\mathrm{d} y}{y} \sim C(\delta, \eta) M^{-(2+\delta)},
$$

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

## 6 Another 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
or

$$
\frac{D_{i}^{B}(n)+\delta}{\sum_{i=1}^{n}\left(\not \prod_{i}^{B}(n)+\delta\right)(+1+\delta}=\frac{D_{i}^{B}(n)+\delta}{(2+\delta) n+1+\delta}
$$

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

$$
\frac{1+\delta}{\sum_{i=1}^{n}\left(\not p_{i}^{B}(n)+\delta\right)}\left(\begin{array}{l}
1+\delta \\
\text { Hofstad (2017). Chapter 8 A A }
\end{array}\right.
$$

This is the model studiedin van der Hofstad (2017), Chapter 8. All proofs are omitted in this section, 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 $\left\{B I_{i}(t): t \geq 0\right\}_{i \geq 1}$ initialized as in (4.1), as well as the definition of $\left\{\tau_{k}^{(i)}: k \geq 1\right\}$ for $i \geq 1$.

Set $T_{0}^{B}=T_{1}^{B}=0$ and start two B.I. processes $B I_{1}(\cdot)$ and $B I_{2}(\cdot)$ at $T_{1}^{B}$. At time $T_{n}^{B}$ with $n \geq 1$, 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 $\left\{B I_{i}\left(t-T_{i-1}^{B}\right): t \geq T_{i-1}^{B}\right\}_{1 \leq i \leq n+1}$ jumps, i.e.

$$
\begin{equation*}
T_{n+1}^{B}:=\min \left\{T_{i-1}^{B}+\tau_{k}^{(i)}: k \geq 1, T_{i-1}^{B}+\tau_{k}^{(i)}>T_{n}^{B}, 1 \leq i \leq n+1\right\} \tag{6.1}
\end{equation*}
$$

and start a new, independent B.I. process $\left\{B I_{n+2}\left(t-T_{n+1}^{B}\right)\right\}_{t \geq T_{n+1}^{B}}$ at $T_{n+1}^{B}$.
Under similar arguments as in Theorem 3, we have the following embedding results.
Corollary 11 Let $\mathcal{D}^{B}(n):=\left(D_{1}^{B}(n), \ldots, D_{n}^{B}(n)\right)$ be the degree sequence in Model $B$ and define

$$
\widetilde{\mathcal{D}}^{B}(n):=\left(B I_{1}\left(T_{n}^{B}\right), B I_{2}\left(T_{n}^{B}\right), \ldots, B I_{n-1}\left(T_{n}^{B}-T_{n-2}^{B}\right), B I_{n}\left(T_{n}^{B}-T_{n-1}^{B}\right)\right)
$$

Then $\mathcal{D}^{B}(n)$ and $\widetilde{\mathcal{D}}^{B}(n)$ 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 $\left\{B_{i}: i \geq 1\right\}$, the branching times $\left\{T_{i}^{B}: i \geq 1\right\}$ satisfies

$$
\left\{T_{i}^{B}: i \geq 2\right\} \stackrel{d}{=}\left\{\sum_{\sum_{(=2}^{i}}^{\left.\frac{B_{j}}{(2+\delta) j-1}: i \geq 2\right\}( }\right.
$$

and $\frac{n}{e^{(2+\delta) T_{n}^{B}}} \xrightarrow{\text { a.s. }} W_{B}$, with $W_{B} \sim$ Camma $(\nmid+2 \delta, 1)$.
(ii) In $M_{p}(0, \infty]$, we have for $\delta \geq 0$,

$$
\sum_{i=1}^{n} \epsilon_{D_{i}^{B}(n) / n^{1 /(2+\delta)}} \Rightarrow \sum_{i=1}^{\infty} \epsilon \epsilon_{\sigma_{i} e^{-T_{i-1}^{B} / W_{B}^{1 /(2+\delta)}}}
$$

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

$$
D_{(1)}^{B}(n) \geq D_{(2)}^{B}(n) \geq \cdots \geq D_{(n)}^{B}(n)
$$

be the order statistics for $\mathcal{D}^{B}(n)$.

Theorem 13 Suppose that $\left\{k_{n}\right\}$ is some intermediate sequence satisfying (5.1), then

$$
\frac{1}{k_{n}} \sum_{i=1}^{n} \oint_{D_{i}^{B}(n) / D_{\left(k_{n}\right)}^{B}(n)}(\cdot) \Rightarrow \nu_{2+\delta}
$$

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 (5.5). Using van der Hofstad (2017, Proposition 8.4) and Stirling's formula, we know that $I \xrightarrow{P} 0$ and $I I I \rightarrow 0$, respectively, for $\left\{k_{n}\right\}$ satisfying (5.1). So we are left with showing

$$
\frac{1}{k_{n}} \mathbf{E}\left(N_{>\left[b\left(n / k_{n}\right) y\right]}(n)\right)-n p_{>\left[b\left(n / k_{n}\right) y\right]} \rightarrow 0, \quad n \rightarrow \infty .
$$

Note that in Model B, we have

$$
\left\{T_{i}^{B}: i \geq 2\right\} \stackrel{d}{=}\left\{\sum_{x=2}^{i} \frac{B_{j}}{(2+\delta) j-1}: i \geq 2\right\}(
$$

where $\left\{B_{i}: i \geq 1\right\}$ 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 1.2) to conclude that for some constant $C_{\delta}>0$,

$$
\begin{aligned}
\frac{n}{k_{n}} & \mathbf{E}\left(\ell_{n} N_{>\left[b\left(n / k_{n}\right) y\right]}(n)\right)-p_{>\left[b\left(n / k_{n}\right) y\right]} \\
& \leq \frac{n}{k_{n}} \sup _{x \geq 0} \mathbf{E}\left(\ell_{n} N_{>x}(n)\right)-p_{>x}=\frac{n}{k_{n}} \sup _{x \geq 0} \mathbf{E}\left(\not \not\left(D_{U_{n}}^{B}(n)>x\right)\right)-p_{>x} \\
& \leq \frac{n}{k_{n}} C_{\delta} \frac{\log n}{n}=C_{\delta} \frac{\log n}{k_{n}},
\end{aligned}
$$

and $\log n / k_{n} \rightarrow 0$ as $n \rightarrow \infty$, for $\left\{k_{n}\right\}$ satisfying (5.1). This completes the proof of the theorem.

Now define the Hill estimator for Model B as

$$
H_{k_{n}, n}^{B}=\frac{1}{k_{n}} \sum_{i=1}^{k_{n}} \log \frac{D_{(i)}^{B}(n)}{D_{\left(k_{n}+1\right)}^{B}(n)} .
$$

The consistency of $H_{k_{n}, n}^{B}$ 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 $\left\{k_{n}\right\}$ be an intermediate sequence satisfying (5.1), then

$$
H_{k_{n}, n}^{B} \xrightarrow{P} \frac{1}{2+\delta} \quad \text { as } n \rightarrow \infty
$$

## 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.

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 computer science and network analyses; see the KONECT website (Kunegis, 2013). This method is encoded in the plfit script, which can be found at 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 $\alpha$, 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

$$
\left.H_{k, n}^{-1}=\hat{\alpha}(k):=\frac{1}{k} \sum_{i=1}^{k} \log \frac{Z_{(i)}}{Z_{(k+1)}}\right)^{-1}, \quad k \geq 1
$$

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 $\hat{\alpha}(k)$ :

$$
d_{k}:=\sup _{y \geq 1} \frac{1}{k} \sum_{i=1}^{n} \epsilon_{Z_{i} / Z_{(k+1)}}(y, \infty]-y^{-\hat{\alpha}(k)}, \quad 1 \leq k \leq n .
$$

Then the chosen $k^{*}$ is the one that minimizes the KS distance, i.e.

$$
k^{*}:=\underset{1 \leq k \leq n}{\operatorname{argmin}} d_{k},
$$

and we estimate the tail index and threshold by $\hat{\alpha}\left(k^{*}\right)$ and $Z_{\left(k^{*}+1\right)}$ 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, 2018).

In the iid case, under some second order condition, we know that for some intermediate sequence $\left\{k_{n}\right\}$,

$$
\begin{aligned}
& \sqrt{k_{n}}\left(\not \mathscr{f}_{k_{n}, n}-\alpha^{-1}\right) \nRightarrow N\left(0, \alpha^{-2}\right), \quad \text { as } n \rightarrow \infty, \\
& \text { Ferreita (2006): Resnick (2007), Chapter 9.1). Is }
\end{aligned}
$$

(de Haan and Ferreita (2006); Resnick (2007), Chapter 9.1). Is this also true in the preferential attachment setup? We only consider Model A.

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

$$
\sqrt{k^{*}}\left(\hat{\alpha}\left(k^{*}\right)-\alpha\right),(
$$

and examine whether it is close to some normal distribution, provided that we have a large preferential attachment model. To do this, we chose $\alpha=$


Fig. 7.1 QQ plots of $\sqrt{k^{*}}\left(\hat{\alpha}\left(k^{*}\right)-\alpha\right)$ 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.


Fig. 7.2 QQ plots of $\sqrt{k_{n}}\left(\hat{\alpha}\left(k_{n}\right)-\alpha\right)$ 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.
$1.5,2,2.5,3,3.5$ and $n=10^{5}$, generated 500 replications of Model A for each value of $\alpha$ and computed $\sqrt{k^{*}}\left(\hat{\alpha}\left(k^{*}\right)-\alpha\right)$ for each replication. QQ plots corresponding to different values of $\alpha$ are gi en in Figure 7.1. We see that for small values of $\alpha$, the distribution of $\sqrt{k^{*}}\left(\alpha\left(k^{*}\right)-\alpha\right)$ /'s 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

$$
k_{n}=(n \log n)^{1 / 2}, n^{0.55}, n^{0.58}, n^{0.6}, n^{0.65}, n^{0.68}, n^{0.7}
$$

Using the 500 replications obtained under different values of $\alpha$, we computed $\hat{\alpha}\left(k_{n}\right)$ with $k_{n}$ varied as above. Also, for each value of $\alpha$, we recorded the MSE of $\hat{\alpha}\left(k_{n}\right)$ under different choices of $k_{n}$, picked $k_{n}$ giving the smallest MSE and plotted the distribution of the corresponding $\sqrt{k_{n}}\left(\hat{\alpha}\left(k_{n}\right)-\alpha\right) / n$ Figure 7.2. Different from Figure 7.1, most of the QQ plots in Figure 7.2 confirm the normality of $\sqrt{k_{n}}\left(\hat{\alpha}\left(k_{n}\right)-\alpha\right)$. (This suggests that the non-normality displayed in Figure 7.1 is possibly the result of the minimum distance method. Further work on the minimum distance method is ongoing in Drees et al (2018). At the moment, neither the normality of $\sqrt{k_{n}}\left(\hat{\alpha}\left(k_{n}\right)-\alpha\right)$ for that of $\sqrt{k^{*}}\left(\hat{\alpha}\left(k^{*}\right)-\alpha\right)($ has been properly analyzed and this requires more investigation.

## References

K.B. Athreya. Preferential attachment random graphs with general weight function. Internet Mathematics, 4(4):401-418, 2007. doi: 10.1080/15427951. 2007.10129150. http://dx.doi.org/10.1080/15427951.2007.10129150.
K.B. Athreya, A.P. Ghosh, and S. Sethuraman. Growth of preferential attachment random graphs via continuous-time branching processes. Proceedings Mathematical Sciences, 118(3):473494, August 2008.
S. Bhamidi. Universal techniques to analyze preferential attachment trees: Global and local analysis. Available: http://www. unc. edu/~bhamidi/ preferent. pdf, 2007.
B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. Directed scale-free graphs. In Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, 2003), pages 132-139, New York, 2003. ACM.
A. Clauset, C.R. Shalizi, and M.E.J. Newman. Power-law distributions in empirical data. SIAM Rev., 51(4):661-703, 2009. ISSN 0036-1445. doi: 10.1137/070710111. http://dx.doi.org/10.1137/070710111.
S. Csörgö, E. Haeusler, and D.M. Mason. The quantile-transform-empiricalprocess approach to limit theorems for sums of order statistics. In Sums, Trimmed Sums and Extremes, volume 23 of Progr. Probab., pages 215-267. Birkhäuser Boston, Boston, MA, 1991a.
L. de Haan and A. Ferreira. Extreme Value Theory: An Introduction. SpringerVerlag, New York, 2006.
L. de Haan and S.I. Resnick. On asymptotic normality of the Hill estimator. Stochastic Models, 14:849-867, 1998.
H. Drees, A. Janßen, T. Wang and S.I. Resnick. Threshold selection by distance minimization. In preparation.
R.T. Durrett. Random Graph Dynamics. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.
F. Gao and A. van der Vaart. On the asymptotic normality of estimating the affine preferential attachment network models with random initial degrees. Stoch. Process. Appl., 127(11):3754-3775, 2017. ISSN: 0304-4149.
F. Gao, A. van der Vaart, R. Castro, and R. van der Hofstad. Consistent estimation in general sublinear preferential attachment trees Electron. J. Statist., 11(2):3979-3999, 2017. ISSN: 0304-4149. doi: 10.1214/17-EJS1356
P. Hall. On some simple estimates of an exponent of regular variation. J. Roy. Statist. Soc. Ser. B, 44(1):37-42, 1982. ISSN 0035-9246.
B.M. Hill. A simple general approach to inference about the tail of a distribution. Ann. Statist., 3:1163-1174, 1975.
T. Hsing. On tail estimation using dependent data. Ann. Statist., 19:15471569, 1991.
O. Kallenberg. Random Measures, Theory and Applications. Springer Series in Probability Theory and Stochastic Modelling. Springer eBooks, 2017. ISBN:978-3-319-41598-7.
D.G. Kendall. Branching processes since 1873. J. London Math. Soc., 41: 385-406, 1966. ISSN 0024-6107. https://doi.org/10.1112/jlms/s1-41. 1.385.

Konect. KONECT - The Koblenz Network Collection. http://konect. uni-koblenz.de.
P. Krapivsky, G. Rodgers, and S. Redner. Degree distributions of growing networks. Phys. Rev. Lett, 86, 2001. doi: 10.1103/PhysRevLett.86.5401. http://dx.doi.org/10.1103/PhysRevLett.86.5401.
P.L. Krapivsky and S. Redner. Organization of growing random networks. Physical Review E, 63(6):066123:1-14, 2001.
J. Kunegis. Konect: the Koblenz network collection. Proceedings of the 22nd International Conference on World Wide Web, 1343-1350. ACM, 2013.
D. Mason. Laws of large numbers for sums of extreme values. Ann. Probab., 10:754-764, 1982.
D. Mason and T. Turova. Weak convergence of the Hill estimator process. In J. Galambos, J. Lechner, and E. Simiu, editors, Extreme Value Theory and Applications, pages 419-432. Kluwer Academic Publishers, Dordrecht, Holland, 1994.
R. Oliveira and J. Spencer. Connectivity transitions in networks with superlinear preferential attachment. Internet Mathematics, 2(2):121163, 2005.
S.I. Resnick. Extreme Values, Regular Variation and Point Processes. Springer-Verlag, New York, 1987.
S.I. Resnick. Adventures in Stochastic Processes. Birkhäuser, Boston, 1992.
S.I. Resnick. Heavy Tail Phenomena: Probabilistic and Statistical Modeling. Springer Series in Operations Research and Financial Engineering. SpringerVerlag, New York, 2007. ISBN: 0-387-24272-4.
S.I. Resnick and G. Samorodnitsky. Tauberian theory for multivariate regularly varying distributions with application to preferential attachment networks. Extremes, 18(3):349-367, 2015. doi: 10.1007/s10687-015-0216-2.
S.I. Resnick and G. Samorodnitsky. Asymptotic normality of degree counts in a preferential attachment model. Advances in Applied Probability, 48:

283-299, 7 2016. ISSN 1475-6064. doi: 10.1017/apr.2016.56.
S.I. Resnick and C. Stărică. Consistency of Hill's estimator for dependent data. J. Appl. Probab., 32(1):139-167, 1995. ISSN 0021-9002.
S.I. Resnick and C. Stărică. Tail index estimation for dependent data. Ann. Appl. Probab., 8(4):1156-1183, 1998. ISSN 1050-5164.
H. Rootzén, M.R. Leadbetter, and L. de Haan. Tail and quantile estimation for strongly mixing stationary sequences. Technical Report 292, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260, 1990.
N. Ross. Power laws in preferential attachment graphs and Stein's method for the negative binomial distribution. Advances in Applied Probability, 45(3): 876-893, 2013.
A. Rudas, B. Tóth, and B. Valkó. Random trees and general branching processes. Random Structures Algorithms, 31(2):186-202, 2007.
G. Samorodnitsky, S. Resnick, D. Towsley, R. Davis, A. Willis, and P. Wan. Nonstandard regular variation of in-degree and out-degree in the preferential attachment model. Journal of Applied Probability, 53(1):146-161, 2016.
S. Tavaré. The birth process with immigration, and the genealogical structure of large populations. Journal of Mathematical Biology, 25(2):161-168, 1987.
R. van der Hofstad. 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. doi: 10.1017/ 9781316779422. http://dx.doi.org/10.1017/9781316779422.
P. Wan, T. Wang, R. A. Davis, and S. I. Resnick. Fitting the linear preferential attachment model. Electron. J. Statist., 11(2):3738-3780, 2017. ISSN 19357524. doi: 10.1214/17-EJS1327.
P. Wan, T. Wang, R. A. Davis, and S. I. Resnick. Are extreme value estimation methods useful for network data? Submitted, 2017.
T. Wang and S. I. Resnick. Asymptotic normality of in- and out-degree counts in a preferential attachment model. Stochastic Models, 33(2):229255, 2017. doi: 10.1080/15326349.2016.1256219. http://dx.doi.org/10. 1080/15326349.2016.1256219.
T. Wang and S.I. Resnick. Multivariate regular variation of discrete mass functions with applications to preferential attachment networks. Methodology and Computing in Applied Probability, pages 1-14, 2016. ISSN 1573-7713. doi: 10.1007/s11009-016-9503-x. http://dx.doi.org/10.1007/ s11009-016-9503-x.
W. A. O'N. Waugh. Transformation of a birth process into a Poisson process. J. Roy. Statist. Soc. Ser. B, 32:418-431, 1970. ISSN 00359246. http://links.jstor.org/sici?sici=0035-9246(1970)32:3<418: TOABPI>2.0.CO;2-A\&origin=MSN.

