Annals of the Institute of Statistical Mathematics (2022) 74:957–986
https://doi.org/10.1007/s10463-022-00827-5

Directed hybrid random networks mixing preferential attachment with uniform attachment mechanisms

Tiandong Wang[1] · **Panpan Zhang**[2]

Received: 17 May 2021 / Revised: 5 January 2022 / Accepted: 17 January 2022 /
Published online: 23 April 2022

Abstract
Motivated by the complexity of network data, we propose a directed hybrid random network that mixes preferential attachment (PA) rules with uniform attachment rules. When a new edge is created, with probability $p \in (0, 1)$, it follows the PA rule. Otherwise, this new edge is added between two uniformly chosen nodes. Such mixture makes the in- and out-degrees of a fixed node grow at a slower rate, compared to the pure PA case, thus leading to lighter distributional tails. For estimation and inference, we develop two numerical methods which are applied to both synthetic and real network data. We see that with extra flexibility given by the parameter p, the hybrid random network provides a better fit to real-world scenarios, where lighter tails from in- and out-degrees are observed.

Keywords Preferential attachment · Uniform attachment · In- and out-degrees · Power laws · Random networks

1 Introduction

The *preferential attachment* (PA) mechanism (Barabási and Albert 1999) has been widely used to model interactions or communications among the entities in a network-based system, especially evolving networks. A precursory study of PA networks was conducted by de Sollar Price (1965) to model the growth of citation networks, where the research outcome coincides with a sociological theory called the *Matthew Effect* (Merton 1968), inducing a well known economic manifestation—"The rich get richer; the poor get poorer".

✉ Tiandong Wang
twang@stat.tamu.edu

[1] Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843, USA

[2] Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, 423 Guardian Drive, 501, Philadelphia, PA 19104, USA

One of the most appealing properties of the PA network is *scale-free* (i.e., the node degree distribution follows a *power law*), rendering that the PA rule has become an attractive choice for real network modeling, such as the World Wide Web (Henzinger and Lawrence 2004) and collaboration networks Newman (2001). We refer the readers to Durrett (2006); van der Hofstad (2017) for some text-style elaborations of the elementary descriptions and probabilistic properties of PA networks. Recent studies have extended classical PA networks to directed counterparts, where degree distributions and maximum degrees (Cooper and Frieze 2003), asymptotic theories (Wang and Resnick 2015, 2018, 2020) and the *maximum likelihood estimators* (MLEs) of the parameters (Wan et al. 2017) have been developed. Other recent works on the mathematical treatments of PA networks and their variants include Gao and van der Vaart (2017); Alves et al. (2019); Mahmoud (2019); Wang and Resnick (2020); Zhang and Mahmoud (2020).

However, classical (either directed or undirected) PA networks do not always fit the real network data well, nor are they able to precisely capture some key attributes of the networks. Alternatively, Atalay et al. (2011) proposed a model mixing PA and *uniform attachment* (UA) to investigate the buyer-supplier network in the United States, showing that the proposed model has outperformed the pure PA model. In this paper, we consider a class of directed *hybrid random networks* (HRNs) presenting PA and UA mechanisms simultaneously, governed by a tuning parameter $p \in (0, 1)$. The presence of UA in the proposed model effectively leverages the heavy tail produced by the PA mechanism, rendering the model tentatively better fits the real networks whose degree distributions are less heavier.

In the literature, there is a limited amount of work on the random structures that integrate PA and UA during the evolution. Cooper and Frieze (2003) looked into the degree sequences in an undirected random network model mixing PA and UA. Shao et al. (2006) carried out a simulation study of the degree distribution in a standard mixed attachment growing network. More recently, Pachon et al. (2018) investigated the scale-free property of the degree distribution in an analogous model through recursive formulations, and Medina et al. (2019) established an Expectation-Maximization (EM) algorithm for parameter estimation of a similar model. The rest of the manuscript is organized as follows. We describe the construction of a hybrid random network in Sect. 2, and study theoretical properties of its degree distributions in Sect. 3. We then propose estimation methods and explore properties of the estimators in Sect. 4, which facilitate the numerical studies on both synthetic datasets (cf. Sect. 5) and real network data (cf. Sect. 6). With all results available, we also provide some interesting direction for future research in Sect. 7.

2 Hybrid random networks

Let $\mathcal{H}_n(V_n, E_n; \alpha, \beta, \gamma, p, \delta_{\text{in}}, \delta_{\text{out}})$ denote the structure of a class of HRNs consisting of a vertex set V_n and an edge set E_n at time n, parameterized by a set of parameters $\boldsymbol{\theta} := (\alpha, \beta, \gamma, p, \delta_{\text{in}}, \delta_{\text{out}})$ subject to $\alpha + \beta + \gamma = 1$, $\delta_{\text{in}}, \delta_{\text{out}} > 0$. Specifically, the parameters, α, β and γ, represent the probabilities of presenting one of the three edge-creation scenarios at each step. With probability α, there emerges

a directed edge from the newcomer to an existing node. With probability γ, there emerges a directed edge from an existing node to the newcomer. With probability $\beta = 1 - \alpha - \gamma$, a directed edge is added between two existing nodes. See Fig. 1 for a graphical illustration. The offset parameters δ_{in} and δ_{out} respectively control the growth rate of in-degree and out-degree in the network. Another parameter $0 \le p \le 1$ specifies the probability of executing PA when sampling the node(s) at the end(s) of the newly added edge at each timestamp. The functionality of p is to balance PA and UA in the model, and accordingly the proposed HRN becomes more flexible than pure PA network model for characterizing the in-degree and out-degree tail distributions of real network data.

We start the network with $\mathcal{H}_0$, which is a self-looped single node labeled with 1. At any subsequent point $n \ge 1$, flip a three-sided coin, for which the probabilities of landing the three faces up are respectively α (associated with scenario 1), β (associated with scenario 2) and γ (associated with scenario 3). Let $J_n \in \{1, 2, 3\}$ indicate the occurrence of the scenario type at time n, i.e. J_n is a tri-nomial random variable on $\{1, 2, 3\}$ with cell probability α, β and γ, respectively. The network evolves as below.

1. For $J_n = 1$, we add a new node u to the network, connecting it to an existing node $i \in V_{n-1}$ by a directed edge pointing to i with probability

$$p \times \frac{D_i^{\text{in}}(n-1) + \delta_{\text{in}}}{\sum_{k \in V_{n-1}} (D_k^{\text{in}}(n-1) + \delta_{\text{in}})} + (1-p) \times \frac{1}{|V_{n-1}|}, \tag{1}$$

   where $D_i^{\text{in}}(n)$ is the in-degree of i in $\mathcal{H}_n$, and $|V_n|$ denotes the number of nodes at time n.
2. For $J_n = 2$, we add a directed edge between two existing nodes $i, j \in V_{n-1}$, where i and j are sampled independently. Suppose that the newly added edge is pointed (from j) to i, then the associated probability is given by

$$\left(p \times \frac{D_j^{\text{out}}(n-1) + \delta_{\text{out}}}{\sum_{k \in V_{n-1}} (D_k^{\text{out}}(n-1) + \delta_{\text{out}})} + (1-p) \times \frac{1}{|V_{n-1}|} \right) \times \left(p \times \frac{D_i^{\text{in}}(n-1) + \delta_{\text{in}}}{\sum_{k \in V_{n-1}} (D_k^{\text{in}}(n-1) + \delta_{\text{in}})} + (1-p) \times \frac{1}{|V_{n-1}|} \right), \tag{2}$$

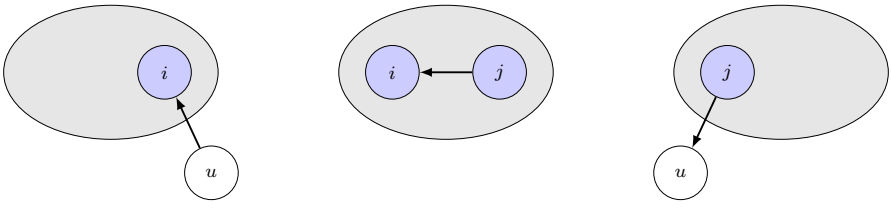

Fig. 1 Three edge-addition scenarios respectively corresponding to α, β and γ (from left to right)

where $D_i^{\text{out}}(n)$ is analogously defined as the out-degree of node i in $\mathcal{H}_n$. Note that no new node is added to the network under this scenario, hence $V_n = V_{n-1}$. Besides, there is a positive probability that a node is sampled twice; If so, a self loop is created.

3. For $J_n = 3$, a new node u is appended to the network by a directed edge emanating out from $j \in V_{n-1}$ with probability

$$p \times \frac{D_j^{\text{out}}(n-1) + \delta_{\text{out}}}{\sum_{k \in V_{n-1}} (D_k^{\text{out}}(n-1) + \delta_{\text{out}})} + (1-p) \times \frac{1}{|V_{n-1}|}. \tag{3}$$

Some simplifications can be made to the conditional probabilities in Eqs. (1), (2) and (3) after observing $\sum_{k \in V_{n-1}} (D_k^{\text{in}}(n-1) + \delta_{\text{in}}) = n + \delta_{\text{in}}|V_{n-1}|$ and $\sum_{k \in V_{n-1}} (D_k^{\text{out}}(n-1) + \delta_{\text{out}}) = n + \delta_{\text{out}}|V_{n-1}|$ (since our initial time is $n = 0$). Meanwhile, the fact that the two fractions have different denominators in each of the conditional probabilities would have brought a great deal of challenges to both analytical computations and parameter estimations.

3 Asymptotic results for the degree distributions

In this section, we investigate the in-degree and out-degree distributions of $\mathcal{H}_n$, and we will focus on both the degree growth of a fixed node and the empirical in- and out-degree distributions.

3.1 Degree growth for a fixed node

Let $\mathcal{F}_n$ be the sigma field generated by the evolution of a hybrid random network up to time n, i.e., $\{\mathcal{H}_k : k \le n\}$. According to the evolutionary scenarios described in Sect. 2, we have for $i \in V_n$,

$$\begin{aligned} &\mathbb{P}\left(D_i^{\text{in}}(n+1) - D_i^{\text{in}}(n) = 1 \mid \mathcal{F}_n\right) \\ &\quad = (\alpha + \beta)\left[\frac{p\left(D_i^{\text{in}}(n) + \delta_{\text{in}}\right)}{\sum_{k \in V_n}\left(D_k^{\text{in}}(n) + \delta_{\text{in}}\right)} + \frac{(1-p)}{|V_n|}\right], \end{aligned} \tag{4}$$

$$\begin{aligned} &\mathbb{P}\left(D_i^{\text{out}}(n+1) - D_i^{\text{out}}(n) = 1 \mid \mathcal{F}_n\right) \\ &\quad = (\beta + \gamma)\left[\frac{p\left(D_i^{\text{out}}(n) + \delta_{\text{out}}\right)}{\sum_{k \in V_n}\left(D_k^{\text{out}}(n) + \delta_{\text{out}}\right)} + \frac{(1-p)}{|V_n|}\right]. \end{aligned} \tag{5}$$

We present important theoretical results on the in- and out-degree sequences in an HRN. We also include detailed proofs for Theorem 2 in Appendix A, and other technical proofs are collected in the supplement.

We start with the asymptotic properties of $(D_i^{\text{in}}(n), D_i^{\text{out}}(n))$ for a fixed node i, by utilizing *martingale* formulations (Durrett 2006, Chapter 4). By the conditional probability in Eq. (4), we have for $i \in V_n$,

$$\mathbb{P}\big(D_i^{\text{in}}(n+1) - D_i^{\text{in}}(n) = 1 \mid \mathcal{F}_n\big) \geq \frac{p\big(D_i^{\text{in}}(n) + \delta_{\text{in}}\big)(\alpha+\beta)}{1+n+\delta_{\text{in}}|V_n|},$$

which implies that for a fixed i,

$$M_{n+1}^{\text{in}} := \frac{D_i^{\text{in}}(S_i+n+1)+\delta_{\text{in}}}{\prod_{k=0}^{n}\left(1+\frac{p(\alpha+\beta)}{S_i+k+1+\delta_{\text{in}}|V_{S_i+k}|}\right)} \tag{6}$$

is a *sub-martingale* with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$, where $S_i := \inf\{n \geq 0 : |V_n| = i\}$. Analogously, based on Eq. (5), we construct another sub-martingale sequence for out-degrees:

$$M_{n+1}^{\text{out}} := \frac{D_i^{\text{out}}(S_i+n+1)+\delta_{\text{out}}}{\prod_{k=0}^{n}\left(1+\frac{p(\beta+\gamma)}{S_i+k+1+\delta_{\text{out}}|V_{S_i+k}|}\right)}. \tag{7}$$

In the proof of Theorem 1, we specify the asymptotic orders of the denominator in Eq. (6) (a similar argument also applies to the denominator in Eq. (7)). Then applying the martingale convergence theorem (Durrett 2006, Theorem 4.2.11) gives the following convergence results for the in- and out-degrees of a fixed node. Technical details of the proof of Theorem 1 are given in Sect. 1 of the supplement.

Theorem 1 *Set*

$$C_1 = \frac{(\alpha+\beta)p}{1+\delta_{\text{in}}(1-\beta)} \qquad \text{and} \qquad C_2 = \frac{(\beta+\gamma)p}{1+\delta_{\text{out}}(1-\beta)}.$$

Then for a fixed node i, there exist finite random variables ζ_i and ξ_i such that as $n \to \infty$,

$$\left(\frac{D_i^{\text{in}}(n)}{n^{C_1}}, \frac{D_i^{\text{out}}(n)}{n^{C_2}}\right) \xrightarrow{a.s.} (\zeta_i, \xi_i).$$

It is worth noting that the growth rates C_1 and C_2 are smaller than those in a pure directed PA model (i.e., $p = 1$). This suggests that incorporating a non-negligible number of uniformly added edges create lighter distributional tails for both in- and out-degrees.

Further, in the next proposition, we specify the growth rates of $\mathbb{E}\left[\left(D_i^{\text{in}}(n)\right)^2\right]$ and $\mathbb{E}\left[\left(D_i^{\text{out}}(n)\right)^2\right]$ for a fixed node $i \in V_n$.

Proposition 1 *There exist $M_1, M_2 > 0$ such that*

$$\sup_{i\geq 1}\frac{\mathbb{E}\left[\left(D_i^{\text{in}}(n)\right)^2\right]}{n^{2C_1}}\leq M_1 \qquad \textit{and} \qquad \sup_{i\geq 1}\frac{\mathbb{E}\left[\left(D_i^{\text{out}}(n)\right)^2\right]}{n^{2C_2}}\leq M_2,$$

where C_1 and C_2 are identical to those specified in Theorem 1.

The full proof of Proposition 1 is collected in Sect. 2 of the supplement. Then with Proposition 1 developed, we apply Theorem 4.6.2 in Durrett (2019) to conclude that both $\{D_i^{\text{in}}(n)/n^{C_1} : n\geq 1\}$ and $\{D_i^{\text{out}}(n)/n^{C_2} : n\geq 1\}$ are uniformly integrable. Therefore, using Theorem 1, we also have

$$\frac{D_i^{\text{in}}(n)}{n^{C_1}}\xrightarrow{L_1}\zeta_i, \qquad \frac{D_i^{\text{out}}(n)}{n^{C_2}}\xrightarrow{L_1}\xi_i.$$

3.2 Degree counts

Let $N_m^{\text{in}}(n)$ and $N_m^{\text{out}}(n)$ respectively denote the number of nodes of in-degree m and out-degree m in an HRN at time n. We develop the asymptotics for $N_m^{\text{in}}(n)/n$, and $N_m^{\text{out}}(n)/n$, $m\geq 0$, i.e., the empirical proportional of nodes with in- or out-degree m in $\mathcal{H}_n$. Let $\text{NB}(n,q)$ represent a negative binomial random variable with generating function

$$(s+(1-s)/q)^{-n}, \qquad s\in[0,1].$$

Theorem 2 *Define* $\tilde{\delta}_{\text{in}} := \delta_{\text{in}}/p+(1-p)/\big(p(1-\beta)\big)$ *and* $\tilde{\delta}_{\text{out}} := \delta_{\text{out}}/p+(1-p)/\big(p(1-\beta)\big)$. *Let* $\text{NB}(\tilde{\delta}_{\text{in}},p_1)$, $\text{NB}(1+\tilde{\delta}_{\text{in}},p_1)$, $\text{NB}(\tilde{\delta}_{\text{out}},p_2)$ *and* $\text{NB}(1+\tilde{\delta}_{\text{out}},p_2)$, *be four independent negative binomial random variables, and set* T_{in} *and* T_{out} *to be two independent exponential random variables with rates*

$$\frac{1+\delta_{\text{in}}(1-\beta)}{p(\alpha+\beta)} \qquad \textit{and} \qquad \frac{1+\delta_{\text{out}}(1-\beta)}{p(\beta+\gamma)},$$

respectively (*which are also independent from* $\text{NB}(\tilde{\delta}_{\text{in}},p_1)$, $\text{NB}(1+\tilde{\delta}_{\text{in}},p_1)$, $\text{NB}(\tilde{\delta}_{\text{out}},p_2)$ *and* $\text{NB}(1+\tilde{\delta}_{\text{out}},p_2)$). *As* $n\to\infty$, *we have*

$$\frac{N_m^{\text{in}}(n)}{n}\xrightarrow{p}\tilde{\psi}_m^{\text{in}} \qquad \textit{and} \qquad \frac{N_m^{\text{out}}(n)}{n}\xrightarrow{p}\tilde{\psi}_m^{\text{out}}, \tag{8}$$

where

$$\begin{aligned}\tilde{\psi}_m^{\text{in}} = {}& \alpha\,\mathbb{P}\big(\text{NB}\big(\tilde{\delta}_{\text{in}},e^{-T_{\text{in}}}\big)=m\big)\\ &+\gamma\,\mathbb{P}\big(1+\text{NB}\big(1+\tilde{\delta}_{\text{in}},e^{-T_{\text{in}}}\big)=m\big),\end{aligned} \tag{9}$$

$$\tilde{\psi}_m^{\text{out}} = \alpha\, \mathbb{P}\big(\text{NB}\big(\tilde{\delta}_{\text{out}}, e^{-T_{\text{out}}}\big) = m\big) + \gamma\, \mathbb{P}\big(1 + \text{NB}\big(1 + \tilde{\delta}_{\text{out}}, e^{-T_{\text{out}}}\big) = m\big). \tag{10}$$

In the proof of Theorem 2, we will first show the concentration of $N_m^{\text{in}}(n)$ around $\mathbb{E}(N_m^{\text{in}}(n))$, i.e. there exists some constant $C > 2\sqrt{2}$ such that

$$\mathbb{P}\left(\left|N_m^{\text{in}}(n) - \mathbb{E}(N_m^{\text{in}}(n))\right| \geq C\sqrt{n \log n}(1 + \log n)\right) = o(1/n).$$

This further implies

$$\mathbb{P}\left(\max_{m \geq 0}\left|N_m^{\text{in}}(n) - \mathbb{E}(N_m^{\text{in}}(n))\right| \geq C\sqrt{n \log n}(1 + \log n)\right) = o(1),$$

as $m \leq n$ in $G(n)$. Then we approximate $\mathbb{E}(N_m^{\text{in}}(n))/n$ by $\tilde{\psi}_m^{\text{in}}$. A similar argument is also applicable to $N_m^{\text{out}}(n)$.

We conclude this section by remarking that the limit functions in Eqs. (9) and (10) coincide with those from a pure PA network with parameters $(\alpha, \beta, \gamma, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}})$. In fact, when $\beta = 0$, the HRN is identical to a pure PA network with $(\alpha, 0, \gamma, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}})$, where all established results for the pure PA model can be readily applied. The major goal in the proof of Theorem 2 is to show that the discrepancy caused by having random number of edges is negligible when n is large.

4 Parameter estimation

In this section, we propose our estimation scheme for the parameters in the HRN model described in Sect. 2, under a few regularity conditions given as follows. We assume the evolution history of the entire network is available since the beginning, recorded in the edge list $\boldsymbol{E} := \{E_k\}_{k=0}^{n-1}$, where $E_0 = (1, 1)$ is deterministic. Notice that $\gamma = 1 - (\alpha + \beta)$ completely depends on α and β so that the model is parametrized in terms of $(\alpha, \beta, p, \delta_{\text{in}}, \delta_{\text{out}})$. We also assume $0 < p < 1$, $0 \leq \alpha, \beta < 1$ and $0 < \alpha + \beta \leq 1$, where the latter two jointly ensure the exclusion of the trivial cases of either α, β or γ taking value 1. Besides, the value of p is not allowed to take 0 or 1 such that UA and PA mechanisms co-exist in the model to ensure identifiability. The offset parameters δ_{in} and δ_{out} are assumed to be positive and finite.

4.1 Maximum likelihood estimation

In a slight abuse of notation, let $E_k = (v_{k,1}, v_{k,2})$ represent the edge (from $v_{k,1}$ to $v_{k,2}$) added at time k, $v_{k,1}$ and $v_{k,2}$ can be the nodes from the existing network or newcomers. According to Eqs. (1), (2) and (3), the likelihood of the model is given by

$$
\begin{aligned}
L(\boldsymbol{\theta}\,|\,\boldsymbol{E}) = \prod_{k=1}^{n}\left[\alpha\left(\frac{p(D^{\text{in}}_{v_{k,2}}(k-1)+\delta_{\text{in}})}{k+\delta_{\text{in}}|V_{k-1}|}+\frac{1-p}{|V_{k-1}|}\right)\right]^{\mathbb{I}_{\{J_k=1\}}} \\
\times\prod_{k=1}^{n}\left[\beta\left(\frac{p(D^{\text{out}}_{v_{k,1}}(k-1)+\delta_{\text{out}})}{k+\delta_{\text{out}}|V_{k-1}|}+\frac{1-p}{|V_{k-1}|}\right)\right. \\
\left.\times\left(\frac{p(D^{\text{in}}_{v_{k,2}}(k-1)+\delta_{\text{in}})}{k+\delta_{\text{in}}|V_{k-1}|}+\frac{1-p}{|V_{k-1}|}\right)\right]^{\mathbb{I}_{\{J_k=2\}}} \\
\times\prod_{k=1}^{n}\left[(1-\alpha-\beta)\left(\frac{p(D^{\text{out}}_{v_{k,1}}(k-1)+\delta_{\text{out}})}{k+\delta_{\text{out}}|V_{k-1}|}+\frac{1-p}{|V_{k-1}|}\right)\right]^{\mathbb{I}_{\{J_k=3\}}},
\end{aligned}
$$

then the log-likelihood becomes

$$
\begin{aligned}
&\log L(\boldsymbol{\theta}\,|\,\boldsymbol{E}) \\
&= \log\alpha\sum_{k=1}^{n}\mathbb{I}_{\{J_k=1\}}+\log\beta\sum_{k=1}^{n}\mathbb{I}_{\{J_k=2\}}+\log(1-\alpha-\beta)\sum_{k=1}^{n}\mathbb{I}_{\{J_k=3\}} \\
&\quad+\sum_{k=1}^{n}\log\left[\left(pD^{\text{in}}_{v_{k,2}}(k-1)+\delta_{\text{in}}\right)|V_{k-1}|+(1-p)k\right]\mathbb{I}_{\{J_k\in\{1,2\}\}} \\
&\quad+\sum_{k=1}^{n}\log\left[\left(pD^{\text{out}}_{v_{k,1}}(k-1)+\delta_{\text{out}}\right)|V_{k-1}|+(1-p)k\right]\mathbb{I}_{\{J_k\in\{2,3\}\}} \\
&\quad-\sum_{k=1}^{n}\log\left[k|V_{k-1}|+|V_{k-1}|^2\delta_{\text{in}}\right]\mathbb{I}_{\{J_k\in\{1,2\}\}} \\
&\quad-\sum_{k=1}^{n}\log\left[k|V_{k-1}|+|V_{k-1}|^2\delta_{\text{out}}\right]\mathbb{I}_{\{J_k\in\{2,3\}\}},
\end{aligned}
\tag{11}
$$

from which we see that the score functions of α and β are independent of those of the other parameters.

Carrying out an analogous analysis as in (Wan et al. 2017, Sect. 3.1), we find that the MLEs for α and β are respectively given by

$$
\hat{\alpha}^{MLE}=\frac{1}{n}\sum_{k=1}^{n}\mathbb{I}_{\{J_k=1\}}\qquad\text{and}\qquad\hat{\beta}^{MLE}=\frac{1}{n}\sum_{k=1}^{n}\mathbb{I}_{\{J_k=2\}}.
$$

To develop the MLEs for δ_{in}, δ_{out} and p, we solve their score equations (cf. (16), (17) and (18)). In Appendix B, we present the approximation procedure for solving the score equations, and come up with a conclusion that previously adopted methods in Wan et al. (2017) fail to attain the desired MLE solutions. Alternatively, in the next section, we formulate the MLE searching procedure as a nonlinear optimization problem with constraint $p\in(0,1)$.

4.2 Nonlinear optimization

Noticing that a former approach that approximates the MLE does not produce reasonable estimates, we consider the MLE-searching procedure as a nonlinear optimization problem with properly identified constraints such that $\alpha, \beta, p \in (0, 1)$ and $\delta_{\text{in}}, \delta_{\text{out}} > 0$. Specifically, we adopt the *Nelder-Mead* (N-M) algorithm (Nelder and Mead 1965), which is appealing for efficiency and fast convergence. Despite limited knowledge about the theoretical results of the N-M algorithm (Lagarias et al. 1998), its utilization is widespread in the community since it generally performs well in practice. One practical issue of the algorithm is that its convergence is quite sensitive to the choice of the initial simplex. An improper initial simplex is usually the main cause of the algorithm breakdown.

Having this in mind, we back up with an alternative—a *Bayesian* estimation based on *Markov chain Monte Carlo* (MCMC) algorithms. Specifically, we consider a *Metropolis-Hastings* (M-H) algorithm (Metropolis et al. 1953; Hastings 1970). Being a classical approach, the fundamentals of the M-H algorithms have been extensively elaborated in a wide range of texts (Chen et al. 2010; Liang et al. 2010; Gelman et al. 2013). Hence we only present a few essential steps.

Let $\pi(\boldsymbol{\theta};\boldsymbol{\psi})$ be the prior distribution of $\boldsymbol{\theta}$, where $\boldsymbol{\psi}$ is a collection of hyper-parameters. Under this setting, the likelihood function $L(\boldsymbol{\theta} \mid \boldsymbol{E})$ is the posterior distribution of $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}^{(t)}$ be the estimates from the t-th iteration, and let $Q(\boldsymbol{\theta}^{\text{prop}} \mid \boldsymbol{\theta}^{(t)})$ denote the proposal density governing the transition probability from the current estimates to a proposed set of candidates. Suppose that the distribution Q is symmetric, then the acceptance rate $a(\boldsymbol{\theta}^{\text{prop}} \mid \boldsymbol{\theta}^{(t)})$ is given by

$$\min\left\{\frac{L(\boldsymbol{\theta}^{\text{prop}} \mid \boldsymbol{E})Q(\boldsymbol{\theta}^{\text{prop}} \mid \boldsymbol{\theta}^{(t)})}{L(\boldsymbol{\theta}^{(t)} \mid \boldsymbol{E})Q(\boldsymbol{\theta}^{(t)} \mid \boldsymbol{\theta}^{\text{prop}})}, 1\right\} = \min\left\{\frac{L(\boldsymbol{\theta}^{\text{prop}} \mid \boldsymbol{E})}{L(\boldsymbol{\theta}^{(t)} \mid \boldsymbol{E})}, 1\right\}.$$

This can be done by generating a standard uniform random variable U such that $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{\text{prop}}$ if $U < a(\boldsymbol{\theta}^{\text{prop}} \mid \boldsymbol{\theta}^{(t)})$; $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$, otherwise.

There are multiple ways of selecting proposal distribution $Q(\boldsymbol{\theta}^{\text{prop}} \mid \boldsymbol{\theta}^{(t)})$, where a simple approach based on random walk that ensures symmetric Q is adopted. One drawback of MCMC algorithms is the lack of theoretical foundation for the assessment of convergence. Since the initial samples of $\boldsymbol{\theta}$ from a proper prior distribution may fall into a low density of the target posterior distribution, a sufficiently large *burn-in* period is always necessary. The number of iterations needed for the algorithm to converge is closely related to its convergence rate (Mengersen and Tweedie 1996), which is practically unwieldy in general. Here we will rely on a few widely-accepted graphical diagnostics to assess the convergence of MCMC algorithms, such as time-series plots and running mean plots (Smith 2007).

5 Simulations

In this section, we carry out an extensive simulation study along with a sensitivity analysis for estimating the parameters of HRN. We focus on the performance of the N-M and M-H algorithms under different combinations of (α, β, p). Specifically, we consider $p \in \{0.8, 0.6, 0.2\}$ (respectively corresponding to dominant PA, roughly even PA and UA and dominant UA) paired with $(\alpha, \beta) \in \{(0.8, 0.1), (0.45, 0.1), (0.1, 0.8)\}$. The other two offset parameters are set to be $\delta_{\text{in}} = 1.3$ and $\delta_{\text{out}} = 0.7$.

For each setting, we generate $R = 100$ replicates of independent HRNs with $n = 10^4$ edges. The optimization functions from R package `optimx` (Nash 2014) are used to implement the N-M algorithm, where all the maximum iteration is set to 500 (default). When applying the M-H algorithm, we use non-informative priors. The burn-in number is set to 10,000, and the number of iterations after burn-in is 20,000. To avoid auto-correlation in the posterior sample, a *thinning* sampling of gap 500 is used. In Tables 1, 2 and 3, we present the point estimates, the absolute percentages of bias, and the standard errors based on the simulation results. Note that we only report the statistics for p, δ_{in} and δ_{out}, as the MLEs of α and β are in closed forms, allowing us to compute their estimates and corresponding sample standard deviations mathematically. For all considered scenarios, the estimates of α and β are unbiased, and the standard errors are infinitesimal.

Overall, both algorithms provide estimates with low bias for p, across all combinations of the parameters, but we do observe that the N-M algorithm outperforms the M-H algorithm.

In particular, for p which controls the percentage of edges produced by the PA rule, the N-M algorithm is preferred as the standard errors are consistently smaller. Estimates for δ_{in} and δ_{out} from both algorithms tend to be biased, especially when p is small (cf. Table 3; when the UA part dominates), though the estimated δ_{in} is slightly less biased than the estimated δ_{out}. When $p = 0.8$, Table 1 reveals that the N-M method produces the best (small bias and small standard errors) estimated

Table 1 Simulation results with large $p = 0.8$

Parameters		Nelder-Mead Algorithm			Metropolis-Hastings Algorithm		
		$\hat{p}$	$\hat{\delta}_{\text{in}}$	$\hat{\delta}_{\text{out}}$	$\hat{p}$	$\hat{\delta}_{\text{in}}$	$\hat{\delta}_{\text{out}}$
$\alpha = 0.1$	Est.	0.7908	1.2304	0.6345	0.8183	1.4513	0.8331
$\beta = 0.8$	Bias(%)	1.1653	5.6584	10.3172	2.2357	10.4235	15.9719
$p = 0.8$	S.E.	0.0010	0.0103	0.0058	0.0079	0.0640	0.0567
$\alpha = 0.8$	Est.	0.7611	1.1776	0.6251	0.8128	1.3351	0.8302
$\beta = 0.1$	Bias(%)	5.1091	10.3971	11.9915	1.5728	2.6264	15.6826
$p = 0.8$	S.E.	0.0048	0.0154	0.0138	0.0118	0.0361	0.0449
$\alpha = 0.45$	Est.	0.8153	1.3608	0.7339	0.8305	1.4193	0.7714
$\beta = 0.1$	Bias(%)	1.8762	4.6677	4.6199	3.6774	8.4022	9.2520
$p = 0.8$	S.E.	0.0049	0.0156	0.0113	0.0098	0.0324	0.0236

Table 2 Simulation results with moderate $p = 0.6$

Parameters		Nelder-Mead Algorithm			Metropolis-Hastings Algorithm		
		$\hat{p}$	$\hat{\delta}_{\text{in}}$	$\hat{\delta}_{\text{out}}$	$\hat{p}$	$\hat{\delta}_{\text{in}}$	$\hat{\delta}_{\text{out}}$
$\alpha = 0.1$	Est.	0.6001	1.3340	0.7012	0.6216	1.5576	0.9057
$\beta = 0.8$	Bias(%)	0.0101	2.5520	0.1694	3.4725	16.5359	22.7067
$p = 0.6$	S.E.	0.0016	0.0162	0.0134	0.0077	0.0819	0.0704
$\alpha = 0.8$	Est.	0.5713	1.1725	0.6319	0.6537	1.4981	1.0724
$\beta = 0.1$	Bias(%)	5.0156	10.8786	10.7706	8.2084	13.2215	34.7240
$p = 0.6$	S.E.	0.0034	0.0115	0.0146	0.0144	0.0564	0.0752
$\alpha = 0.45$	Est.	0.6227	1.4097	0.7746	0.6643	1.5867	0.9064
$\beta = 0.1$	Bias(%)	3.6493	7.7825	9.6363	9.6789	18.0667	22.7719
$p = 0.6$	S.E.	0.0034	0.0147	0.0104	0.0154	0.0637	0.0468

Table 3 Simulation results with small $p = 0.2$

Parameters		Nelder-Mead Algorithm			Metropolis-Hastings Algorithm		
		$\hat{p}$	$\hat{\delta}_{\text{in}}$	$\hat{\delta}_{\text{out}}$	$\hat{p}$	$\hat{\delta}_{\text{in}}$	$\hat{\delta}_{\text{out}}$
$\alpha = 0.1$	Est.	0.2021	1.2424	0.8122	0.2079	1.4151	1.0505
$\beta = 0.8$	Bias(%)	1.0335	4.6396	13.8146	3.8097	8.1307	33.3683
$p = 0.2$	S.E.	0.0010	0.0249	0.0268	0.0023	0.0788	0.0750
$\alpha = 0.8$	Est.	0.1971	1.2281	0.8923	0.2321	1.6970	1.2663
$\beta = 0.1$	Bias(%)	1.4966	5.8588	21.5482	13.8237	23.3961	44.7213
$p = 0.2$	S.E.	0.0039	0.0410	0.0436	0.0071	0.0902	0.0996
$\alpha = 0.45$	Est.	0.2125	1.4816	0.8271	0.2398	1.7302	1.2277
$\beta = 0.1$	Bias(%)	5.8757	12.2570	15.3696	16.6028	24.8628	42.9845
$p = 0.2$	S.E.	0.0026	0.0265	0.0249	0.0072	0.0703	0.1108

$(\delta_{\text{in}}, \delta_{\text{out}})$ for the combination $(\alpha, \beta, \gamma) = (0.45, 0.1, 0.45)$. When $p = 0.2$ or 0.6, from Tables 2 and 3, we see the most accurately estimated $(\delta_{\text{in}}, \delta_{\text{out}})$ appear in case $(\alpha, \beta, \gamma) = (0.1, 0.8, 0.1)$. Especially for $p = 0.6$, the N-M algorithm provides the most accurate estimation overall.

The simulation results reveal that $\hat{p}$ is unbiased from both algorithms. However, there is a noticeable efficiency gain in estimating the tuning parameter p by using the N-M algorithm, rendering it a preferred approach. Besides, we look into the distributions of the estimates. The standard central limit theorem ensures that the limiting distributions of $\hat{\alpha}^{MLE}$ and $\hat{\beta}^{MLE}$ are normal. However, since the score functions for δ_{in}, δ_{out} and p are not separable, no standard approach can be readily used to uncover their limiting distributions. Nonetheless, the approximation method developed in Theorem 2 plausibly suggests that they may follow a Gaussian law asymptotically as well. To verify, we show the quantile-quantile (Q-Q) plots for the estimates from the case of $\alpha = 0.1$, $\beta = 0.8$ and $p = 0.8$ as an example; see Fig. 2. The Q-Q plots imply

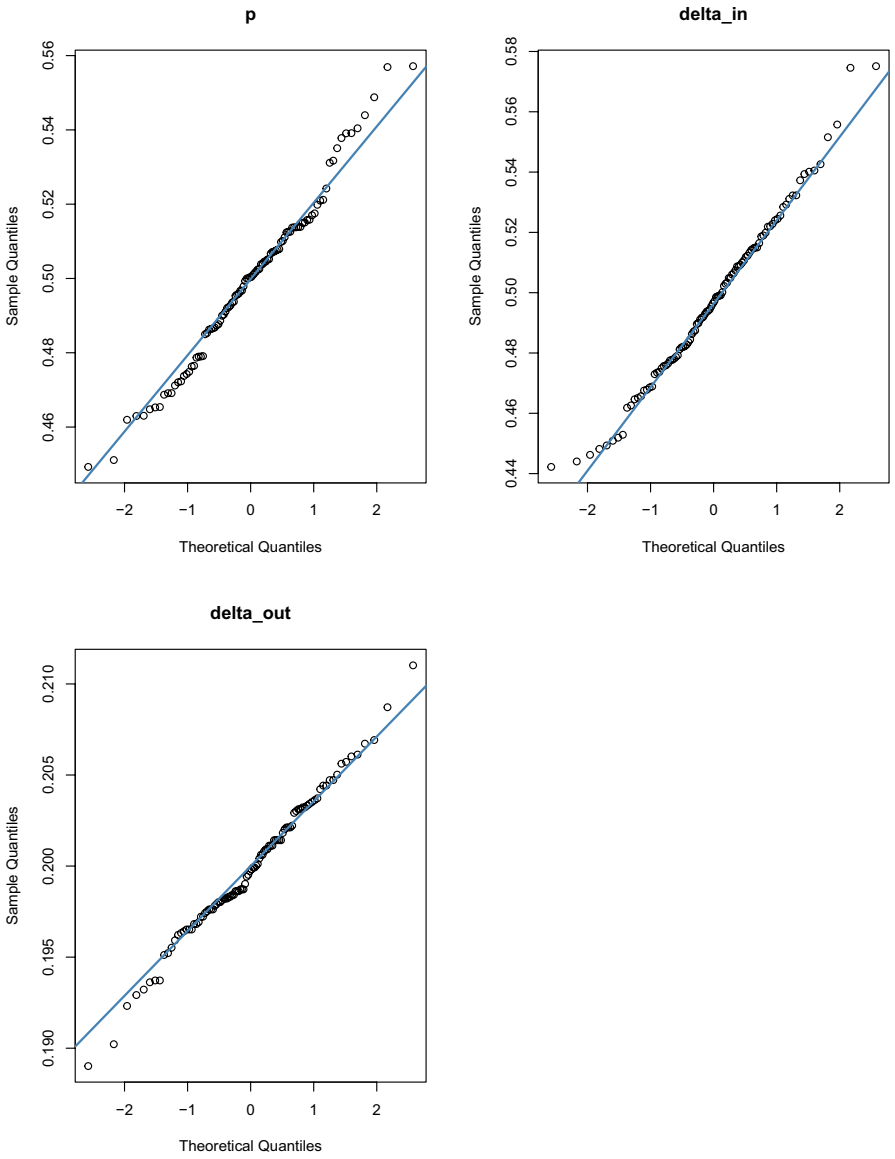

Fig. 2 Q-Q plots for the estimates based on 100 independently simulated HRNs with $\alpha = 0.1$, $\beta = 0.8$, $p = 0.8$, $\delta_{\text{in}} = 1.3$ and $\delta_{\text{out}} = 0.7$

that each of the estimates seems to follow a normal distribution marginally. We then run analogous analyses on the simulated HRNs under different parameter settings and obtain the same pattern. So, those Q-Q plots are not repeatedly presented.

Additionally, the N-M algorithm appears to be more efficient with average computation time (over valid simulation runs) 2.2 seconds versus 340.2 seconds for the

M-H algorithm. Albeit the better overall performance of the N-M method, the algorithm, as mentioned, undergoes the limitation of sensitivity to the initial value. Under the current setting, the N-M algorithm is able to provide estimation results over 90% of the simulation run with a fixed initial start given by $\alpha^{(0)} = \beta^{(0)} = \gamma^{(0)} = 1/3$, $p^{(0)} = 1/2$ and $\delta_{\text{in}}^{(0)} = \delta_{\text{out}}^{(0)} = 1$, where all the initial values are relatively close to the true values. When we use a random initial start (e.g., spacing α, β and γ randomly on the unit interval, sampling p from a standard uniform distribution and sampling δ_{in} and δ_{out} independently from a standard exponential distribution), the success rates (for all the scenarios) drop significantly to 60% or less. The failure of the algorithm is primarily due to the inaccurate start of δ_{in} and δ_{out}. We have also run some experiments on smaller networks. When reducing the simulated network size to 5,000, the success rate of the N-M algorithm declines to 35% or less even with a fixed initial simplex.

In contrast, the M-H algorithm is more robust, as it is always able to produce estimation results regardless of the size of the network. Specifically, we consider a random spacing of α, β and γ on the unit interval, sample δ_{in} and δ_{out} independently from a standard exponential distribution, and let p start from a value close to 1 (e.g., $1 - 10^{-4}$), which indicates an almost perfect linear PA. Simulation results show that the choice of a non-informative prior of p (i.e., sampling it uniformly from [0, 1]) has negligible impact on the final estimation results. Though we observe bias in the estimates of δ_{in} and δ_{out}, the 100% success rate renders the M-H method a competitive alternative.

To summarize, we recommend the N-M algorithm for parameter estimation if there is auxiliary information available to decide reasonable initial values. In addition, the M-H algorithm is a possible backup when the N-M algorithm fails. In practice, we may consider an integration of the two algorithms. Although the estimates of δ_{in} and δ_{out} from the M-H algorithm may not be very accurate, they are close enough to true values allowing for a successful implementation of the N-M algorithm. Therefore, we may first adopt the M-H algorithm to get coarse estimates of the model parameters and use them as initial values for the N-M algorithm, which ultimately leads to finer estimation results. It is worth mentioning that this initial value selection procedure does not actually affect the estimation results by the N-M algorithm, but effectively increases the probability of the successful implementation of the N-M algorithm. If the target parameter is p only but not δ_{in} or δ_{out}, the estimation results from the M-H algorithm may have been acceptable.

5.1 Model identifiability

It is worth highlighting that the proposed mixture model assumes $p \in (0, 1)$. Whenever $p = 1$, the proposed model coincides with the standard linear PA model. Based on the theoretical results given in Theorem 2, we notice that the limiting out- and in-degree distributions from an HRN with parameters $(\alpha, \beta, \gamma, p, \delta_{\text{in}}, \delta_{\text{out}})$ agrees with those from a standard PA model with parameters $(\alpha, \beta, \gamma, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}})$. This raises potential concerns on the issue of non-identifiability for our proposed model.

To verify, we generate 100 independent HRN replicates of size 10^4 with parameters given by $(\alpha, \beta, \gamma, p, \delta_{\text{in}}, \delta_{\text{out}}) = (0.1, 0.8, 0.1, 0.5, 1, 2)$. Pertaining to the relations developed in Theorem 2, we have $(\tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}}) = (7, 9)$. In Table 4, we present the estimated results based on the proposed algorithm, and see that the estimates of p, δ_{in} and δ_{out} are all close to the true parameters for simulated HRNs, rather than $(p, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}}) = (1, 7, 9)$. For comparison, we also use $(\alpha, \beta, \gamma, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}}) = (0.1, 0.8, 0.1, 7, 9)$ to generate 100 independent linear PA networks. Our numerical results show that if we do not impose the criterion $p = 1$, i.e. we do not have strong belief on a pure PA model, the estimated parameters are close to those from the HRN counterparts. Therefore, the condition of $p < 1$ is crucial to avoid the non-identifiability issue of the proposed model.

6 Real data analysis

In this section, we fit the proposed HRN model to two real network datasets: the Dutch Wikipedia talk network and the Facebook wall posts, both of which are retrieved from the KONECT network data repository (http://konect.cc/). By investigating the timestamp information from both datasets, we see the existence of two additional edge creation scenarios at each step:

1. Set $J_n = 4$ (with probability $0 < \xi < 1$) if a new node with a self-loop is added to the network;
2. Set $J_n = 5$ (with probability $0 < \eta < 1$) if two new nodes with a directed edge connecting them are added to the network.

Note that these two additional scenarios require minor modifications of the log-likelihood given in Sect. 4, but they do not impose direct effect on the score functions of p, δ_{in} and δ_{out}. The MLEs of ξ and η are straightforward:

$$\hat{\xi}^{MLE} = \frac{1}{n}\sum_{k=1}^{n} \mathbb{I}_{\{J_k=4\}} \qquad \text{and} \qquad \hat{\eta}^{MLE} = \frac{1}{n}\sum_{k=1}^{n} \mathbb{I}_{\{J_k=5\}}.$$

The assessment of the goodness-of-fit of the proposed HRN model is based on simulations. This is a valid approach which has been theoretically validated and widely used in the community (Hunter et al. 2008).

Table 4 Estimates of p, δ_{in} and δ_{out} for HRNs given $(\alpha, \beta, \gamma) = (0.1, 0.8, 0.1)$

Model	$\hat{\alpha}$	$\hat{\beta}$	$\hat{p}$	$\hat{\delta}_{\text{in}}$	$\hat{\delta}_{\text{out}}$
Est.	0.0997	0.8002	0.4802	0.7905	1.7385
SE	0.0003	0.0004	0.0040	0.0462	0.0579

6.1 Dutch wikipedia talk

In the Dutch Wikipedia talk dataset, every single node represents a specific user of the Dutch Wikipedia, and the creation of a directed edge from node u to node v refers to the event that user u leaves a message on user v's talk page. The dataset includes the communication among the users of Dutch Wikipedia talk pages from 10/18/2002 and 11/23/2015, consisting of three columns. The first two columns represent users' ID and the third column gives a UNIX timestamp with the time of a message posted on one's Wikipedia talk page. For each row, the first user writes a message on the talk page of the second user at a timestamp given in the third column.

According to findings in Wang and Resnick (2021), this network does not enter the stable phase of network growth until year 2008. Then for illustration purposes, we choose a sub-network according to the timestamp information from 01/01/2013 to 03/31/2013. The sub-network is directed, consisting of 3288 nodes and 21,724 directed edges. We fit the proposed hybrid model to the data by using the integration of the M-H algorithm and the N-M algorithm. For the M-H algorithm, the burn-in number and iteration sample size are both set at 100,000, and the gap for thinning sampling is 500. The convergence of the estimates is checked via the time-series plots in Fig. 3. Using the estimates from the M-H algorithm as the initial values for the N-M algorithm, we get almost identical estimates given by

$$\widehat{\theta}_{\mathrm{WK}} := (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\xi}, \hat{p}, \hat{\delta}_{\mathrm{in}}, \hat{\delta}_{\mathrm{out}}) = (0.014, 0.745, 0.227, 0.006, 0.999, 0.186, 0.149),$$

where the value of $\hat{p}$ is extremely close to 1. The large $\hat{p}$ suggests PA dominates the evolutionary process, thus leading to little difference between the proposed HRN and that proposed in Wan et al. (2017). For comparison, we compute the MLEs from the pure PA network model to get

$$\widehat{\theta}^{(PA)}_{\mathrm{WK}} := (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\xi}, \hat{\delta}_{\mathrm{in}}, \hat{\delta}_{\mathrm{out}}) = (0.012, 0.784, 0.189, 0.006, 0.156, 0.152),$$

from which we see the estimates from the two models are close to each other.

We generate 50 independent replications of the HRNs and pure PA networks with their respective estimates. The out-degree and in-degree tail distributions of these simulated networks are presented in Fig. 4, where the top two panels are for HRNs and the bottom two are for pure PA networks. Since the estimate $\hat{p}$ for HRN is extremely close to 1, the empirical out-degree and in-degree tail distributions from the two classes of simulated networks and their coverage region shapes are alike as expected. The left two-panel show that the out-degree tail distribution of the selected sub-network is well covered by both overlaid plots. However, in the right two panels, we observe negligible discrepancies between the overlaid plots and the in-degree distribution of the real data. Overall, our analysis suggests the evolution of the selected sub-network of the Dutch Wikipedia talk data follows a linear PA mechanism, so it is well fit by both HRN and pure PA model (Wan et al. 2017). Thus, the proposed hybrid model is flexible to fit the real network data only presenting linear PA mechanism.

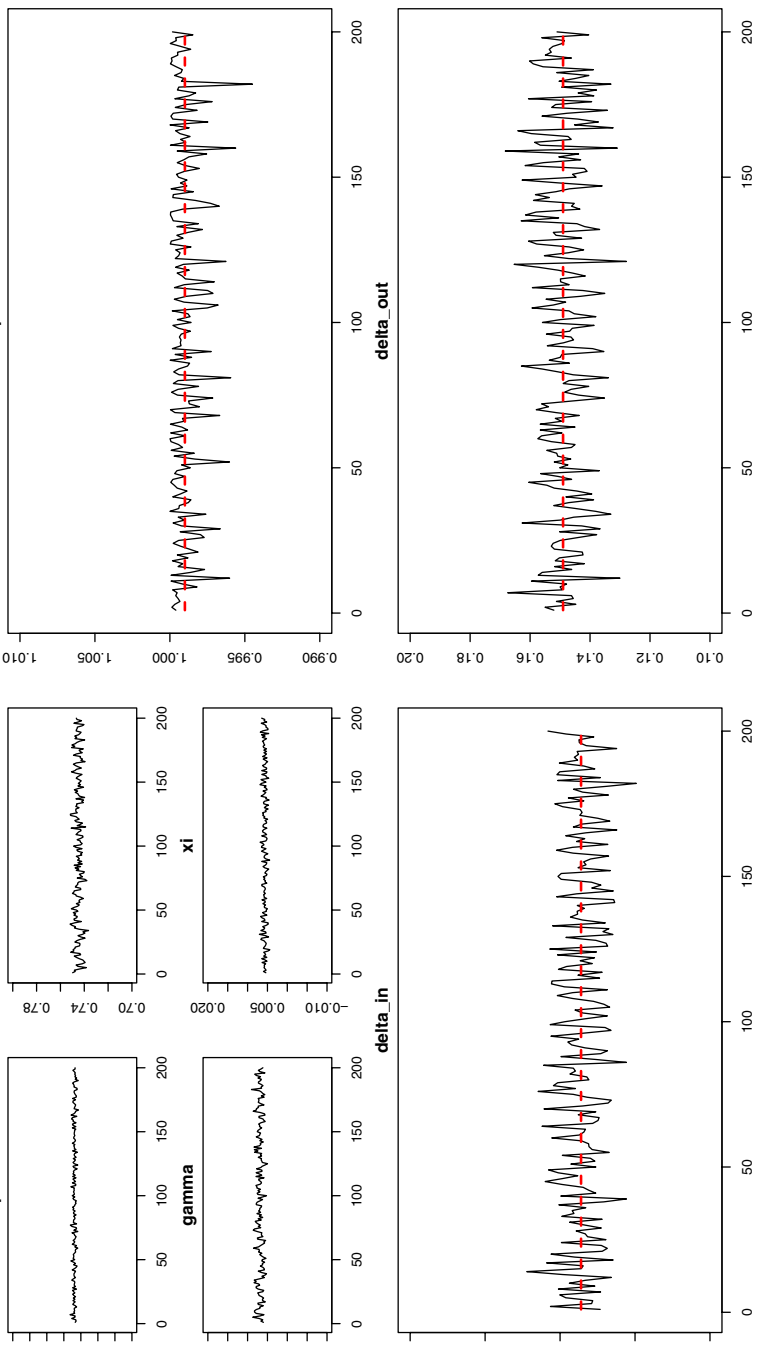

Fig. 3 Graphical diagnostics of the convergence for the M-H algorithm applied to the sub-network of Dutch Wikipedia talk dataset; red dashed lines represent the estimated values

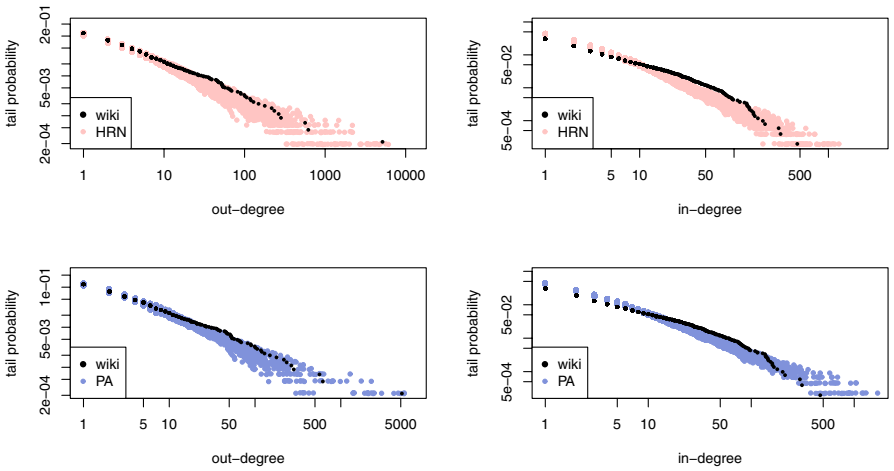

Fig. 4 Top two panels: Out-degree and in-degree tail distributions of 50 independent HRNs (red), compared with those from the Wikipedia data (black); Bottom two panels: Out-degree and in-degree tail distributions of 50 independent pure PA networks (blue), compared with those from the Wikipedia data (black)

6.2 Facebook wall posts

The Facebook wall post dataset collects data from a regional network of users in New Orleans from 09/13/2004 to 01/21/2009. The data forms a directed graph where the nodes are Facebook users and each directed edge represents a post from one node to another node's page. Like the Dutch Wikipedia Talk example, the Facebook dataset also contains three columns, where the first two contain the identifiers of the individual users, while the third records the timestamp of the corresponding post.

We select a sub-network of the Facebook wall post based on timestamps from 01/01/2006 to 06/30/2006. The sub-network consists of 4,200 nodes and 11,422 directed edges. Fitting the proposed hybrid model to the data, we get

$$\widehat{\theta}_{\mathrm{FB}} := (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\xi}, \hat{\rho}, \hat{\delta}_{\mathrm{in}}, \hat{\delta}_{\mathrm{out}}) = (0.071, 0.714, 0.114, 0.077, 0.830, 0.172, 0.007).$$

The estimates are obtained via the integration of the N-M and M-H algorithms, where the settings of burn-in number, iteration size and thinning gap are identical to the previous example. In Fig. 5, we verify the convergence of the estimates based on the M-H algorithm. Once again, we do not observe significant difference between the corresponding estimates from the two algorithms. For the purpose of comparison, we also fit the data with pure PA model to get

$$\widehat{\theta}^{(PA)}_{\mathrm{FB}} := (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\xi}, \hat{\delta}_{\mathrm{in}}, \hat{\delta}_{\mathrm{out}}) = (0.073, 0.711, 0.115, 0.078, 0.604, 0.341).$$

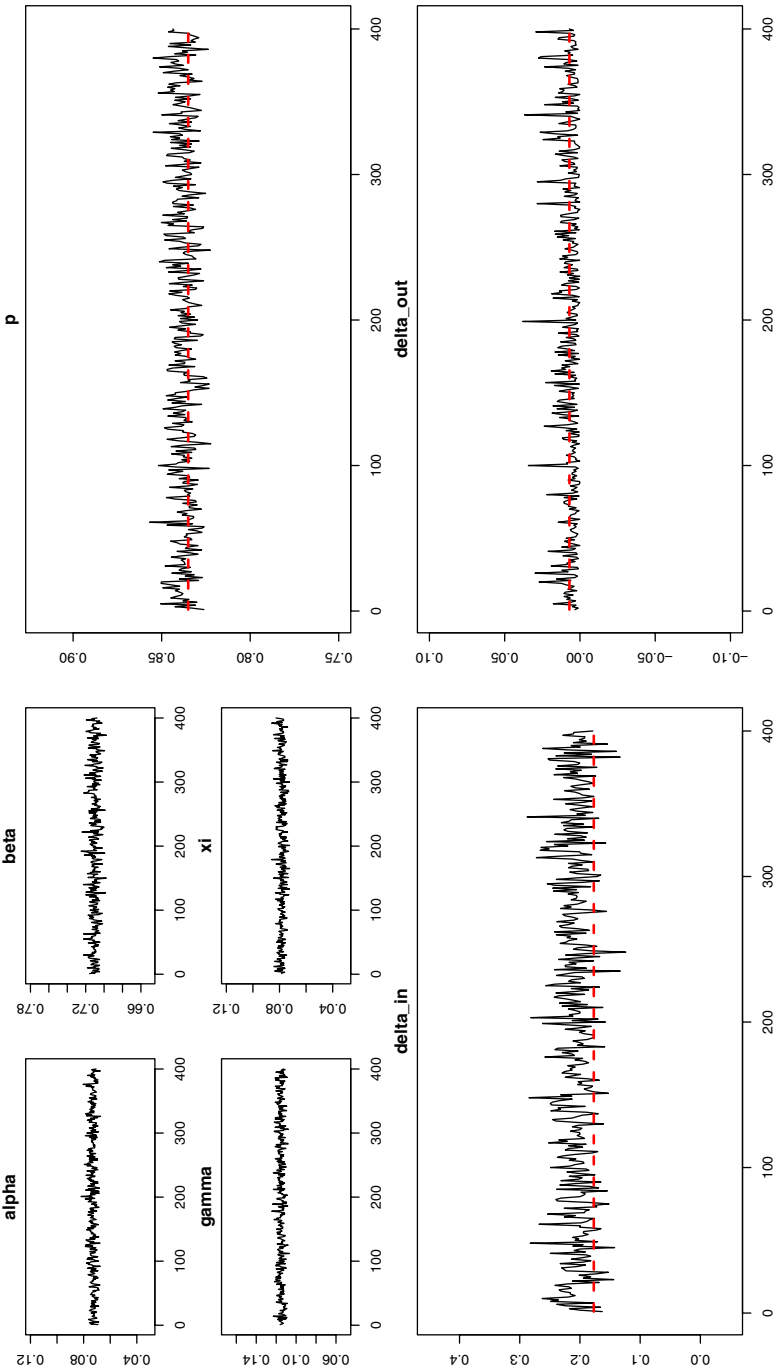

Fig. 5 Graphical diagnostics of the convergence for the M-H algorithm applied to the sub-network of Facebook wall posts dataset; red dashed lines represent the estimated values

Analogous to the previous example, we assess the goodness-of-fit of the model through the tail distributions of in-degree and out-degree. Specifically, we generate 50 independent HRNs and pure PA networks by using $\widehat{\theta}_{\text{FB}}$ and $\widehat{\theta}^{(PA)}_{\text{FB}}$, respectively. We then overlay the empirical out-degree and in-degree distributions of the two classes of simulated networks in Fig. 6. The graphical results show that the out-degree tail distribution is better captured by the HRN than the in-degree tail distribution, as it appears on the lower bound of the overlaid tail distributions of the simulated networks. The in-degree tail distribution of the Facebook sub-network is not well covered by the coverage region formed by the counterparts of the simulated networks, though the shapes look similar and the deviation is not large. Nonetheless, the goodness of fit of the proposed HRN is better than the pure PA model. The out-degree tail of the Facebook sub-network obviously deviates away from the coverage region, and for in-degree tail, we observe much more discrepancy in pure PA model (than the proposed HRN).

The discrepancy in Fig. 6 may be due to the high reciprocity feature in the Facebook wall posts as well as the fact that the collected data is for users in New Orleans only. The wall posts activities among the Facebook users in a specific region tend to be reciprocated: when a friend posts a message on one's wall, he/she is likely to reply quickly. In fact, using the `reciprocity()` function in the `igraph` package (Csardi and Nepusz 2006), we see that the proportion of reciprocated edges in the sub-network is over 0.18. Indeed, the reciprocated wall posts are certainly not uniform, thus not very well characterized by the parameter p. To better study the reciprocity feature, we may consider other variants of the PA model, which are left for future work.

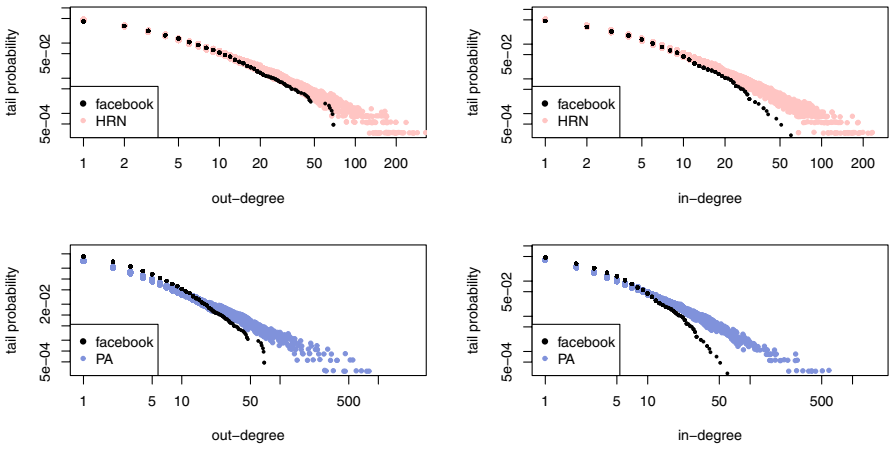

Fig. 6 Top two panels: Out-degree and in-degree tail distributions of 50 independent HRNs (red), compared with those from the Facebook data (black); Bottom two panels: Out-degree and in-degree tail distributions of 50 independent pure PA networks (blue), compared with those from the Facebook data (black)

6.3 Multiple time window fit

Recently in Wang and Resnick (2021), the detailed timestamp information for these two datasets has been carefully studied, and the actual network evolution may have different phases of growth. Therefore, we now divide the network evolution into small sub-networks and fit the proposed model according to each sub-network. A similar study has been done in (Wan et al. 2017, Sect. 6) to obtain a reasonable fit for the pure PA model in the Dutch Wikipedia Talk data. Since we have $\hat{p} \approx 1$ in Sect. 6.1, here we will only consider applying the multiple-window approach to the Facebook data.

By Wang and Resnick (2021), the Facebook wall post network does not enter the stable phase of network growth until May 2006, suggesting that this particular network may reach its local stability in different time windows. To characterize the evolutionary feature of the entire dataset, we use the proposed HRN model to fit the sub-networks from different time windows. More specifically, we split the whole network into 87 sub-networks, each of which contains 10^4 edges (where the last sub-network is slightly smaller than the rest in size). By applying the HRN model, we get the estimates of α, β, γ, ξ, p, δ_{in} and δ_{out} for each sub-network. The estimation results for p, δ_{in} and δ_{out} are shown in Fig. 7, whereas the estimates of the rest remain almost identical over time.

The top panel in Fig. 7 suggests an overall decreasing trend in the estimates of p. Especially after entering 2009, the estimate of p stays lower than 0.7, rendering that fitting the data via the proposed HRN model is significantly different from by fitting a pure PA model. As remarked in (Wang and Resnick 2021, Sect. 3.1) and Viswanath et al. (2009), Facebook's new site design was unveiled on July 20, 2008, allowing users to directly view wall posts through friends' feeds, and we speculate this new design as one possible reason to increase the part of uniform attachment in the fitted network.

In addition, the downward trend also shows that it may be inappropriate to fit the entire network data via one model, regardless of HRN model or pure PA model. In contrast, the estimates of δ_{in} and δ_{out} are relatively stable overtime in spite of ups and downs. Therefore, fitting massive scale networks, especially those growing with multiple phases, by fitting locally stable sub-networks seems a promising approach, which deserves further investigation in our future study.

7 Discussions

In this paper, we propose a class of hybrid model simultaneously presenting the preferential attachment (PA) and uniform attachment (UA) mechanisms, which are governed by a tuning parameter $p \in (0, 1)$. We would like to point out that the degree distributions of the proposed model are asymptotically equivalent to those of the linear PA model under the setup of $\tilde{\delta}_{\text{in}}$ and $\tilde{\delta}_{\text{out}}$ given in Theorem 2. Therefore, the value of p is assumed to be strictly less than 1 to avoid the issue of identifiability. Two standard methods, the Nelder-Mead (N-M) algorithm and the Metropolis-Hastings (M-H) algorithm, are adopted for parameter estimation. Through extensive

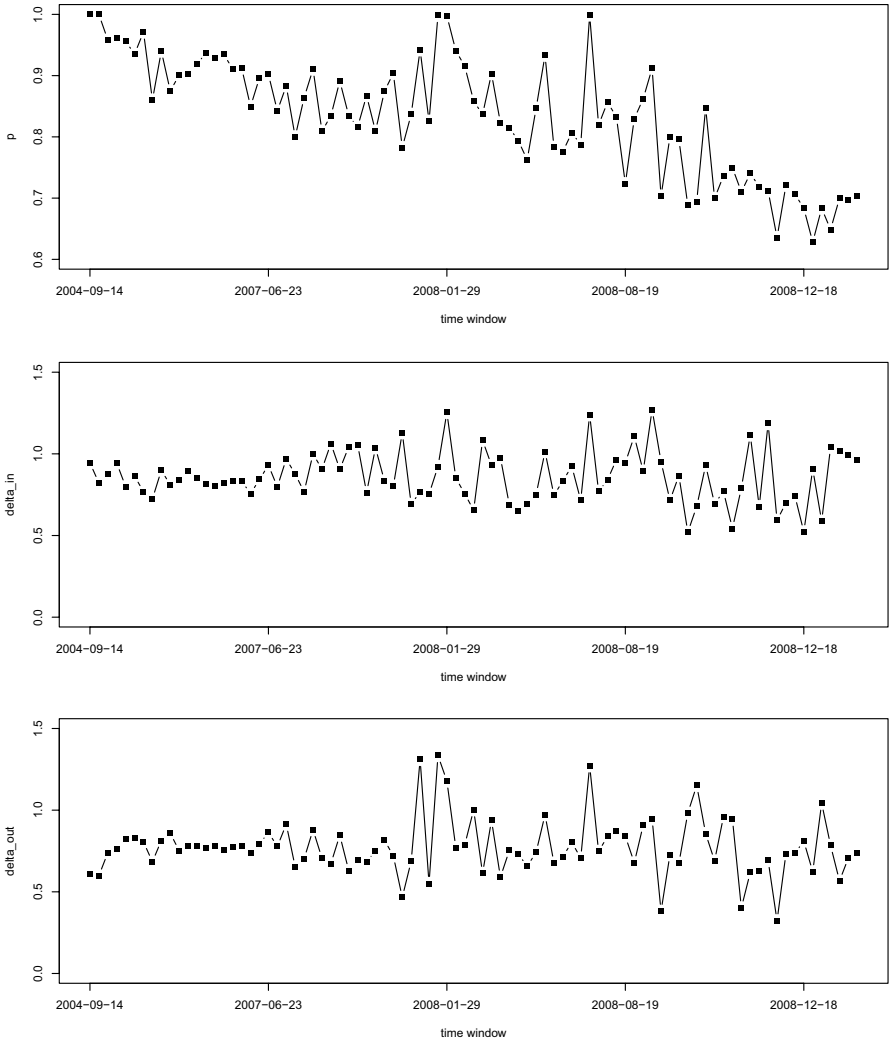

Fig. 7 Estimates of p, δ_{in} and δ_{out} over time windows

simulations and a sensitivity study, we find that the N-M algorithm is preferred, but the corresponding success rate of producing estimation result depends heavily on the selection of the initial simplex. We thus consider an integrated approach where we use the more robust M-H algorithm to get the initial values for the target parameters, followed by the implementation of the N-M algorithm.

In addition, we fit the HRN model to two real network datasets: the Dutch Wikipedia talk and Facebook wall posts, where we see that the proposed hybrid model provides a more flexible modeling framework compared with the directed PA network model as in Wan et al. (2017). The extra tuning parameter p helps correct the tail distributions of out-and in-degrees.

From the Facebook example, it is worth noting that even though the heaviness of the tail distributions (for both out- and in-degrees) has been weakened by p in the HRN, the proposed UA part is not able to fully capture the reciprocity property in the real network. We here provide three classes of possible remedies: (1) It may be worthwhile to try some algorithmic approaches, such as network rewiring, to wash off the nodes of large in-degree or out-degree in the simulated networks; (2) We may extend the model allowing for multi-edge addition at each timestamp; (3) We may consider modifying the model directly by introducing another parameter measuring the rate of reciprocation; (4) We may consider a more realistic mixer (some suitable light-tailed distribution) rather than simple UA in the present hybrid setting. We will report our research outcomes elsewhere in the future.

A. Proof of Theorem 2

Analogous to the previous proofs, we present the major steps of the proof for in-degree. To show the convergence of $\frac{N_m^{\text{in}}(n)}{n}$, we take two steps. The first is to prove the concentration of $\frac{N_m^{\text{in}}(n)}{n}$ around $\mathbb{E}\big(N_m^{\text{in}}(n)\big)/n$, then it suffices to find the asymptotic limit of $\mathbb{E}\big(N_m^{\text{in}}(n)\big)/n$.

Note that when $\beta = 0$, then the number of nodes in graph $G(n)$ is deterministic, so the concentration results in van der Hofstad (2017), Proposition 8.4 are applicable, and we have for $C > 2\sqrt{2}$,

$$\mathbb{P}\left(\left|N_m^{\text{in}}(n) - \mathbb{E}(N_m^{\text{in}}(n))\right| \geq C\sqrt{n\log n}\right) = o(1/n).$$

When $\beta > 0$, the total number of nodes in graph $G(n)$ is random, and detailed proofs are needed. We claim that for $\beta > 0$, there exists some constant $C > 2\sqrt{2}$ such that.

$$\mathbb{P}\left(\left|N_m^{\text{in}}(n) - \mathbb{E}(N_m^{\text{in}}(n))\right| \geq C\sqrt{n\log n}(1+\log n)\right) = o(1/n). \tag{12}$$

The proof of (12) relies on rewriting $N_m^{\text{in}}(n) - \mathbb{E}(N_m^{\text{in}}(n))$ in terms of a Doob's martingale, similar to the argument in the corrected version of Deijfen et al. (2009) (available at https://arxiv.org/pdf/0705.4151.pdf, and cited as Deijfen et al. (2020). But here since the number of nodes created at each step is random, we need to modify the proof machinery outlined in Deijfen et al. (2020). Recall the notation in Sect. 2 that $\{J_n : n \geq 1\}$ is a sequence of iid tri-nomial random variable on $\{1, 2, 3\}$ with cell probability α, β and γ, respectively. Write $\{J_k : 1 \leq k \leq n\} =: J_{[n]}$, and for $1 \leq t \leq n$, define

$$Z_t := \mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_t, J_{[t]}\right],$$

and $Z_0 = \mathbb{E}\big[N_m^{\text{in}}(n)\big]$. Then

$$N_m^{\text{in}}(n) - \mathbb{E}(N_m^{\text{in}}(n)) = Z_n - Z_0,$$

and $\{Z_n : n \geq 0\}$ is a martingale with $\mathbb{E}(|Z_t|) = \mathbb{E}(N_m^{\text{in}}(n)) \leq n$.

Then consider

$$\begin{aligned} Z_t - Z_{t-1} &= \mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_t, J_{[t]}\right] - \mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[t-1]}\right] \\ &= \mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_t, J_{[t]}\right] - \mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[t]}\right] \\ &\quad + \mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[t]}\right] - \mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[t-1]}\right] \\ &=: I + II. \end{aligned}$$

For I, the only change in the conditioning is the extra information contained in $G(t)$, which, compared with that in $G(t-1)$, specifies how the edge created at the t-th -step is constructed. This has the potential to affect the in-degrees of at most 2 nodes, thus leading to $|I| \leq 2$.

For the second term, II, we define $\bar{J}_t$ to be an independent copy of J_t, which is also independent from $J_{[n]}$. Write $\bar{J}_{[n]} := \{J_1, \ldots, J_{t-1}, \bar{J}_t, J_{t+1}, \ldots, J_n\}$. Let $\bar{N}_m^{\text{in}}(n)$ and $\bar{D}_v^{\text{in}}(n)$ be the number of nodes with in-degree m, and the in-degree of node v in the hybrid PA graph, $\bar{G}(n) = (\bar{V}_n, \bar{E}_n)$, constructed from $\bar{J}_{[n]}$, respectively. Then we have

$$\begin{aligned} II &= \mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[t]}\right] - \mathbb{E}\left[\bar{N}_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[t-1]}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}\right] \,\middle|\, \mathcal{F}_{t-1}, J_{[t]}\right] - \mathbb{E}\left[\bar{N}_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[t]}\right] \\ &= \mathbb{E}\left\{\mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}\right] - \mathbb{E}\left[\bar{N}_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, \bar{J}_{[n]}\right] \,\middle|\, \mathcal{F}_{t-1}, J_{[t]}\right\} \\ &= \mathbb{E}\left\{\mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right] - \mathbb{E}\left[\bar{N}_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right] \,\middle|\, \mathcal{F}_{t-1}, J_{[t]}\right\}. \end{aligned}$$

Therefore, it suffices to consider

$$\begin{aligned} &\left|\mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right] - \mathbb{E}\left[\bar{N}_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right]\right| \\ &\leq \left|\sum_{v=1}^{|V_n|} \mathbb{P}\left[D_v^{\text{in}}(n) = m \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right] - \sum_{v=1}^{|\bar{V}_n|} \mathbb{P}\left[\bar{D}_v^{\text{in}}(n) = m \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right]\right| \end{aligned} \tag{13}$$

where potential differences will occur only if $J_n \neq \bar{J}_n$.

We start by assuming that $J_n, \bar{J}_n \in \{1, 3\}$, i.e. the total numbers of nodes in the two graphs remain unchanged. Then the quantity in (13) is bounded above by

$$\begin{aligned} &\sum_{v=1}^{|V_n|} \left|\mathbb{P}\left[D_v^{\text{in}}(n) = m \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right] - \mathbb{P}\left[\bar{D}_v^{\text{in}}(n) = m \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right]\right| \\ &\quad \leq \sum_{v=1}^{|V_n|} \mathbb{E}\left[\mathbf{1}_{\{D_v^{\text{in}}(n) \neq \bar{D}_v^{\text{in}}(n)\}} \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right]. \end{aligned}$$

When we either have $J_t = 1, \bar{J}_t = 3$ or $J_t = 3, \bar{J}_t = 1$, there are at most 2 nodes whose in-degrees will be different. Therefore, when $J_t, \bar{J}_t \in \{1, 3\}$,

$$\left| \mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]} \right] - \mathbb{E}\left[\bar{N}_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]} \right] \right| \leq 2.$$

If $(J_t, \bar{J}_t) \in \{(2,1), (2,3), (1,2), (3,2)\}$, then $\left| |V_s| - |\bar{V}_s| \right| = 1$, for all $s \geq n$, and we need to consider nodes created before and after step t separately. In particular, the difference in the total number of nodes will also lead to different attachment probabilities in the two graphs. Without loss of generality, we assume $J_t = 2$ and $\bar{J}_t \neq 2$. For comparison purpose, we will relabel the extra node added at the t-th step as t', and keep the labeling of the other nodes identical in the two graphs. Then the quantity in (13) is bounded above by

$$1 + \sum_{v=1}^{|V_n|} \left| \mathbb{P}\left[D_v^{\text{in}}(n) = m \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]} \right] - \mathbb{P}\left[\bar{D}_v^{\text{in}}(n) = m \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]} \right] \right|.$$

Let N be the first time after t that a new node is created, i.e. $N := \inf\{k \geq t+1 : J_k \neq 2\}$. Note that for $s \in \{t+1, \dots N\}$, every edge that is added at step s and pointing to the node t' will lead to a potential difference in the in-degree of nodes in V_{t-1}. Hence, apart from node t', there are at most $N - t - 1$ number of nodes in V_{t-1} having different in-degrees in the two graphs. If no edge between step $t+1$ and step N has been pointing to the node t', then possible differences in the in-degree of one particular node may occur due to the change in the attachment probabilities. This is also the case for those nodes added at N and afterward. To deal with different in-degrees due to changes in the attachment probabilities, we will apply a similar treatment as given in (Deijfen et al. 2020, Eq. (2.17)).

We now rewrite

$$\begin{aligned}
&\mathbb{P}\left[D_v^{\text{in}}(n) = m \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]} \right] - \mathbb{P}\left[\bar{D}_v^{\text{in}}(n) = m \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]} \right] \\
&= \mathbb{P}\left[D_v^{\text{in}}(n) = m, \bar{D}_v^{\text{in}}(n) \neq m \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]} \right] \\
&\quad - \mathbb{P}\left[D_v^{\text{in}}(n) \neq m, \bar{D}_v^{\text{in}}(n) = m \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]} \right].
\end{aligned}$$

Therefore, at least one of the attachments to node $v \in V_n$ needs to have been made for one of the graphs but not the other. Let s denote the first time where such an attachment was made differently in the two graphs. Then we have $D_v^{\text{in}}(s-1) = \bar{D}_v^{\text{in}}(s-1) \leq m$. Hence,

$$
\begin{aligned}
&\mathbb{P}\left[D_v^{\text{in}}(n) = m, \bar{D}_v^{\text{in}}(n) \neq m \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right] \\
&\leq \sum_{s=t+1}^{n} \mathbb{E}\left[p(D_v^{\text{in}}(s-1)+\delta_{\text{in}})\left|\frac{1}{s+\delta_{\text{in}}|V_{s-1}|} - \frac{1}{s+\delta_{\text{in}}|\bar{V}_{s-1}|}\right| \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right] \\
&\quad + \sum_{s=t+1}^{n}(1-p)\left|\frac{1}{|V_{s-1}|} - \frac{1}{|\bar{V}_{s-1}|}\right| \\
&= \sum_{s=t+1}^{n} \mathbb{E}\left[\frac{p\delta_{\text{in}}(D_v^{\text{in}}(s-1)+\delta_{\text{in}})}{(s+\delta_{\text{in}}|V_{s-1}|)(s+\delta_{\text{in}}|\bar{V}_{s-1}|)} \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right] \\
&\quad + \sum_{s=t+1}^{n}(1-p)\frac{1}{|V_{s-1}||\bar{V}_{s-1}|}.
\end{aligned}
\tag{14}
$$

Since $\sum_{v\in V_{s-1}}(D_v^{\text{in}}(s-1)+\delta_{\text{in}}) = s+\delta_{\text{in}}|V_{s-1}|$, then (14) implies that

$$
\begin{aligned}
&\sum_{v} \mathbb{P}\left[D_v^{\text{in}}(n) = m, \bar{D}_v^{\text{in}}(n) \neq m \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right] \\
&\leq \sum_{s=t+1}^{n}\sum_{v\in V_{s-1}}\left(\mathbb{E}\left[\frac{p\delta_{\text{in}}(D_v^{\text{in}}(s-1)+\delta_{\text{in}})}{(s+\delta_{\text{in}}|V_{s-1}|)(s+\delta_{\text{in}}|\bar{V}_{s-1}|)} \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right]\right. \\
&\qquad \left. + \sum_{s=t+1}^{n}(1-p)\frac{1}{|V_{s-1}||\bar{V}_{s-1}|}\right) \\
&\leq \sum_{s=t+1}^{n}\left(\frac{p\delta_{\text{in}}}{s} + \frac{1-p}{|V_{s-1}|}\right).
\end{aligned}
$$

Thus, combining all scenarios together gives that

$$
\begin{aligned}
|II| &\leq \mathbb{E}\left[\left|\mathbb{E}\left[N_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right] - \mathbb{E}\left[\bar{N}_m^{\text{in}}(n) \,\middle|\, \mathcal{F}_{t-1}, J_{[n]}, \bar{J}_{[n]}\right]\right| \,\middle|\, \mathcal{F}_{t-1}, J_{[t]}\right] \\
&\leq 2\left(1 + \mathbb{E}\left[N - t - 1 + \sum_{s=t+1}^{n}\left(\frac{p\delta_{\text{in}}}{s} + \frac{1-p}{|V_{s-1}|}\right)\middle|\mathcal{F}_{t-1}, J_{[t]}\right]\right) \\
&= 2\left(\frac{1}{\beta} + \sum_{s=t+1}^{n}\frac{p\delta_{\text{in}}}{s} + \sum_{s=t+1}^{n}\mathbb{E}\left[\frac{1-p}{|V_{s-1}|}\middle|\mathcal{F}_{t-1}, J_{[t]}\right]\right).
\end{aligned}
$$

Applying the bound in Eq. (1.2) of the supplement gives that there exists some constant $C' > 0$ such that

$$
|II| \leq 2/\beta + C'\sum_{s=t+1}^{n} s^{-1} \leq C'\log(n/t),
$$

which further implies

$$|Z_t - Z_{t-1}| \le 1 + 2/\beta + C' \log(n/t).$$

Then by the Azuma-Hoeffding's inequality, we have

$$\mathbb{P}\Big(\Big|N_m^{\text{in}}(n) - \mathbb{E}(N_m^{\text{in}}(n))\Big| \ge b\Big) \le 2\exp\left\{-\frac{b^2}{8\sum_{t=1}^{n}(1+2/\beta + C'\log(n/t))^2}\right\} \le 2\exp\left\{-\frac{b^2}{8n(1+2/\beta+\log n)^2}\right\}.$$

Then the claim in (12) follows by setting $b = C\sqrt{n\log n}(1+2/\beta+\log n)$, with $C > 2\sqrt{2}$.

Then we are left with identifying the asymptotic limit of $\mathbb{E}(N_m^{\text{in}}(n))/n$. Consider the following approximation of the attachment probability:

$$\frac{p\big(D_i^{\text{in}}(n)+\delta_{\text{in}}\big)}{\big(1+\delta_{\text{in}}(1-\beta)\big)n} + \frac{1-p}{(1-\beta)n} = \frac{D_i^{\text{in}}(n)+\delta_{\text{in}}+\frac{1-p}{p(1-\beta)}\big(1+\delta_{\text{in}}(1-\beta)\big)}{\big(1+\delta_{\text{in}}(1-\beta)\big)n/p}.$$

Recall that

$$\tilde{\delta}_{\text{in}} = \delta_{\text{in}} + \frac{1-p}{p(1-\beta)}\big(1+\delta_{\text{in}}(1-\beta)\big) = \frac{\delta_{\text{in}}}{p} + \frac{1-p}{p(1-\beta)}.$$

Applying Chernoff bound again gives

$$\left|\mathbb{E}\left[\frac{p\big(D_i^{\text{in}}(n)+\delta_{\text{in}}\big)}{n+1+|V_n|\delta_{\text{in}}} + \frac{1-p}{|V_n|}\right] - \mathbb{E}\left[\frac{D_i^{\text{in}}(n)+\tilde{\delta}_{\text{in}}}{\big(1+\delta_{\text{in}}(1-\beta)\big)n/p}\right]\right| \le Cn^{-3/2}\sqrt{\log n}, \tag{15}$$

for some constant $C > 0$. Consider a in-degree sequence $\big\{\tilde{D}_i^{\text{in}}(n)\big\}$ from a directed PA network with set of parameters $(\alpha, \beta, \gamma, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}})$, as studied in Samorodnitsky et al. (2016); Wan et al. (2017). Establish an argument similar to Eq. (1.1) in the supplement as follows:

$$\left|\mathbb{E}\left[\frac{p\big(D_i^{\text{in}}(n)+\delta_{\text{in}}\big)}{n+1+|V_n|\delta_{\text{in}}} + \frac{1-p}{|V_n|}\right] - \mathbb{E}\left[\frac{\tilde{D}_i^{\text{in}}(n)-\tilde{\delta}_{\text{in}}}{n+1+|V_n|\tilde{\delta}_{\text{in}}}\right]\right| \le \tilde{C}n^{-3/2}\sqrt{\log n},$$

for some constant $\tilde{C} > 0$. Note

$$\mathbb{P}\big(D_i^{\text{in}}(n) = m\big) = \sum_{j=m-1}^{m} \mathbb{P}\big(D_i^{\text{in}}(n) = m \,|\, D_i^{\text{in}}(n-1) = j\big)\mathbb{P}\big(D_i^{\text{in}}(n-1) = j\big).$$

By the developed Chernoff bounds, we have

$$\mathbb{P}\big(D_i^{\text{in}}(n)=m\big) \le \mathbb{P}\big(\tilde{D}_i^{\text{in}}(n)=m\big) + (C+\tilde{C})\sum_{k=i}^{n} k^{-3/2}\sqrt{\log k}.$$

Noticing that $\sum_{k=i}^{n} k^{-3/2}\sqrt{\log k} < \infty$ as $n \to \infty$, we complete the proof by applying the results derived in Wang and Resnick (2020). □

B. Validation of MLE

From the log-likelihood function in (11), we have the following score functions for δ_{in}, δ_{out} and p, respectively.

$$\begin{aligned}\frac{\partial}{\partial \delta_{\text{in}}}\log L(\boldsymbol{\theta}\,|\,\boldsymbol{E}) = &\sum_{k=1}^{n}\frac{|V_{k-1}|}{\big(pD^{\text{in}}_{v_{k,2}}(k-1)+\delta_{\text{in}}\big)|V_{k-1}|+(1-p)k}\,\mathbb{I}_{\{J_k=\{1,2\}\}}\\ &-\sum_{k=1}^{n}\frac{|V_{k-1}|}{k+|V_{k-1}|\delta_{\text{in}}}\,\mathbb{I}_{\{J_k=\{1,2\}\}},\end{aligned} \tag{16}$$

$$\begin{aligned}\frac{\partial}{\partial \delta_{\text{out}}}\log L(\boldsymbol{\theta}\,|\,\boldsymbol{E}) = &\sum_{k=1}^{n}\frac{|V_{k-1}|}{\big(pD^{\text{out}}_{v_{k,1}}(k-1)+\delta_{\text{out}}\big)|V_{k-1}|+(1-p)k}\,\mathbb{I}_{\{J_k=\{2,3\}\}}\\ &-\sum_{k=1}^{n}\frac{|V_{k-1}|}{k+|V_{k-1}|\delta_{\text{out}}}\,\mathbb{I}_{\{J_k=\{2,3\}\}},\end{aligned} \tag{17}$$

$$\begin{aligned}\frac{\partial}{\partial p}\log L(\boldsymbol{\theta}\,|\,\boldsymbol{E}) = &\sum_{k=1}^{n}\frac{\big(D^{\text{in}}_{v_{k,2}}(k-1)|V_{k-1}|-k\big)\mathbb{I}_{\{J_k=\{1,2\}\}}}{\big(pD^{\text{in}}_{v_{k,2}}(k-1)+\delta_{\text{in}}\big)|V_{k-1}|+(1-p)k}\\ &+\sum_{k=1}^{n}\frac{\big(D^{\text{out}}_{v_{k,1}}(k-1)|V_{k-1}|-k\big)\mathbb{I}_{\{J_k=\{2,3\}\}}}{\big(pD^{\text{out}}_{v_{k,1}}(k-1)+\delta_{\text{out}}\big)|V_{k-1}|+(1-p)k}.\end{aligned} \tag{18}$$

We then set the score function (16) to 0. Note that due to the randomness of $|V_{k-1}|$, the methodology given in Wan et al. (2017) is not directly applicable. Instead, we approximate the score function (16) as follows:

$$\begin{aligned}&\sum_{k=1}^{n}\frac{|V_{k-1}|}{\big(pD^{\text{in}}_{v_{k,2}}(k-1)+\delta_{\text{in}}\big)|V_{k-1}|+(1-p)k}\,\mathbb{I}_{\{J_k=\{1,2\}\}}\\ &\quad=\sum_{k=1}^{n}\frac{\mathbb{I}_{\{J_k=\{1,2\}\}}}{\big(pD^{\text{in}}_{v_{k,2}}(k-1)+\delta_{\text{in}}\big)+(1-p)k/|V_{k-1}|}\\ &\quad=\frac{1}{p}\sum_{k=1}^{n}\frac{\mathbb{I}_{\{J_k=\{1,2\}\}}}{D^{\text{in}}_{v_{k,2}}(k-1)+\tilde{\delta}_{\text{in}}}+R_{\text{in}}(n),\end{aligned}$$

where

$$R_{\text{in}}(n) = \sum_{k=1}^{n} \frac{\mathbb{I}_{\{J_k=\{1,2\}\}}}{p} \left(\frac{1}{D^{\text{in}}_{v_{k,2}}(k-1) + \delta_{\text{in}}/p + (1-p)k/(p|V_{k-1}|)} - \frac{1}{D^{\text{in}}_{v_{k,2}}(k-1) + \tilde{\delta}_{\text{in}}} \right).$$

Therefore,

$$\begin{aligned} |R_{\text{in}}(n)| &\le \frac{1}{p} \sum_{k=1}^{n} \frac{(1-p)/p\left|k/|V_{k-1}| - 1/(1-\beta)\right| \mathbb{I}_{\{J_k=\{1,2\}\}}}{(D^{\text{in}}_{v_{k,2}}(k-1) + \delta_{\text{in}}/p + (1-p)k/|V_{k-1}|)(D^{\text{in}}_{v_{k,2}}(k-1) + \tilde{\delta}_{\text{in}})} \\ &\le \frac{1-p}{p^2} \sum_{k=1}^{n} \frac{|k/|V_{k-1}| - 1/(1-\beta)|}{(\delta_{\text{in}}/p + (1-p)k/|V_{k-1}|)\tilde{\delta}_{\text{in}}}. \end{aligned}$$

Since $|V_{n-1}|/n \xrightarrow{a.s.} 1/(1-\beta)$, then by the Cesàro convergence of random variables, we have $|R_{\text{in}}(n)|/n \xrightarrow{a.s.} 0$. Then the approximate score equation in (16) becomes

$$\frac{1}{n} \sum_{k=1}^{n} \frac{\mathbb{I}_{\{J_k=\{1,2\}\}}}{D^{\text{in}}_{v_{k,2}}(k-1) + \tilde{\delta}_{\text{in}}} = \frac{1}{n} \sum_{k=1}^{n} \frac{|V_{k-1}|}{k + |V_{k-1}|\delta_{\text{in}}} \mathbb{I}_{\{J_k=\{1,2\}\}}.$$

Applying the method in Wan et al. (2017) further yields the following approximate score function:

$$\sum_{m=0}^{\infty} \frac{N^{\text{in}}_{>m}(n)/n}{m + \tilde{\delta}_{\text{in}}} = \frac{\gamma}{\tilde{\delta}_{\text{in}}} + \frac{(\alpha+\beta)(1-\beta)}{1 + \delta_{\text{in}}(1-\beta)}, \tag{19}$$

where $N^{\text{in}}_{>m}(n)$ denotes the number of nodes with in-degree strictly greater than m in $\mathcal{H}_n$.

Similarly, the score equation with respect to (17) can be approximated by

$$\sum_{m=0}^{\infty} \frac{N^{\text{out}}_{>m}(n)/n}{m + \tilde{\delta}_{\text{out}}} = \frac{\alpha}{\tilde{\delta}_{\text{out}}} + \frac{(\beta+\gamma)(1-\beta)}{1 + \delta_{\text{out}}(1-\beta)}, \tag{20}$$

with $N^{\text{out}}_{>m}(n)$ being the number of nodes with out-degree strictly greater than m in $\mathcal{H}_n$. However, with (19) and (20) available, the approximation to the third score equation in (18) leads to a deterministic solution of $p = 1$. This indicates former methods to find MLE as in Wan et al. (2017) are not able to give us the desirable results.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-022-00827-5.

Acknowledgements We would like to thank two anonymous referees and the handling AE for constructive reports that help improve the quality of the paper.

References

Alves, C., Ribeiro, R., Sanchis, R. (2019). Preferential attachment random graphs with edge-step functions. *Journal of Theoretical Probability, 34*(1), 438–476.

Atalay, E., Hortaçsu, A., Roberts, J., Syverson, C. (2011). Network structure of production. *Proceedings of the National Academy of Sciences of the United States of America, 108*(13), 5199–5202.

Barabási, A.-L., Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512.

Chen, M.-H., Shao, Q.-M., Ibrahim, J. G. (2010). *Monte Carlo Methods in Bayesian Computation*. New York, NY: Springer-Verlag.

Cooper, C., Frieze, A. (2003). A general model of web graphs. *Random Structures and Algorithms, 22*(3), 311–335.

Csardi, G., Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695.

Deijfen, M., van den Esker, H., van der Hofstad, R., Hooghiemstra, G. (2009). A preferential attachment model with random initial degrees. *Arkiv för Matematik, 47*(1), 41–72.

Deijfen, M., van den Esker, H., van der Hofstad, R., Hooghiemstra, G. (2020). A preferential attachment model with random initial degrees. https://arxiv.org/pdf/0705.4151.pdf

de Sollar Price, D. J. (1965). Networks of scientific papers. *Science, 149*(3683), 510–515.

Durrett, R. T. (2006). *Random Graph Dynamics*. Cambridge, U.K.: Cambridge University Press.

Durrett, R. T. (2019). *Probability*: *Theory and Examples* (5 ed.). Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, U.K.: Cambridge University Press.

Gao, F., van der Vaart, A. (2017). On the asymptotic normality of estimating the affine preferential attachment network models with random initial degrees. *Stochastic Processes and their Applications, 127*(11), 3754–3775.

Gelman, A., Carlin, J. B., Dunson, D. B., Behtari, A., Rubin, D. B. (2013). *Bayesian Data Analysis*. Boca Raton, FL, U.S.A.: Chapman and Hall/CRC.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*(1), 97–109.

Henzinger, M., Lawrence, S. (2004). Extracting knowledge from the World Wide Web. *Proceedings of the National Academy of Sciences of the United States of America, 101*(supplement 1), 5186–5191.

Hunter, D. R., Goodreau, S. M., Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association, 103*(481), 248–258.

Lagarias, J. C., Reeds, J. A., Wright, M. H., Wright, P. E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal on Optimization, 9*(1), 112–147.

Liang, F., Liu, C., Carroll, R. J. (2010). *Advanced Markov Chain Monte Carlo Methods*: *Learning from Past Examples*. Hoboken, NJ, U.S.A.: Wiley.

Mahmoud, H. M. (2019). Local and global degree profiles of randomly grown self-similar hooking networks under uniform and preferential attachment. *Advances in Applied Mathematics, 111*, 101930.

Medina, J. A., Finke, J., Rocha, C. (2019). Estimating formation mechanisms and degree distributions in mixed attachment networks. *Journal of Physica A*: *Mathematical and Theoretical*, *52*, 095001.

Mengersen, K. L., Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Annals of Statistics, 24*(1), 101–121.

Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 56–63.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics, 21*, 1087.

Nash, J. C. (2014). On best practice optimization methods in R. *Journal of Statistical Software, 60*(2), 1–14.

Nelder, J. A., Mead, R. (1965). A simple method for function minimization. *The Computer Journal, 7*(4), 308–313.

Newman, M. E. J. (2001). Clustering and preferential attachment in growing networks. *Physical Review E, 65*(1), 025102.

Pachon, A., Sacerdote, L., Yang, S. (2018). Scale-free behavior of networks with the copresence of preferntial and uniform attachment rules. *Physica D*: *Nonliner Phenomena*, *371*, 1–12.

Samorodnitsky, G., Resnick, S., Towsley, D., Davis, R., Willis, A., Wan, P. (2016). Nonstandard regular variation of in-degree and out-degree in the preferential attachment model. *Journal of Applied Probability, 53*(1), 146–161.

Shao, Z.-G., Zou, X.-W., Jin, Z.-Z. (2006). Growing networks withmixed attachment mechanisms. *Journal of Physics A*: *Mathematical and General*, *39*, 9.

Smith, B. J. (2007). boa: An R package for MCMC output convergence assessment and posterior inference. *Journal of Sstatistical Software, 21*(11), 1–37.

van der Hofstad, R. (2017). *Random Graphs and Complex Networks*. Cambridge, U.K.: Cambridge University Press.

Viswanath, B., Mislove, A., Cha, M., Gummadi, K.P. (2009, August). On the evolution of user interaction in Facebook. In J. Crowcroft, & B. Krishnamurthy (Eds.), *Proceedings of the 2nd ACM Workshop on Online Social Networks (WOSN'09)*, New York, NY, U.S.A. (pp. 37–42). Association for Computing Machinery.

Wan, P., Wang, T., Davis, R. A., Resnick, S. I. (2017). Fitting the linear preferential attachment model. *Electronic Journal of Statistics, 11*(2), 3738–3780.

Wang, T., Resnick, S. (2018). Multivariate regular variation of discrete mass functions with applications to preferential attachment networks. *Methodology and Computing in Applied Probability, 20*(3), 1029–1042.

Wang, T., Resnick, S. (2020). Degree growth rates and index estimation in a directed preferential attachment model. *Stochastic Processes and their Applications, 130*(2), 878–906.

Wang, T., Resnick, S. I. (2015). Asymptotic normality of in- and out-degree counts in a preferential attachment model. *Stochastic Models, 33*(2), 229–255.

Wang, T., Resnick, S. I. (2020). A directed preferential attachment model with Poisson measurement. https://arxiv.org/pdf/2008.07005.pdf.

Wang, T., Resnick, S. I. (2021). Common growth patterns for regional social networks: A point process approach. *Journal of Data Science*. https://doi.org/10.6339/21-JDS1021.

Zhang, P., Mahmoud, H. M. (2020). On nodes of small degrees and degree profile in preferential dynamic attachment circuits. *Methodology and Computing in Applied Probability, 22*(2), 625–645.