# Fitting the linear preferential attachment model

## Phyllis Wan

*Department of Statistics*
*Columbia University*
*1255 Amsterdam Avenue, MC 4690*
*New York, NY 10027*
*e-mail:* phyllis@stat.columbia.edu

## Tiandong Wang

*School of Operations Research and Information Engineering*
*Cornell University*
*Ithaca, NY 14853*
*e-mail:* tw398@cornell.edu

## Richard A. Davis

*Department of Statistics*
*Columbia University*
*1255 Amsterdam Avenue, MC 4690*
*New York, NY 10027*
*e-mail:* rdavis@stat.columbia.edu

**and**

## Sidney I. Resnick

*School of Operations Research and Information Engineering*
*Cornell University*
*Ithaca, NY 14853*
*e-mail:* sir1@cornell.edu

**Abstract:** Preferential attachment is an appealing mechanism for modeling power-law behavior of the degree distributions in directed social networks. In this paper, we consider methods for fitting a 5-parameter linear preferential model to network data under two data scenarios. In the case where full history of the network formation is given, we derive the maximum likelihood estimator of the parameters and show that it is strongly consistent and asymptotically normal. In the case where only a single-time snapshot of the network is available, we propose an estimation method which combines method of moments with an approximation to the likelihood. The resulting estimator is also strongly consistent and performs quite well compared to the MLE estimator. We illustrate both estimation procedures through simulated data and explore the usage of this model in a real data example.

**Keywords and phrases:** Power laws, multivariate heavy tail statistics, preferential attachment, estimation.

## 1. Introduction

The preferential attachment mechanism, in which edges and nodes are added to the network based on probabilistic rules, provides an appealing description for the evolution of a network. The rule for how edges connect nodes depends on node degree; large-degree nodes attract more edges. The idea is applicable to both directed and undirected graphs and is often the basis for studying social networks, collaborator and citation networks, and recommender networks. Elementary descriptions of the preferential attachment model can be found in [5] while more mathematical treatments are available in [2, 4, 21]. Also see [10] for a statistical survey of methods for network data, [18] for consideration of statistics of an undirected network and [24] for asymptotics of a directed exponential random graph models. Limit theory for estimates of an undirected preferential attachment model was considered in [6].

For many networks, empirical evidence supports the hypothesis that in- and out-degree distributions follow a power law. This property has been shown to hold in linear preferential attachment models, which makes preferential attachment an attractive choice for network modeling [3, 4, 11, 12, 21]. While the marginal degree power laws in a simple linear preferential attachment model were established in [3, 11, 12], the joint regular variation (see [15, 16]) which is akin to a *joint power law*, was only recently established [17, 19]. In addition, it was shown in [22] that the joint probability mass function of the in- and out-degrees is multivariate regularly varying. This is a key result as the degrees of a network are integer-valued.

In this paper, we discuss methods of fitting a simple linear preferential attachment model, which is parametrized by $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \delta_{\mathrm{in}}, \delta_{\mathrm{out}})$. The first three parameters, $\alpha, \beta, \gamma$, correspond to probabilities of the 3 scenarios for adding an edge and hence sum to 1, i.e., $\alpha + \beta + \gamma = 1$. The other two, $\delta_{\mathrm{in}}$ and $\delta_{\mathrm{out}}$, are tuning parameters related to growth rates. The tail indices of the marginal power laws for the in- and out-degrees can be expressed as explicit functions of $\boldsymbol{\theta}$ (see (2.5) and (2.6) below). The graph $G(n) = (V(n), E(n))$, where $V(n)$ is the set of nodes and $E(n)$ is the set of edges at the $n$th iteration, evolves based on postulates that describe how new edges and nodes are formed. This construction of the network is Markov in the sense that the probabilistic rules for obtaining $G(n + 1)$ once $G(n)$ is known do not require prior knowledge of earlier stages of the construction.

The Markov structure of the model allows us to construct a likelihood function based on observing $G(n_0), G(n_0+1), \ldots, G(n_0+n)$. After deriving the likelihood function, we show that it has a unique maximum at $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}_{\mathrm{in}}, \hat{\delta}_{\mathrm{out}})$ and that the resulting maximum likelihood estimator is strongly consistent and asymptotically normal. The normality is proved using a martingale central limit theorem applied to the score function. The limiting distribution also reveals that $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$, $\hat{\delta}_{\mathrm{in}}$, and $\hat{\delta}_{\mathrm{out}}$ are asymptotically independent. From these results, asymptotic properties of the MLE for the power law indices can be derived.

For some network data, only a snapshot of the nodes and edges is available at a single point in time, that is, only $G(n)$ is available for some $n$. In such cases, we propose an estimation procedure for the parameters of the network using an approximation to the likelihood and method of moments. This also produces strongly consistent estimators. These estimators perform reasonably well compared to the MLE where the entire evolution of the network is known but predictably there is some loss of efficiency.

We illustrate the estimation procedure for both scenarios using simulated data. Simulation plays an important role in the process of modeling networks since it provides a way to assess the performance of model fitting procedures in the idealized setting of knowing the true model. Also, after fitting a model to real data, simulation provides a check on the quality of fit. Departures from model assumptions can often be detected via simulation of multiple realizations from the fitted network. Hence it is important to have efficient simulation algorithms for producing realizations of the preferential attachment network for a given set of parameter values. We adopt a simulation method, learned from Joyjit Roy, that was inspired by [1] and is similar to that of [20].

Our fitting methods are implemented in a real data setting using the Dutch Wiki talk network [14]. While one should not expect the simple 5-parameter (later extended to 7-parameter) linear preferential attachment model to fully explain a network with millions of edges, it does provide a reasonable fit to the tail behavior of the degree distributions. We are also able to detect important structural features in the network through fitting the model over separate time intervals.

Often it is difficult to believe in the existence of a true model, especially one whose parameters remain constant over time. Allowing, as we do, a preferential attachment model with only a few parameters and no possibility for node removal may seem simplistic and unrealistic for social network data. Of course, preferential attachment is only one mechanism for network formation and evidence for its use in fields outside data networks is mixed [8, 9] and we restrict attention to linear preferential attachment. Even imperfect models have the potential to capture salient properties in the data, such as heavy-tailedness of the in-degree and out-degree distributions, and to identify departures from model assumptions. While maximum likelihood estimation is essentially the gold standard for cases when the underlying model is a good representation of the data, it may perform poorly in case the model is far from being appropriate. In forthcoming work, we consider a semi-parametric estimation approach for network models that exhibit heavy-tailed degree distributions. This alternative estimation methodology borrows ideas from extreme value theory.
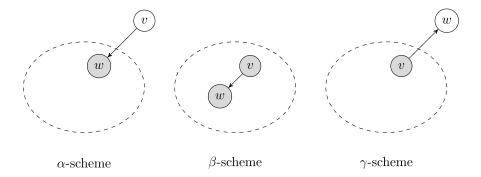
The rest of the paper is structured as follows. In Section 2, we formulate the linear preferential attachment network model and present an efficient simulation method for the network. Section 3 gives parameter estimators when either the full history is known or when only a single snapshot in time is available. We test these estimators against simulated data in Section 5 and then explore the Wiki talk network in Section 6.

## 2. Model specification and simulation

In this section, we present the linear preferential attachment model in detail and provide a fast simulation algorithm for the network.

### 2.1. The linear preferential attachment model

The directed edge preferential attachment model [3, 12] constructs a growing directed random graph $G(n) = (V(n), E(n))$ whose dynamics depend on five non-negative real numbers $\alpha, \beta, \gamma$, $\delta_{in}$ and $\delta_{out}$, where $\alpha + \beta + \gamma = 1$ and $\delta_{in}, \delta_{out} > 0$. To avoid degenerate situations, assume that each of the numbers $\alpha, \beta, \gamma$ is strictly smaller than 1. We obtain a new graph $G(n)$ by adding one edge to the existing graph $G(n-1)$ and index the constructed graphs by the number $n$ of edges in $E(n)$. We start with an arbitrary initial finite directed graph $G(n_0)$ with at least one node and $n_0$ edges. For $n > n_0$, $G(n) = (V(n), E(n))$ is a graph with $|E(n)| = n$ edges and a random number $|V(n)| = N(n)$ of nodes. If $u \in V(n)$, $D_{in}^{(n)}(u)$ and $D_{out}^{(n)}(u)$ denote the in- and out-degree of $u$ respectively in $G(n)$. There are three scenarios that we call the $\alpha$, $\beta$ and $\gamma$-schemes, which are activated by flipping a 3-sided coin whose outcomes are $1, 2, 3$ with probabilities $\alpha, \beta, \gamma$. More formally, we have an iid sequence of multinomial random variables $\{J_n, n > n_0\}$ with cells labelled $1, 2, 3$ and cell probabilities $\alpha, \beta, \gamma$. Then the graph $G(n)$ is obtained from $G(n-1)$ as follows.



$\alpha$-scheme $\qquad\qquad$ $\beta$-scheme $\qquad\qquad$ $\gamma$-scheme

- If $J_n = 1$ (with probability $\alpha$), append to $G(n-1)$ a new node $v \in V(n) \setminus V(n-1)$ and an edge $(v, w)$ leading from $v$ to an existing node $w \in V(n-1)$. Choose the existing node $w \in V(n-1)$ with probability depending on its in-degree in $G(n-1)$:

$$\mathbf{P}[\text{choose } w \in V(n-1)] = \frac{D_{in}^{(n-1)}(w) + \delta_{in}}{n - 1 + \delta_{in} N(n-1)}. \qquad (2.1)$$

- If $J_n = 2$ (with probability $\beta$), add a directed edge $(v, w)$ to $E(n-1)$ with $v \in V(n-1) = V(n)$ and $w \in V(n-1) = V(n)$ and the existing nodes $v, w$

are chosen independently from the nodes of $G(n-1)$ with probabilities

$$\mathbf{P}[\text{choose }(v,w)] = \Big(\frac{D_{\text{out}}^{(n-1)}(v) + \delta_{out}}{n - 1 + \delta_{out}N(n-1)}\Big)\Big(\frac{D_{\text{in}}^{(n-1)}(w) + \delta_{in}}{n - 1 + \delta_{in}N(n-1)}\Big).$$

- If $J_n = 3$ (with probability $\gamma$), append to $G(n-1)$ a new node $w \in V(n) \setminus V(n-1)$ and an edge $(v,w)$ leading from the existing node $v \in V(n-1)$ to the new node $w$. Choose the existing node $v \in V(n-1)$ with probability

$$\mathbf{P}[\text{choose }v \in V(n-1)] = \frac{D_{\text{out}}^{(n-1)}(v) + \delta_{out}}{n - 1 + \delta_{out}N(n-1)}. \tag{2.2}$$

Note that this construction allows the possibility of having self loops in the case where $J_n = 2$, but the proportion of edges that are self loops goes to 0 as $n \to \infty$. Also, multiple edges are allowed between two nodes.

### 2.2. Power law of degree distributions

Given an observed network with $n$ edges, let $N_{ij}(n)$ denote the number of nodes with in-degree $i$ and out-degree $j$. If the network is generated from the linear preferential attachment model described above, then from [3], there exists a proper probability distribution $\{f_{ij}\}$ such that almost surely

$$\frac{N_{ij}(n)}{N(n)} \to f_{ij} =: \frac{p_{ij}}{1 - \beta}, \quad n \to \infty. \tag{2.3}$$

Consider the limiting marginal in-degree distribution $p_i^{\text{in}} := \sum_j p_{ij}$. It is calculated from [3, Equation (3.10)] that

$$p_0^{\text{in}} = \frac{\alpha}{1 + a_1(\delta_{\text{in}})\delta_{\text{in}}},$$

and for $i \geq 1$,

$$p_i^{\text{in}} = \frac{\Gamma(i + \delta_{\text{in}})\Gamma(1 + \delta_{\text{in}} + a_1(\delta_{\text{in}})^{-1})}{\Gamma(i + 1 + \delta_{\text{in}} + a_1(\delta_{\text{in}})^{-1})\Gamma(1 + \delta_{\text{in}})} \left(\frac{\alpha\delta_{\text{in}}}{1 + a_1(\delta_{\text{in}})\delta_{\text{in}}} + \frac{\gamma}{a_1(\delta_{\text{in}})}\right),$$

where

$$a_1(\lambda) := \frac{\alpha + \beta}{1 + \lambda(1 - \beta)}, \quad \lambda > 0.$$

Moreover, $p_i^{\text{in}}$ satisfies

$$p_i^{\text{in}} := \sum_{j=0}^{\infty} p_{ij} \sim C_{\text{in}}i^{-\iota_{\text{in}}} \text{ as } i \to \infty, \quad \text{as long as } \quad \alpha\delta_{\text{in}} + \gamma > 0, \tag{2.4}$$

for some finite positive constant $C_{\text{in}}$, where the power index

$$\iota_{\text{in}} = 1 + \frac{1 + \delta_{\text{in}}(\alpha + \gamma)}{\alpha + \beta}. \tag{2.5}$$

Similarly, the limiting marginal out-degree distribution has the same property:

$$p_j^{\text{out}} := \sum_{i=0}^{\infty} p_{ij} \sim C_{\text{out}} i^{-\iota_{\text{out}}} \text{ as } j \to \infty, \quad \text{as long as } \gamma\delta_{\text{out}} + \alpha > 0,$$

for $C_{\text{out}}$ positive and

$$\iota_{\text{out}} = 1 + \frac{1 + \delta_{\text{out}}(\alpha + \gamma)}{\beta + \gamma}. \tag{2.6}$$

### *2.3. Simulation algorithm*

We describe an efficient simulation procedure for the preferential attachment network given the parameter values $(\alpha, \beta, \gamma, \delta_{in}, \delta_{out})$, where $\alpha + \beta + \gamma = 1$. The simulation cost of the algorithm is linear in time. This algorithm, which was provided by Joyjit Roy during his graduate work at Cornell University, is presented below for completeness. Note that this simulation algorithm is specifically designed for the case where the preferential attachment probabilities (2.1)–(2.2) are linear in the degrees. A similar idea for the simulation of the Yule-Simon process appeared in [20]. Efficient simulation methods for the case where the preferential attachment probabilities are non-linear are studied in [1], where their algorithm trades some efficiency for the flexibility to model non-linear preferential attachment.

Using the notation from the introduction, at time $t = 0$, we initiate with an arbitrary graph $G(n_0) = (V(n_0), E(n_0))$ of $n_0$ edges, where the elements of $E(n_0)$ are represented in form of $(v_i^{(1)}, v_i^{(2)}) \in V(n_0) \times V(n_0)$, $i = 1, \ldots, n_0$, with $v_i^{(1)}, v_i^{(2)}$ denoting the outgoing and incoming vertices of the edge, respectively. To grow the network, we update the network at each stage from $G(n - 1)$ to $G(n)$ by adding a new edge $(v_n^{(1)}, v_n^{(2)})$. Assume that the nodes are labeled using positive integers starting from 1 according to the time order in which they are created, and let the random number $N(n) = |V(n)|$ denote the total number of nodes in $G(n)$.

Let us consider the situation where an existing node is to be chosen from $V(n)$ as the vertex of the new edge. Naively sampling from the multinomial distribution requires $O(N(n))$ evaluations, where $N(n)$ increases linearly with $n$. Therefore the total cost to simulate a network of $n$ edges is $O(n^2)$. This is significantly burdensome when $n$ is large, which is usually the case for observed networks. Algorithm 1 describes a simulation algorithm which uses the alias method [13] for node sampling. Here sampling an existing node from $V(n)$ requires only constant execution time, regardless of $n$. Hence the cost to simulate $G(n)$ is only $O(n)$. This method allows generation of a graph with $10^7$ nodes on a personal laptop in less than 5 seconds.

To see that the algorithm indeed produces the intended network, it suffices to consider the case of sampling an existing node from $V(n-1)$ as the incoming vertex of the new edge. In the function Node_Sample in Algorithm 1, we generate

---

**Algorithm 1:** Simulating a directed edge preferential attachment network

**Algorithm**

**Input**: $\alpha, \beta, \delta_{\mathrm{in}}, \delta_{\mathrm{out}}$, the parameter values; $G(n_0) = (V(n_0), E(n_0))$, the initialization graph; $n$, the targeted number edges

**Output**: $G(n) = (V(n), E(n))$, the resulted graph

$t \leftarrow n_0$
**while** $t < n$ **do**
  $N(t) \leftarrow |V(t)|$
  Generate $U \sim Uniform(0, 1)$
  **if** $U < \alpha$ **then**
    $v^{(1)} \leftarrow N(t) + 1$
    $v^{(2)} \leftarrow \mathsf{Node\_Sample}(E(t), 2, \delta_{\mathrm{in}})$
    $V(t) \leftarrow \mathsf{Append}(V(t), N(t) + 1)$
  **else if** $\alpha \leq U < \alpha + \beta$ **then**
    $v^{(1)} \leftarrow \mathsf{Node\_Sample}(E(t), 1, \delta_{out})$
    $v^{(2)} \leftarrow \mathsf{Node\_Sample}(E(t), 2, \delta_{\mathrm{in}})$
  **else if** $U \geq \alpha + \beta$ **then**
    $v^{(1)} \leftarrow \mathsf{Node\_Sample}(E(t), 1, \delta_{out})$
    $v^{(2)} \leftarrow N(t) + 1$
    $V(t) \leftarrow \mathsf{Append}(V(t), N(t) + 1)$
  $E(t + 1) \leftarrow \mathsf{Append}(E(t), (v^{(1)}, v^{(2)}))$
  $t \leftarrow t + 1$
**end**
**return** $G(n) = (V(n), E(n))$

---

**Function** Node_Sample

**Input**: $E(t)$, the edge list up to time $t$; $j = 1, 2$, the node to be sample, representing outgoing and incoming nodes, respectively; $\delta \in \{\delta_{in}, \delta_{out}\}$, the offset parameter

**Output**: the sampled node, $v$

Generate $W \sim Uniform(0, t + N(t)\delta)$
**if** $W \leq t$ **then**
  $v \leftarrow v^{(j)}_{\lceil W \rceil}$
**else if** $W > t$ **then**
  $v \leftarrow \left\lceil \frac{W - t}{\delta} \right\rceil$
**return** $v$

---

$W \sim \mathrm{Uniform}(0, n - 1 + N(n-1)\delta_{\mathrm{in}})$ and set

$$v \leftarrow v^{(j)}_{\lceil W \rceil} \mathbf{1}_{\{W \leq n-1\}} + \left\lceil \frac{W - (n-1)}{\delta_{\mathrm{in}}} \right\rceil \mathbf{1}_{\{W > n-1\}}.$$

Then

$$\begin{aligned}
\mathbf{P}\left(v = w\right) &= \mathbf{P}\left(v^{(j)}_{\lceil W \rceil} = w\right) \mathbf{P}\left(W \leq n - 1\right) \\
&\quad + \mathbf{P}\left(\left\lceil \frac{W - (n-1)}{\delta_{\mathrm{in}}} \right\rceil = w\right) \mathbf{P}\left(W > n - 1\right) \\
&= \frac{D^{(n-1)}_{\mathrm{in}}(w)}{n - 1} \frac{n - 1}{n - 1 + N(n-1)\delta_{\mathrm{in}}}
\end{aligned}$$

$$+ \frac{1}{N(n-1)} \frac{N(n-1)\delta_{in}}{n-1+N(n-1)\delta_{in}}$$

$$= \frac{D_{in}^{(n-1)}(w) + \delta_{in}}{n-1+N(n-1)\delta_{in}},$$

which corresponds to the desired selection probability (2.1).

## 3. Parameter estimation: MLE based on the full network history

In this section, we estimate the preferential attachment parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \delta_{in}, \delta_{out})$ under two assumptions about what data is available. In the first scenario, the full evolution of the network is observed, from which the likelihood function can be computed. The resulting MLE is strongly consistent and asymptotically normal. For the second scenario, the data only consist of one snapshot of the network with $n$ edges, without the knowledge of the network history that produced these edges. For this scenario we give an estimation approach through approximating the score function and moment matching, which produces parameter estimators that are also strongly consistent but less efficient than those based on the full evolution of the network. In both cases, the estimators are uniquely determined.

### 3.1. Likelihood calculation

Assume the network begins with the graph $G(n_0)$ (consisting of $n_0$ edges) and then evolves according to the description in Section 2.1 with parameters $(\alpha, \beta, \delta_{in}, \delta_{out})$, where $\delta_{in}, \delta_{out} > 0$ and $\alpha, \beta$ are non-negative probabilities. The $\gamma$ is implicitly defined by $\gamma = 1 - \alpha - \beta$. To avoid trivial cases, we will also assume $\alpha, \beta, \gamma < 1$ for the rest of the paper. For MLE estimation we restrict the parameter space for $\delta_{in}, \delta_{out}$ to be $[\epsilon, K]$, for some sufficiently small $\epsilon > 0$ and large $K$. In particular, the true value of $\delta_{in}, \delta_{out}$ is assumed to be contained in $(\epsilon, K)$. Let $e_t = (v_t^{(1)}, v_t^{(2)})$ be the newly created edge when the random graph evolves from $G(t-1)$ to $G(t)$. We sometimes refer to $t$ as the time rather than the number of edges.

Assume we observe the initial graph $G(n_0)$ and the edges $\{e_t\}_{t=n_0+1}^n$ in the order of their formation. For $t = n_0 + 1, \ldots, n$, the values of the following variables are known:

- $N(t)$, the number of nodes in graph $G(t)$;
- $D_{in}^{(t-1)}(v)$, $D_{out}^{(t-1)}(v)$, the in- and out-degree of node $v$ in $G(t-1)$, for all $v \in V(t-1)$;
- $J_t$, the scenario under which $e_t$ is created.

Then the likelihood function is

$$L(\alpha, \beta, \delta_{in}, \delta_{out} | \ G(n_0), (e_t)_{t=n_0+1}^n)$$

$$= \prod_{t=n_0+1}^{n} \left( \alpha \frac{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}}{t-1+\delta_{\text{in}}N(t-1)} \right)^{\mathbf{1}_{\{J_t=1\}}}$$

$$\times \prod_{t=n_0+1}^{n} \left( \beta \left( \frac{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}}{t-1+\delta_{\text{in}}N(t-1)} \right) \left( \frac{D_{\text{out}}^{(t-1)}(v_t^{(1)}) + \delta_{\text{out}}}{t-1+\delta_{\text{out}}N(t-1)} \right) \right)^{\mathbf{1}_{\{J_t=2\}}}$$

$$\times \prod_{t=n_0+1}^{n} \left( (1-\alpha-\beta) \frac{D_{\text{out}}^{(t-1)}(v_t^{(1)}) + \delta_{\text{out}}}{t-1+\delta_{\text{out}}N(t-1)} \right)^{\mathbf{1}_{\{J_t=3\}}} \tag{3.1}$$

and the log likelihood function is

$$\log L(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}} |\ G(n_0), (e_t)_{t=n_0+1}^{n}) \tag{3.2}$$

$$= \log \alpha \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=1\}} + \log \beta \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=2\}} + \log(1-\alpha-\beta) \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}}$$

$$+ \sum_{t=n_0+1}^{n} \log \left( D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}} \right) \mathbf{1}_{\{J_t \in \{1,2\}\}}$$

$$+ \sum_{t=n_0+1}^{n} \log \left( D_{\text{out}}^{(t-1)}(v_t^{(1)}) + \delta_{\text{out}} \right) \mathbf{1}_{\{J_t \in \{2,3\}\}}$$

$$- \sum_{t=n_0+1}^{n} \log(t-1+\delta_{\text{in}}N(t-1)) \mathbf{1}_{\{J_t \in \{1,2\}\}}$$

$$- \sum_{t=n_0+1}^{n} \log(t-1+\delta_{\text{out}}N(t-1)) \mathbf{1}_{\{J_t \in \{2,3\}\}}.$$

The score functions for $\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}$ are calculated as follows:

$$\frac{\partial}{\partial \alpha} \log L(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}} |\ G(n_0), (e_t)_{t=n_0+1}^{n})$$

$$= \frac{1}{\alpha} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=1\}} - \frac{1}{1-\alpha-\beta} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}}, \tag{3.3}$$

$$\frac{\partial}{\partial \beta} \log L(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}} |\ G(n_0), (e_t)_{t=n_0+1}^{n})$$

$$= \frac{1}{\beta} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=2\}} - \frac{1}{1-\alpha-\beta} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}}, \tag{3.4}$$

$$\frac{\partial}{\partial \delta_{\text{in}}} \log L(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}} |\ G(n_0), (e_t)_{t=n_0+1}^{n})$$

$$= \sum_{t=n_0+1}^{n} \frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} \mathbf{1}_{\{J_t \in \{1,2\}\}}$$

$$- \sum_{t=n_0+1}^{n} \frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)} \mathbf{1}_{\{J_t \in \{1,2\}\}}, \tag{3.5}$$

$$\frac{\partial}{\partial \delta_{\text{out}}} \log L(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}} | G(n_0), (e_t)_{t=n_0+1}^n)$$

$$= \sum_{t=n_0+1}^n \frac{1}{D_{\text{out}}^{(t-1)}(v_t^{(1)}) + \delta_{\text{out}}} \mathbf{1}_{\{J_t \in \{2,3\}\}}$$

$$- \sum_{t=n_0+1}^n \frac{N(t-1)}{t-1+\delta_{\text{out}}N(t-1)} \mathbf{1}_{\{J_t \in \{2,3\}\}}.$$

Note that the score functions (3.3), (3.4) for $\alpha$ and $\beta$ do not depend on $\delta_{\text{in}}$ and $\delta_{\text{out}}$. One can show that the Hessian matrix of the log-likelihood for $(\alpha, \beta)$ is positive definite. Setting (3.3) and (3.4) to zero gives the unique MLE estimates for $\alpha$ and $\beta$,

$$\hat{\alpha}^{MLE} = \frac{1}{n-n_0} \sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=1\}}, \tag{3.6}$$

$$\hat{\beta}^{MLE} = \frac{1}{n-n_0} \sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=2\}}. \tag{3.7}$$

These estimates are strongly consistent by applying the strong law of large numbers for the $\{J_t\}_{t \geq n_0+1}$ sequence.

Next, consider the first term of the score function for $\delta_{\text{in}}$ in (3.5), and we have

$$\sum_{t=n_0+1}^n \frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} \mathbf{1}_{\{J_t \in \{1,2\}\}}$$

$$= \sum_{i=0}^\infty \frac{1}{i+\delta_{\text{in}}} \sum_{t=n_0+1}^n \mathbf{1}_{\left\{ D_{\text{in}}^{(t-1)}(v_t^{(2)})=i, J_t \in \{1,2\} \right\}}.$$

Observe that $\left\{ D_{\text{in}}^{(t-1)}(v_t^{(2)}) = i, J_t \in \{1,2\} \right\}$ describes the event that the in-degree of node $v_t^{(2)} \in V(t-1)$ is $i$ at time $t-1$ and is augmented to $i+1$ at time $t$. For each $i \geq 1$, such an event happens at some stage $t \in \{n_0+1, n_0+2, \ldots, n\}$ only for those nodes with in-degree $\leq i$ at time $n_0$ and in-degree $> i$ at time $n$. Let $N_{ij}(n)$ denote the number of nodes with in-degree $i$ and out-degree $j$ at time $n$, and $N_i^{\text{in}}(n)$ and $N_{>i}^{\text{in}}(n)$ to be the number of nodes with in-degree equal to $i$ and greater than $i$, respectively, i.e.,

$$N_i^{\text{in}}(n) = \sum_{j=0}^\infty N_{ij}(n), \quad N_{>i}^{\text{in}}(n) = \sum_{k>i} N_k^{\text{in}}(n).$$

Then

$$\sum_{t=n_0+1}^n \mathbf{1}_{\left\{ D_{\text{in}}^{(t-1)}(v_t^{(2)})=i, J_t \in \{1,2\} \right\}} = N_{>i}^{\text{in}}(n) - N_{>i}^{\text{in}}(n_0), \quad i \geq 1.$$

On the other hand, when $i = 0$, $\left\{ D_{\text{in}}^{(t-1)}(v_t^{(2)}) = 0, J_t \in \{1,2\} \right\}$ occurs for some $t$ if and only if all of the following three events happen:

(i) $v_t^{(2)}$ has in-degree $> 0$ at time $n$;
(ii) $v_t^{(2)}$ does not have in-degree $> 0$ at time $n_0$;
(iii) $v_t^{(2)}$ was not created under the $\gamma$-scheme (otherwise it would have been born with in-degree 1).

This implies:

$$\sum_{t=n_0+1}^{n} \mathbf{1}_{\left\{ D_{\text{in}}^{(t-1)}(v_t^{(2)}) = 0, J_t \in \{1,2\} \right\}} = N_{>0}^{\text{in}}(n) - N_{>0}^{\text{in}}(n_0) - \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}},$$

since there are, in total, $\sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}}$ nodes created under the $\gamma$-scheme. Therefore,

$$\sum_{t=n_0+1}^{n} \frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} \mathbf{1}_{\{J_t \in \{1,2\}\}}$$
$$= \sum_{i=0}^{\infty} \frac{1}{i + \delta_{\text{in}}} \sum_{t=n_0+1}^{n} \mathbf{1}_{\left\{ D_{\text{in}}^{(t-1)}(v_t^{(2)}) = i, J_t \in \{1,2\} \right\}}$$
$$= \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n) - N_{>i}^{\text{in}}(n_0)}{i + \delta_{\text{in}}} - \frac{\sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}}}{\delta_{\text{in}}}. \qquad (3.8)$$

Setting the score function (3.5) for $\delta_{\text{in}}$ to 0 and dividing both sides by $n - n_0$ leads to

$$\frac{1}{n - n_0} \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n) - N_{>i}^{\text{in}}(n_0)}{i + \delta_{\text{in}}} - \frac{1}{\delta_{\text{in}}(n - n_0)} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}}$$
$$- \frac{1}{n - n_0} \sum_{t=n_0+1}^{n} \frac{N(t-1)}{t - 1 + \delta_{\text{in}} N(t-1)} \mathbf{1}_{\{J_t \in \{1,2\}\}} = 0, \qquad (3.9)$$

where the only unknown parameter is $\delta_{\text{in}}$. In Section 3.2, we show that the solution to (3.9) actually maximizes the likelihood function in $\delta_{\text{in}}$. Similarly, the MLE for $\delta_{\text{out}}$ can be solved from

$$\frac{1}{n - n_0} \sum_{j=0}^{\infty} \frac{N_{>j}^{\text{out}}(n) - N_{>j}^{\text{out}}(n_0)}{j + \delta_{\text{out}}} - \frac{\frac{1}{n-n_0} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=1\}}}{\delta_{\text{out}}}$$
$$- \frac{1}{n - n_0} \sum_{t=n_0+1}^{n} \frac{N(t-1)}{t - 1 + \delta_{\text{out}} N(t-1)} \mathbf{1}_{\{J_t \in \{2,3\}\}} = 0,$$

where $N_{>j}^{\text{out}}(n)$ is defined in the same fashion as $N_{>i}^{\text{in}}(n)$.

**Remark 3.1.** *The arguments leading to* (3.8) *allow us to rewrite the likelihood function* (3.1):

$$L(\alpha, \beta, \delta_{in}, \delta_{out}|\ G(n_0), (e_t)_{t=n_0+1}^n)$$

$$= \alpha^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=1\}}}\ \beta^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=2\}}}\ (1-\alpha-\beta)^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}}}$$

$$\times \prod_{t=n_0+1}^n (t-1+\delta_{in}N(t-1))^{-\mathbf{1}_{\{J_t \in \{1,2\}\}}}\ (t-1+\delta_{out}N(t-1))^{-\mathbf{1}_{\{J_t \in \{2,3\}\}}}$$

$$\times \prod_{t=n_0+1}^n \left[\prod_{i=0}^\infty (i+\delta_{in})^{\mathbf{1}\left\{D_{in}^{(t-1)}(v_t^{(2)})=i, J_t \in \{1,2\}\right\}}\right.$$

$$\left. \prod_{j=0}^\infty (j+\delta_{out})^{\mathbf{1}\left\{D_{out}^{(t-1)}(v_t^{(1)})=j, J_t \in \{2,3\}\right\}}\right]$$

$$= \alpha^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=1\}}}\ \beta^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=2\}}}\ (1-\alpha-\beta)^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}}}$$

$$\times \prod_{t=n_0+1}^n \left[(t-1+\delta_{in}N(t-1))^{-\mathbf{1}_{\{J_t \in \{1,2\}\}}}\ (t-1+\delta_{out}N(t-1))^{-\mathbf{1}_{\{J_t \in \{2,3\}\}}}\right.$$

$$\left. \delta_{in}^{-\mathbf{1}_{\{J_t=3\}}}\ \delta_{out}^{-\mathbf{1}_{\{J_t=1\}}}\right]$$

$$\times \prod_{i=0}^\infty (i+\delta_{in})^{N_{>i}^{in}(n)-N_{>i}^{in}(n_0)}\ \prod_{j=0}^\infty (j+\delta_{out})^{N_{>j}^{out}(n)-N_{>j}^{out}(n_0)}.$$

*Hence by the factorization theorem,* $N(n_0)$, $(J_t)_{t=n_0+1}^n$, $(N_{>i}^{in}(n)-N_{>i}^{in}(n_0))_{i \geq 0}$, $(N_{>j}^{out}(n)-N_{>j}^{out}(n_0))_{j \geq 0}$ *are sufficient statistics for* $(\alpha, \beta, \delta_{in}, \delta_{out})$.

### 3.2. Consistency of MLE

We remarked after (3.6) and (3.7) that $\hat{\alpha}^{MLE}$ and $\hat{\beta}^{MLE}$ converge almost surely to $\alpha$ and $\beta$. We now prove that the MLE of $(\delta_{\text{in}}, \delta_{\text{out}})$ is also strongly consistent. Note that if we initiate the network with $G(n_0)$ (for both $n_0$ and $N(n_0)$ finite), then almost surely for all $i, j \geq 0$,

$$\frac{N_{>i}^{\text{in}}(n_0)}{n} \leq \frac{N(n_0)}{n} \to 0, \quad \frac{N_{>j}^{\text{out}}(n_0)}{n} \leq \frac{N(n_0)}{n} \to 0, \quad \text{as } n \to \infty,$$

and $(n - n_0)/n \to 1$. In other words, $n_0$, $N_{>i}^{\text{in}}(n_0)$, $N_{>j}^{\text{out}}(n_0)$ are all $o(n)$. So for simplicity, we assume that the graph is initiated with finitely many nodes and no edges, that is, $n_0 = 0$ and $N(0) \geq 1$. In particular, these assumptions imply the sum of the in-degrees at time $n$ is equal to $n$.

Let $\Psi_n(\cdot), \Phi_n(\cdot)$ be the functional forms of the terms in the log-likelihood function (3.2) involving $\delta_{\text{in}}$ and $\delta_{\text{out}}$ respectively, normalized by $1/n$, i.e.,

$$\Psi_n(\lambda) := \sum_{i=0}^\infty \frac{N_{>i}^{\text{in}}(n)}{n} \log(i+\lambda) - \frac{\log \lambda}{n} \sum_{t=1}^n \mathbf{1}_{\{J_t=3\}}$$

$$-\frac{1}{n}\sum_{t=1}^{n}\log\left(t-1+\lambda N(t-1)\right)\mathbf{1}_{\{J_t\in\{1,2\}\}},$$

$$\Phi_n(\mu) := \sum_{j=0}^{\infty}\frac{N_{>j}^{\text{out}}(n)}{n}\log(j+\mu) - \frac{\log\mu}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t=1\}}$$

$$-\frac{1}{n}\sum_{t=1}^{n}\log\left(t-1+\mu N(t-1)\right)\mathbf{1}_{\{J_t\in\{2,3\}\}}.$$

The following theorem gives the consistency of the MLE of $\delta_{\text{in}}$ and $\delta_{\text{out}}$.

**Theorem 3.2.** *Suppose* $\delta_{in}, \delta_{out} \in (\epsilon, K) \subset (0, \infty)$. *Define*

$$\hat{\delta}_{in}^{MLE} = \hat{\delta}_{in}^{MLE}(n) := \operatorname*{argmax}_{\epsilon \le \lambda \le K} \Psi_n(\lambda), \quad \hat{\delta}_{out}^{MLE} = \hat{\delta}_{out}^{MLE}(n) := \operatorname*{argmax}_{\epsilon \le \mu \le K} \Phi_n(\mu).$$

*Then these are the MLE estimators of* $\delta_{in}, \delta_{out}$ *and they are strongly consistent; that is,*

$$\hat{\delta}_{in}^{MLE} \xrightarrow{a.s.} \delta_{in}, \qquad \hat{\delta}_{out}^{MLE} \xrightarrow{a.s.} \delta_{out}, \qquad n \to \infty.$$

*Proof of Theorem 3.2.* We only verify the consistency of $\hat{\delta}_{\text{in}}^{MLE}$ since similar arguments apply to $\hat{\delta}_{\text{out}}^{MLE}$. Define

$$\psi_n(\lambda) := \Psi_n'(\lambda) = \sum_{i=0}^{\infty}\frac{N_{>i}^{\text{in}}(n)/n}{i+\lambda} - \frac{\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t=3\}}}{\lambda}$$

$$-\frac{1}{n}\sum_{t=1}^{n}\frac{N(t-1)}{t-1+\lambda N(t-1)}\mathbf{1}_{\{J_t\in\{1,2\}\}}.$$

Let us consider a limit version of $\psi_n$:

$$\psi(\lambda) := \sum_{i=0}^{\infty}\frac{p_{>i}^{\text{in}}(\delta_{\text{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - (1-\beta)a_1(\lambda), \tag{3.10}$$

where $p_{>i}^{\text{in}}(\delta_{\text{in}}) := \sum_{k>i}p_k^{\text{in}}(\delta_{\text{in}})$ with $p_k^{\text{in}}(\delta_{\text{in}}) := p_k^{\text{in}}$ as defined in (2.4), and

$$a_1(\lambda) := \frac{\alpha+\beta}{1+\lambda(1-\beta)}, \qquad \lambda > 0.$$

Here we write $p_i^{\text{in}}(\delta_{\text{in}})$ to emphasize the dependence on $\delta_{\text{in}}$. In Lemmas A.1 and A.2, provided in the appendix, it is shown that $\psi(\cdot)$ has a unique zero at $\delta_{\text{in}}$, where $\psi(\lambda) > 0$ when $\lambda < \delta_{\text{in}}$ and $\psi(\lambda) < 0$ when $\lambda > \delta_{\text{in}}$, and

$$\sup_{\lambda \ge \epsilon}|\psi_n(\lambda) - \psi(\lambda)| \to 0. \tag{3.11}$$

Since $\psi$ is continuous, for any $\kappa > 0$ arbitrarily small, there exists $\varepsilon_\kappa > 0$ such that $\psi(\lambda) > \varepsilon_\kappa$ for $\lambda \in [\epsilon, \delta_{\text{in}} - \kappa]$ and $\psi(\lambda) < -\varepsilon_\kappa$ for $\lambda \in [\delta_{\text{in}} + \kappa, K]$. From (3.11),

$$\mathbf{P}\left(\exists N_\kappa \ s.t. \sup_{n>N_\kappa}\sup_{\lambda\in[\epsilon,K]}|\psi_n(\lambda) - \psi(\lambda)| < \varepsilon_\kappa/2\right) = 1. \tag{3.12}$$

Note $\sup_{\lambda \in [\epsilon, K]} |\psi_n(\lambda) - \psi(\lambda)| < \varepsilon_\kappa/2$ implies

$$\psi_n(\lambda) \geq \psi(\lambda) - \sup_{\lambda \in [\epsilon, K]} |\psi_n(\lambda) - \psi(\lambda)| \geq \varepsilon_\kappa - \varepsilon_\kappa/2 > 0, \quad \lambda \in [\epsilon, \delta_{\text{in}} - \kappa),$$

and

$$\psi_n(\lambda) \leq \psi(\lambda) + \sup_{\lambda \in [\epsilon, K]} |\psi_n(\lambda) - \psi(\lambda)| \leq -\varepsilon_\kappa + \varepsilon_\kappa/2 < 0, \quad \lambda \in (\delta_{\text{in}} + \kappa, K].$$

These jointly indicate that $\delta_{\text{in}} - \kappa \leq \hat{\delta}_{\text{in}}^{MLE} \leq \delta_{\text{in}} + \kappa$. Hence (3.12) implies

$$\mathbf{P}\left(\lim_{n \to \infty} |\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}}| \leq \kappa\right) = 1,$$

for arbitrary $\kappa > 0$. That is, $\hat{\delta}_{\text{in}}^{MLE} \xrightarrow{\text{a.s.}} \delta_{\text{in}}$. $\qquad \square$

### 3.3. Asymptotic normality of MLE

In the following theorem, we establish the asymptotic normality for the MLE estimator

$$\hat{\boldsymbol{\theta}}_n^{MLE} = (\hat{\alpha}^{MLE}, \hat{\beta}^{MLE}, \hat{\delta}_{\text{in}}^{MLE}, \hat{\delta}_{\text{out}}^{MLE}).$$

**Theorem 3.3.** *Let $\hat{\boldsymbol{\theta}}_n^{MLE}$ be the MLE estimator for $\boldsymbol{\theta}$, the parameter vector of the preferential attachment model. Then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{MLE} - \boldsymbol{\theta}) \xrightarrow{d} N\left(\mathbf{0}, \Sigma(\boldsymbol{\theta})\right),$$

*where*

$$\Sigma^{-1}(\boldsymbol{\theta}) = I(\boldsymbol{\theta}) := \begin{bmatrix} \frac{1-\beta}{\alpha(1-\alpha-\beta)} & \frac{1}{1-\alpha-\beta} & 0 & 0 \\ \frac{1}{1-\alpha-\beta} & \frac{1-\alpha}{\beta(1-\alpha-\beta)} & 0 & 0 \\ 0 & 0 & I_{in} & 0 \\ 0 & 0 & 0 & I_{out} \end{bmatrix}, \qquad (3.13)$$

*with*

$$I_{in} := \sum_{i=0}^{\infty} \frac{p_{>i}^{in}}{(i + \delta_{in})^2} - \frac{\gamma}{\delta_{in}^2} - \frac{(\alpha + \beta)(1 - \beta)^2}{(1 + \delta_{in}(1 - \beta))^2}, \qquad (3.14)$$

$$I_{out} := \sum_{j=0}^{\infty} \frac{p_{>j}^{out}}{(j + \delta_{out})^2} - \frac{\alpha}{\delta_{out}^2} - \frac{(\gamma + \beta)(1 - \beta)^2}{(1 + \delta_{out}(1 - \beta))^2}.$$

*In particular, $I(\boldsymbol{\theta})$ is the asymptotic Fisher information matrix for the parameters, and hence the MLE estimator is efficient.*

**Remark 3.4.** *From Theorem 3.3, the estimators $(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE})$, $\hat{\delta}_{in}^{MLE}$, and $\hat{\delta}_{out}^{MLE}$ are asymptotically independent.*

*Proof of Theorem 3.3.* We first show the limiting distributions for the MLE's, i.e. $(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE})$, $\hat{\delta}_{\text{in}}^{MLE}$ and $\hat{\delta}_{\text{out}}^{MLE}$. From (3.6) and (3.7),

$$(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE}) = \frac{1}{n} \sum_{t=1}^{n} \left( \mathbf{1}_{\{J_t=1\}}, \mathbf{1}_{\{J_t=2\}} \right),$$

where $\{J_t\}_{t \geq 1}$ is a sequence of iid random variables. Hence the limiting distribution of the pair $\left( \hat{\alpha}^{MLE}, \hat{\beta}^{MLE} \right)$ follows directly from standard central limit theorem for sums of independent random variables.

Next we show the asymptotic normality for $\hat{\delta}_{\text{in}}^{MLE}$; the argument for $\hat{\delta}_{\text{out}}^{MLE}$ is similar. Recall from (3.5) that the score function for $\delta_{\text{in}}$ can be written as

$$\frac{\partial}{\partial \delta_{\text{in}}} \log L(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}) \Big|_{\delta} =: \sum_{t=1}^{n} u_t(\delta),$$

where $u_t$ is defined by

$$u_t(\delta) := \frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta} \mathbf{1}_{\{J_t \in \{1,2\}\}} - \frac{N(t-1)}{t-1+\delta N(t-1)} \mathbf{1}_{\{J_t \in \{1,2\}\}}. \quad (3.15)$$

The MLE estimator $\hat{\delta}_{\text{in}}^{MLE}$ can be obtained by solving $\sum_{t=1}^{n} u_t(\delta) = 0$. By a Taylor expansion of $\sum_{t=1}^{n} u_t(\delta)$,

$$0 = \sum_{t=1}^{n} u_t(\hat{\delta}_{\text{in}}^{MLE}) = \sum_{t=1}^{n} u_t(\delta_{\text{in}}) + (\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}}) \sum_{t=1}^{n} \dot{u}_t(\hat{\delta}_{\text{in}}^*), \quad (3.16)$$

where $\dot{u}_t$ denotes the derivative of $u_t$ and $\hat{\delta}_{\text{in}}^* = \delta_{\text{in}} + \xi(\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}})$ for some $\xi \in [0,1]$. An elementary transformation of (3.16) gives

$$n^{1/2}(\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}}) = \left( -\frac{1}{n^{-1} \sum_{t=1}^{n} \dot{u}_t(\hat{\delta}_{\text{in}}^*)} \right) \left( n^{-1/2} \sum_{t=1}^{n} u_t(\delta_{\text{in}}) \right).$$

To establish

$$n^{1/2}(\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}}) \xrightarrow{d} N(0, I_{\text{in}}^{-1}),$$

where $I_{\text{in}}$ is as defined in (3.13), it suffices to show the following two results:

(i) $n^{-1/2} \sum_{t=1}^{n} u_t(\delta_{\text{in}}) \xrightarrow{d} N(0, I_{\text{in}})$,
(ii) $n^{-1} \sum_{t=1}^{n} \dot{u}_t(\hat{\delta}_{\text{in}}^*) \xrightarrow{p} -I_{\text{in}}$.

These are proved in Lemmas A.3 and A.4 in the appendix, respectively.

To establish the joint asymptotic normality of the MLE estimator $\hat{\boldsymbol{\theta}}_n^{MLE}$, denote the joint score function vector for $\boldsymbol{\theta}$ by

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{\theta}) =: \mathbf{S}_n(\boldsymbol{\theta}) = (S_n(\alpha), S_n(\beta), S_n(\delta_{\text{in}}), S_n(\delta_{\text{out}}))^T,$$

where $S_n(\alpha), S_n(\beta), S_n(\delta_{\text{in}}), S_n(\delta_{\text{out}})$ are the score functions for $\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}$, respectively. A multivariate Taylor expansion gives

$$\mathbf{0} \;=\; \mathbf{S}_n\left(\hat{\boldsymbol{\theta}}_n^{MLE}\right) \;=\; \mathbf{S}_n(\boldsymbol{\theta}) + \dot{\mathbf{S}}_n\left(\hat{\boldsymbol{\theta}}_n^*\right)\left(\hat{\boldsymbol{\theta}}_n^{MLE} - \boldsymbol{\theta}\right), \qquad (3.17)$$

where $\dot{\mathbf{S}}_n$ denotes the Hessian matrix of the log-likelihood function $\log L(\boldsymbol{\theta})$, and $\hat{\boldsymbol{\theta}}_n^* = \boldsymbol{\theta} + \boldsymbol{\xi} \circ \left(\hat{\boldsymbol{\theta}}_n^{MLE} - \boldsymbol{\theta}\right)$ for some vector $\boldsymbol{\xi} \in [0,1]^4$, where "$\circ$" denotes the Hadamard product. From Remark 3.1, the likelihood function $L(\boldsymbol{\theta})$ can be factored into

$$L(\boldsymbol{\theta}) \;=\; f_1(\alpha, \beta) f_2(\delta_{\text{in}}) f_3(\delta_{\text{out}}).$$

Hence

$$\frac{1}{n}\dot{\mathbf{S}}_n(\hat{\boldsymbol{\theta}}_n^*) \;=\; \begin{bmatrix} \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^*)}{\partial \alpha^2} & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^*)}{\partial \alpha \partial \beta} & 0 & 0 \\ \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^*)}{\partial \beta \partial \alpha} & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^*)}{\partial \beta^2} & 0 & 0 \\ 0 & 0 & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^*)}{\partial \delta_{\text{in}}^2} & 0 \\ 0 & 0 & 0 & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^*)}{\partial \delta_{\text{out}}^2} \end{bmatrix} \xrightarrow{p} I(\boldsymbol{\theta})$$

$$(3.18)$$

as implied in the previous part of the proof, where $I(\boldsymbol{\theta})$ (defined in (3.13)) is positive semi-definite.

Note that $(S_n(\alpha), S_n(\beta)), S_n(\delta_{\text{in}}), S_n(\delta_{\text{out}})$ are pairwise uncorrelated. As an example, observe that

$$\begin{aligned}
\mathbf{E}[S_n(\alpha)S_n(\delta_{\text{in}})] &= \int \frac{\partial \log L(\boldsymbol{\theta})}{\partial \alpha} \frac{\partial \log L(\boldsymbol{\theta})}{\partial \delta_{\text{in}}} L(\boldsymbol{\theta}) d\mathbf{x} \\
&= \int \frac{\partial \log f_1(\alpha, \beta)}{\partial \alpha} \frac{\partial \log f_2(\delta_{\text{in}})}{\partial \delta_{\text{in}}} f_1(\alpha, \beta) f_2(\delta_{\text{in}}) f_3(\delta_{\text{out}}) d\mathbf{x} \\
&= \int \frac{\partial f_1(\alpha, \beta)}{\partial \alpha} \frac{\partial f_2(\delta_{\text{in}})}{\partial \delta_{\text{in}}} f_3(\delta_{\text{out}}) \, d\mathbf{x} \\
&= \frac{\partial^2}{\partial \alpha \partial \delta_{\text{in}}} \int L(\boldsymbol{\theta}) d\mathbf{x} \\
&= 0 \;=\; \mathbf{E}[S_n(\alpha)]\mathbf{E}[S_n(\delta_{\text{in}})].
\end{aligned}$$

Using the Cramér-Wold device, the joint convergence of $\mathbf{S}_n(\boldsymbol{\theta})$ follows easily, i.e.,

$$n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, I(\boldsymbol{\theta})).$$

From here, the result of the theorem follows from (3.17) and (3.18). □

## 4. Parameter estimation based on one snapshot

Based only on the single snapshot $G(n)$, we propose a parameter estimation procedure. We assume that the choice of the snapshot does not depend on any endogenous information related to the network. The snapshot merely represents

a point in time where the data is available. Since no information on the initial graph $G(n_0)$ is available, we merely assume $n_0$ and $N(n_0)$ are fixed and $n \to \infty$.

Among the sufficient statistics for $(\alpha, \beta, \delta_{\mathrm{in}}, \delta_{\mathrm{out}})$ derived in Remark 3.1, $\left(N_{>i}^{\mathrm{in}}(n)\right)_{i \geq 0}$, $\left(N_{>j}^{\mathrm{out}}(n)\right)_{j \geq 0}$ are computable from $G(n)$, but the $(J_t)_{t=1}^n$ are not. However, when $n$ is large, we can use the following approximations according to the proof of Lemma A.2:

$$\frac{1}{n} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}} \approx 1 - \alpha - \beta,$$

and

$$\frac{1}{n} \sum_{t=n_0+1}^{n} \frac{N(t)}{t + \delta_{\mathrm{in}} N(t)} \mathbf{1}_{\{J_t \in \{1,2\}\}} \approx (\alpha + \beta) \frac{1 - \beta}{1 + \delta_{\mathrm{in}}(1 - \beta)}.$$

Substituting in (3.9), we estimate $\delta_{\mathrm{in}}$ in terms of $\alpha$ and $\beta$ by solving

$$\sum_{i=0}^{\infty} \frac{N_{>i}^{\mathrm{in}}(n)/n}{i + \delta_{\mathrm{in}}} - \frac{1 - \alpha - \beta}{\delta_{\mathrm{in}}} - \frac{(\alpha + \beta)(1 - \beta)}{1 + (1 - \beta)\delta_{\mathrm{in}}} = 0. \tag{4.1}$$

Note that a strongly consistent estimator of $\beta$ can be obtained directly from $G(n)$:

$$\tilde{\beta} = 1 - \frac{N(n)}{n} \xrightarrow{\text{a.s.}} \beta.$$

To obtain an estimate for $\alpha$, we make use of the recursive formula for $\{p_i^{\mathrm{in}}\}$ in (A.1a):

$$\left(1 + \frac{(\alpha + \beta)\delta_{\mathrm{in}}}{1 + (1 - \beta)\delta_{\mathrm{in}}}\right) p_0^{\mathrm{in}} = \alpha, \tag{4.2}$$

and replace $p_0^{\mathrm{in}}$ by $N_0^{\mathrm{in}}(n)/n$ for large $n$,

$$\left(1 + \frac{(\alpha + \beta)\delta_{\mathrm{in}}}{1 + (1 - \beta)\delta_{\mathrm{in}}}\right) \frac{N_0^{\mathrm{in}}(n)}{n} = \alpha. \tag{4.3}$$

Plug the strongly consistent estimator $\tilde{\beta}$ into (4.1) and (4.3), and we claim that solving the system of equations:

$$\sum_{i=0}^{\infty} \frac{N_{>i}^{\mathrm{in}}(n)/n}{i + \delta_{\mathrm{in}}} - \frac{1 - \alpha - \tilde{\beta}}{\delta_{\mathrm{in}}} - \frac{(\alpha + \tilde{\beta})(1 - \tilde{\beta})}{1 + (1 - \tilde{\beta})\delta_{\mathrm{in}}} = 0, \tag{4.4a}$$

$$\left(1 + \frac{(\alpha + \tilde{\beta})\delta_{\mathrm{in}}}{1 + (1 - \tilde{\beta})\delta_{\mathrm{in}}}\right) \frac{N_0^{\mathrm{in}}(n)}{n} = \alpha, \tag{4.4b}$$

gives the unique solution $(\tilde{\alpha}, \tilde{\delta}_{\mathrm{in}})$ which is strongly consistent for $(\alpha, \delta_{\mathrm{in}})$.

**Theorem 4.1.** *The solution $(\tilde{\alpha}, \tilde{\delta}_{in})$ to the system of equations in (4.4) is unique and strongly consistent for $(\alpha, \delta_{in})$, i.e.,*

$$\tilde{\alpha} \xrightarrow{a.s.} \alpha, \quad \tilde{\delta}_{in} \xrightarrow{a.s.} \delta_{in}.$$

The proof of Theorem 4.1 is given in Section A.3.

The parameters $\tilde{\delta}_{\text{out}}$ and $\tilde{\gamma}$ can be estimated by a mirror argument. We summarize the estimation procedure for $(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}})$ from the snapshot $G(n)$ as follows:

1. Estimate $\beta$ by $\tilde{\beta} = 1 - N(n)/n$.
2. Obtain $\tilde{\delta}_{\text{in}}^0$ by solving (i.e., matching (4.4a) and (4.4b))

$$\sum_{i=1}^{\infty} \frac{N_{>i}^{\text{in}}(n)}{n} \frac{i}{i + \delta_{\text{in}}}(1 + \delta_{\text{in}}(1 - \tilde{\beta})) = \frac{\frac{N_0^{\text{in}}(n)}{n} + \tilde{\beta}}{1 - \frac{N_0^{\text{in}}(n)}{n} \frac{\delta_{\text{in}}}{1 + (1 - \tilde{\beta})\delta_{\text{in}}}}.$$

3. Estimate $\alpha$ by

$$\tilde{\alpha}^0 = \frac{\frac{N_0^{\text{in}}(n)}{n} + \tilde{\beta}}{1 - \frac{N_0^{\text{in}}(n)}{n} \frac{\tilde{\delta}_{\text{in}}^0}{1 + (1 - \tilde{\beta})\tilde{\delta}_{\text{in}}^0}} - \tilde{\beta}.$$

4. Obtain $\tilde{\delta}_{\text{out}}^0$ by solving

$$\sum_{j=1}^{\infty} \frac{N_{>j}^{\text{out}}(n)}{n} \frac{j}{j + \delta_{\text{out}}}(1 + \delta_{\text{out}}(1 - \tilde{\beta})) = \frac{\frac{N_0^{\text{out}}(n)}{n} + \tilde{\beta}}{1 - \frac{N_0^{\text{out}}(n)}{n} \frac{\delta_{\text{out}}}{1 + (1 - \tilde{\beta})\delta_{\text{out}}}}.$$

5. Estimate $\gamma$ by

$$\tilde{\gamma}^0 = \frac{\frac{N_0^{\text{out}}(n)}{n} + \tilde{\beta}}{1 - \frac{N_0^{\text{out}}(n)}{n} \frac{\tilde{\delta}_{\text{out}}^0}{1 + (1 - \tilde{\beta})\tilde{\delta}_{\text{out}}^0}} - \tilde{\beta}.$$

Note that even though all three estimators $\tilde{\alpha}^0, \tilde{\beta}, \tilde{\gamma}^0$ are strongly consistent and hence $\tilde{\alpha}^0 + \tilde{\beta} + \tilde{\gamma}^0 \xrightarrow{\text{a.s.}} 1$, Steps 1–5 do not necessarily imply the strict equality

$$\tilde{\alpha}^0 + \tilde{\beta} + \tilde{\gamma}^0 = 1.$$

We recommend adding the following two steps for a re-normalization to overcome this defect.

6. Re-normalize the probabilities

$$(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) \leftarrow \left( \frac{\tilde{\alpha}^0(1 - \tilde{\beta})}{\tilde{\alpha}^0 + \tilde{\gamma}^0}, \tilde{\beta}, \frac{\tilde{\gamma}^0(1 - \tilde{\beta})}{\tilde{\alpha}^0 + \tilde{\gamma}^0} \right).$$

7. Plug $\tilde{\alpha}$ into (4.4a) to update the estimate of $\delta_{\text{in}}$, i.e., solve for $\tilde{\delta}_{\text{in}}$ from

$$\sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n)/n}{i + \tilde{\delta}_{\text{in}}} - \frac{1 - \tilde{\alpha} - \tilde{\beta}}{\tilde{\delta}_{\text{in}}} - \frac{(\tilde{\alpha} + \tilde{\beta})(1 - \tilde{\beta})}{1 + (1 - \tilde{\beta})\tilde{\delta}_{\text{in}}} = 0.$$

Similarly, solve for $\tilde{\delta}_{\text{out}}$ from

$$\sum_{j=0}^{\infty} \frac{N_{>j}^{\text{out}}(n)/n}{j + \tilde{\delta}_{\text{out}}} - \frac{1 - \tilde{\gamma} - \tilde{\beta}}{\tilde{\delta}_{\text{out}}} - \frac{(\tilde{\gamma} + \tilde{\beta})(1 - \tilde{\beta})}{1 + (1 - \tilde{\beta})\tilde{\delta}_{\text{out}}} = 0.$$
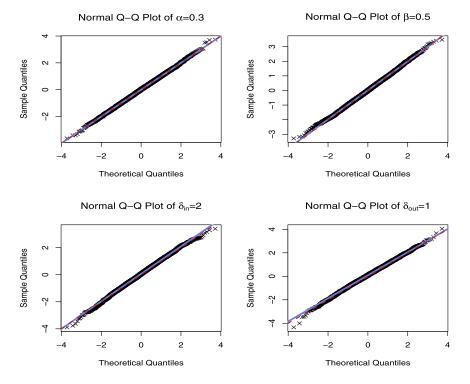
FIG 5.1. *Normal QQ-plots in black for normalized estimates in* (5.2) *under* 5000 *replications of a preferential attachment network with* $10^5$ *edges and* $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$. *The fitted lines in blue are the traditional qq-lines (given by R) used to check normality of the estimates. The red dashed line represents the* $y = x$ *line in all plots.*

## 5. Simulation study

We now apply the estimation procedures described in Sections 3 and 4 to simulated data, which allows us to compare the estimation results using the full history of the network with that using just one snapshot. Algorithm 1 is used to simulate realizations of the preferential attachment network.

### 5.1. MLE

For the scenario of observing the full history of the network, we simulated 5000 independent replications of the preferential attachment network with $10^5$ edges under the true parameter values

$$\boldsymbol{\theta} = (\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}) = (0.3, 0.5, 2, 1). \tag{5.1}$$

For each realization, the MLE estimate of the parameters was computed and standardized as

$$\frac{\sqrt{n}\left((\hat{\boldsymbol{\theta}}_n^{MLE})_i - (\boldsymbol{\theta})_i\right)}{\hat{\sigma}_{ii}}, \tag{5.2}$$

where $(\hat{\boldsymbol{\theta}}_n)_i$ and $(\boldsymbol{\theta})_i$ denote the $i$-th components of $\hat{\boldsymbol{\theta}}_n^{MLE}$ and $\boldsymbol{\theta}$ respectively, and $\hat{\sigma}_{ii}^2$ is the $i$-th diagonal component of the matrix $\hat{\Sigma} := \Sigma(\hat{\boldsymbol{\theta}}_n^{MLE})$. The explicit formula for the entries of $\hat{\Sigma}$ is

$$
\hat{\Sigma} = \begin{bmatrix}
\hat{\alpha}^{MLE}\left(1 - \hat{\alpha}^{MLE}\right) & -\hat{\alpha}^{MLE}\hat{\beta}^{MLE} & 0 & 0 \\
-\hat{\alpha}^{MLE}\hat{\beta}^{MLE} & \hat{\beta}^{MLE}\left(1 - \hat{\beta}^{MLE}\right) & 0 & 0 \\
0 & 0 & \hat{I}_{\mathrm{in}}^{-1} & 0 \\
0 & 0 & 0 & \hat{I}_{\mathrm{out}}^{-1}
\end{bmatrix},
$$

where, see (3.13) and (3.14),

$$
\hat{I}_{\mathrm{in}} = \sum_{i=0}^{\infty} \frac{N_{>i}^{\mathrm{in}}(n)/n}{\left(i + \hat{\delta}_{\mathrm{in}}^{MLE}\right)^2} - \frac{1 - \hat{\alpha}^{MLE} - \hat{\beta}^{MLE}}{\left(\hat{\delta}_{\mathrm{in}}^{MLE}\right)^2}
$$

$$
- \frac{\left(\hat{\alpha}^{MLE} + \hat{\beta}^{MLE}\right)\left(1 - \hat{\beta}^{MLE}\right)^2}{\left(1 + \hat{\delta}_{\mathrm{in}}^{MLE}(1 - \hat{\beta}^{MLE})\right)^2},
$$

$$
\hat{I}_{\mathrm{out}} = \sum_{j=0}^{\infty} \frac{N_{>j}^{\mathrm{out}}(n)/n}{\left(j + \hat{\delta}_{\mathrm{out}}^{MLE}\right)^2} - \frac{\hat{\alpha}^{MLE}}{\left(\hat{\delta}_{\mathrm{out}}^{MLE}\right)^2} - \frac{\left(1 - \hat{\alpha}^{MLE}\right)\left(1 - \hat{\beta}^{MLE}\right)^2}{\left(1 + \hat{\delta}_{\mathrm{out}}^{MLE}(1 - \hat{\beta}^{MLE})\right)^2}.
$$

By the strong consistency of the MLEs combined with Lemma A.2, we have that $\hat{\Sigma} \xrightarrow{\mathrm{a.s.}} \Sigma$.

The QQ-plots of the normalized MLEs are shown in Figure 5.1, all of which line up quite well with the $y = x$ line (the red dashed line). This is consistent with the asymptotic theory described in Theorem 3.3. Confidence intervals for $\boldsymbol{\theta}$ can be obtained using this theorem. Given a single realization, an approximate $(1 - \varepsilon)$-confidence interval for $(\boldsymbol{\theta})_i$ is

$$
(\hat{\boldsymbol{\theta}}_n^{MLE})_i \pm z_{\varepsilon/2}\sqrt{\frac{\hat{\sigma}_{ii}^2}{n}} \quad \text{for } i = 1, \ldots, 4,
$$

where $z_{\varepsilon/2}$ is the upper $\varepsilon/2$ quantile of $N(0, 1)$.

### 5.2. One snapshot

We used the same simulated data as in Section 5.1 to obtain parameter estimates $\tilde{\boldsymbol{\theta}}_n := (\tilde{\alpha}, \tilde{\beta}, \tilde{\delta}_{\mathrm{in}}, \tilde{\delta}_{\mathrm{out}})$ through only the final snapshot, i.e., the set of directed edges without timestamps, following the procedure described at the end of Section 4. For the purpose of comparison with MLE, Figure 5.2 gives the QQ-plots for the normalized estimates from the snapshots using the same standardizations for the MLEs, i.e.,

$$
\frac{\sqrt{n}\left((\tilde{\boldsymbol{\theta}}_n)_i - (\boldsymbol{\theta})_i\right)}{\hat{\sigma}_{ii}}, \quad i = 1, \ldots, 4, \tag{5.3}
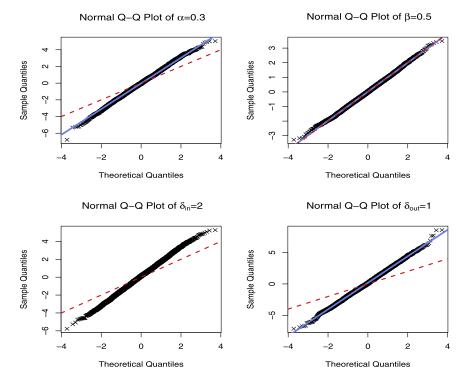$$

FIG 5.2. *Normal QQ-plots for the normalized estimates in (5.3) under* 5000 *replications of a preferential attachment network with* $10^5$ *edges and* $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$. *The fitted lines in blue are the traditional qq-lines used to check normality of the estimates. The red dashed line represents the* $y = x$ *line in all plots.*

where $(\tilde{\boldsymbol{\theta}}_n)_i$ denotes the $i$-th components of $\tilde{\boldsymbol{\theta}}_n$. Again, the fitted lines in blue are the traditional QQ-lines and the red dashed lines are the $y = x$ line. The QQ-plot for $\tilde{\beta}$ exhibits the same shape as for $\hat{\beta}^{MLE}$, since the two estimates are identical.

From Figure 5.2, we see that the snapshot estimates of all four parameters are consistent and approximately normal, i.e., the QQ-plots are linear. However, the slopes of the QQ-lines for $\tilde{\alpha}, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}}$ are much steeper than the diagonal line, indicating a loss of efficiency for $\tilde{\boldsymbol{\theta}}_n$ compared with $\hat{\boldsymbol{\theta}}_n$. Indeed the estimator variance is inflated for all parameters except for $\beta$, where $\tilde{\beta}$ coincides with the true MLE. This is as expected since knowing only the final snapshot provides far less information than the whole network history.

Recall that for a consistent estimator $T_n$ of a one-dimensional parameter $\theta$ constructed from a random sample of size $n$, the asymptotic relative efficiencies (ARE) of $T_n$ is defined by

$$ARE(T_n) := \lim_{n \to \infty} \frac{\text{Var}(\sqrt{n}T_n^*)}{\text{Var}(\sqrt{n}T_n)},$$

where $T_n^*$ denotes the asymptotically efficient estimator. We may compute the ARE's for the snapshot parameter estimates

$$ARE(\tilde{\alpha}) = \lim_{n \to \infty} \frac{n \text{Var}(\hat{\alpha}^{MLE})}{n \text{Var}(\tilde{\alpha})} \approx \frac{\widehat{\text{Var}}(\hat{\alpha}^{MLE})}{\widehat{\text{Var}}(\tilde{\alpha})} \approx 0.398,$$

$$ARE(\tilde{\delta}_{\text{in}}) = \lim_{n \to \infty} \frac{n \text{Var}(\hat{\delta}_{\text{in}}^{MLE})}{n \text{Var}(\tilde{\delta}_{\text{in}})} \approx \frac{\widehat{\text{Var}}(\hat{\delta}_{\text{in}}^{MLE})}{\widehat{\text{Var}}(\tilde{\delta}_{\text{in}})} \approx 0.392,$$

$$ARE(\tilde{\delta}_{\text{out}}) = \lim_{n \to \infty} \frac{n \text{Var}(\hat{\delta}_{\text{out}}^{MLE})}{n \text{Var}(\tilde{\delta}_{\text{out}})} \approx \frac{\widehat{\text{Var}}(\hat{\delta}_{\text{out}}^{MLE})}{\widehat{\text{Var}}(\tilde{\delta}_{\text{out}})} \approx 0.226,$$

where $\widehat{\text{Var}}$ denotes the sample variance of the parameter estimate based on the 5000 replications. Note that $ARE(\tilde{\beta}) = 1$ since $\tilde{\beta} = \hat{\beta}^{MLE}$.

Given a single realization, the variances of the snapshot estimates can be estimated through resampling as follows. Using the estimated parameter $\tilde{\boldsymbol{\theta}}_n$, simulate $10^4$ independent bootstrap replicates of the network with $n = 10^5$ edges. For each simulated network, the snapshot estimate, $\tilde{\boldsymbol{\theta}}_n^* := \left( \tilde{\alpha}^*, \tilde{\beta}^*, \tilde{\delta}_{\text{in}}^*, \tilde{\delta}_{\text{out}}^* \right)$, is computed. The sample variance of these $10^4$ snapshot estimates can then be used as an approximation for the variance of $\tilde{\boldsymbol{\theta}}_n$ so that assuming asymptotic normality, a $(1 - \varepsilon)$-confidence interval for $\boldsymbol{\theta}$ can be approximated by

$$(\tilde{\boldsymbol{\theta}}_n)_i \pm z_{\varepsilon/2} \sqrt{\widehat{\text{Var}}\left( (\tilde{\boldsymbol{\theta}}_n^*)_i \right)} \quad \text{for } i = 1, \ldots, 4,$$

where $z_{\varepsilon/2}$ is the upper $\varepsilon/2$ quantile of $N(0, 1)$.

## 5.3. Sensitivity test

Now we investigate the sensitivity of our estimates while values of the parameters $(n, \alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$ are allowed to vary. First consider the impact of $n$, the number of edges in the network. To do so we held the parameters fixed with values given by (5.1): $(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}) = (0.3, 0.5, 2, 1)$ and varied the value of $n$. The QQ-plots (not presented) for standardized estimates using both full MLE and one-snapshot methods were produced to check the asymptotic normality. When $n = 500, 1000$, diagnostics revealed departures from normality for both the MLE and the snapshot estimates. However, after increasing $n$ to 10000, estimates obtained from both approaches appeared normally distributed as expected.

For each value of $n$ in Table 5.1, 5000 replicates of the network with $n$ edges and parameters $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$ were generated. For each realization, the MLE's $\hat{\boldsymbol{\theta}}_n^{MLE}$ were computed using the full history of the network and the one-snapshot estimates $\tilde{\boldsymbol{\theta}}_n$ were obtained using the 7-step snapshot method proposed in Section 4, pretending that only the last snapshot $G(n)$ was available. The mean for these two estimators were recorded in Table 5.1. There is little bias for both estimates of $\alpha$ and $\beta$, even for small values of $n$. On the other hand, there is some bias for estimated $\delta_{\text{in}}$ and $\delta_{\text{out}}$ for $n \leq 5000$. The magnitudes of

TABLE 5.1

*Mean of $\hat{\boldsymbol{\theta}}_n^{MLE}$ and $\tilde{\boldsymbol{\theta}}_n$ with ARE's of $\tilde{\boldsymbol{\theta}}_n$ relative to $\hat{\boldsymbol{\theta}}_n^{MLE}$ for $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$ under different choices of $n$.*

| $n$ | $Mean(\hat{\boldsymbol{\theta}}_n^{MLE})$ | $Mean(\tilde{\boldsymbol{\theta}}_n)$ | $ARE(\tilde{\boldsymbol{\theta}}_n)$ |
|---|---|---|---|
| 1000 | (0.300, 0.500, 2.076, 1.054) | (0.301, 0.500, 2.128, 1.066) | (0.408, 1.000, 0.397, 0.228) |
| 5000 | (0.300, 0.500, 2.022, 1.013) | (0.301, 0.500, 2.036, 1.010) | (0.414, 1.000, 0.386, 0.236) |
| 10000 | (0.300, 0.500, 2.011, 1.006) | (0.301, 0.500, 2.019, 1.006) | (0.408, 1.000, 0.388, 0.232) |
| 50000 | (0.300, 0.500, 2.003, 1.002) | (0.300, 0.500, 2.005, 1.002) | (0.399, 1.000, 0.393, 0.230) |
| 100000 | (0.300, 0.500, 2.001, 1.001) | (0.300, 0.500, 2.003, 1.000) | (0.392, 1.000, 0.382, 0.223) |

TABLE 5.2

*Mean of $\hat{\boldsymbol{\theta}}_n^{MLE}$ and $\tilde{\boldsymbol{\theta}}_n$ with ARE's of $\tilde{\boldsymbol{\theta}}_n$ relative to $\hat{\boldsymbol{\theta}}_n^{MLE}$ for $(n, \delta_{in}, \delta_{out}) = (10^5, 2, 1)$ under different choices of $(\alpha, \beta)$.*

| $(\alpha, \beta)$ | $Mean(\hat{\boldsymbol{\theta}}_n^{MLE})$ | $Mean(\tilde{\boldsymbol{\theta}}_n)$ | $ARE(\tilde{\boldsymbol{\theta}}_n)$ |
|---|---|---|---|
| (0.001, 0.99) | (0.001, 0.990, 2.034, 1.016) | (0.001, 0.990, 2.071, 1.049) | (0.291, 1.000, 0.147, 0.316) |
| (0.01, 0.9) | (0.010, 0.900, 2.004, 1.001) | (0.010, 0.900, 2.008, 1.004) | (0.331, 1.000, 0.207, 0.381) |
| (0.1, 0.8) | (0.100, 0.800, 2.003, 1.001) | (0.100, 0.800, 2.004, 1.002) | (0.353, 1.000. 0.264, 0.216) |
| (0.2, 0.6) | (0.200, 0.600, 2.002, 1.001) | (0.200, 0.600, 2.003, 1.001) | (0.364, 1.000, 0.309, 0.236) |
| (0.5, 0.3) | (0.500, 0.300, 2.001, 1.001) | (0.500, 0.300, 2.002, 1.000) | (0.472, 1.000, 0.529, 0.202) |
| (0.7, 0.2) | (0.700, 0.200, 2.002, 1.000) | (0.700, 0.200, 2.002, 1.000) | (0.726, 1.000, 0.793, 0.217) |
| (0.1, 0.3) | (0.100, 0.300, 2.001, 1.001) | (0.100, 0.300, 2.002, 1.000) | (0.420, 1.000, 0.313, 0.629) |

the biases for both types of estimates decrease as $n$ increases. Also the ARE's of the snapshot estimator stay within a narrow band as $n$ increases.

Next we held $(n, \delta_{\mathrm{in}}, \delta_{\mathrm{out}}) = (10^5, 2, 1)$ fixed and experimented with various values of $(\alpha, \beta)$ in Table 5.2. For each choice of $(\alpha, \beta)$, 5000 independent realizations of the network were generated and the means of the MLE $\hat{\boldsymbol{\theta}}_n^{MLE}$ and the one-snapshot estimates $\tilde{\boldsymbol{\theta}}_n$ were recorded. Overall, the biases for $\hat{\boldsymbol{\theta}}_n^{MLE}$ are remarkably small for virtually all combinations of parameter values, except for those parameter choices where one of $(\alpha, \beta)$ is extremely small. The biases for the snapshot estimates $\tilde{\boldsymbol{\theta}}_n$ exhibit a similar property, but the magnitudes of the biases are consistently larger than those in the MLE case.

In general, the snapshot estimators are able to achieve 20%–50% efficiency over the range of parameters considered. The loss of efficiency might be less than one would expect given the substantial reduction in the data available to produce the snapshot estimates. It is worth noting that in the case where $(\alpha, \beta) = (0.7, 0.2)$, the efficiencies of the snapshot estimators for $\alpha$ and $\delta_{\mathrm{in}}$ are much larger (0.73 and 0.79, respectively). A heuristic explanation for this increase is that the parameter $\gamma = 1 - \alpha - \beta = 0.1$ is relatively small. By the implicit constraints used for the snapshot estimates, we have

$$\tilde{\alpha} + \tilde{\gamma} = 1 - \tilde{\beta} = 1 - \hat{\beta}^{MLE} = \hat{\alpha}^{MLE} + \hat{\gamma}^{MLE},$$

that is, the snapshot estimate of the sum $\alpha + \gamma$ is the same as the MLE for the sum. Now if $\gamma$ is small, one would expect the resulting estimates to also be small so that $\tilde{\alpha}$ would be nearly the same as $\hat{\alpha}^{MLE}$. Hence the ARE would be close to 1. On the other hand, in the case of a larger $\gamma$, see the bottom row of Table 5.2 in which $\gamma = 0.6$, the ARE for $\alpha$ is not as large (0.42), but the ARE for $\tilde{\delta}_{\mathrm{out}}$ is (0.63).

## 6. Real network example

In this section, we explore fitting a preferential attachment model to a social network. As an illustration, we chose the Dutch Wiki talk network dataset, available on KONECT [14] ([http://konect.uni-koblenz.de/networks/wiki_talk_nl](http://konect.uni-koblenz.de/networks/wiki_talk_nl)). The nodes represent users of Dutch Wikipedia, and an edge from node A to node B refers to user A writing a message on the talk page of user B at a certain time point. The network consists of 225,749 nodes (users) and 1,554,699 edges (messages). All edges are recorded with timestamps.

In order to accommodate all the edge formulation scenarios appearing in the dataset, we extend our model by appending the following two interaction schemes ($J_n = 4, 5$) in addition to the existing three ($J_n = 1, 2, 3$) described in Section 2.1.

- If $J_n = 4$ (with probability $\xi$), append to $G(n-1)$ two new nodes $v, w \in V(n) \setminus V(n-1)$ and an edge connecting them $(v, w)$.
- If $J_n = 5$ (with probability $\rho$), append to $G(n-1)$ a new node $v \in V(n) \setminus V(n-1)$ with self loop $(v, v)$.

These scenarios have been observed in other social network data, such as the network that models Facebook wall posts ([http://konect.uni-koblenz.de/networks/facebook-wosn-wall](http://konect.uni-koblenz.de/networks/facebook-wosn-wall)). They occur in small proportions and can be easily accommodated by a slight modification in the model fitting procedure. The new model has parameter vector $(\alpha, \beta, \gamma, \xi, \delta_{\text{in}}, \delta_{\text{out}})$, and $\rho$ is implicitly defined through $\rho = 1 - (\alpha + \beta + \gamma + \xi)$. Similar to the derivations in Section 3, the MLE estimators for $\alpha, \beta, \gamma, \xi$ are

$$\hat{\alpha}^{MLE} = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{J_t=1\}}, \quad \hat{\beta}^{MLE} = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{J_t=2\}},$$

$$\hat{\gamma}^{MLE} = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{J_t=3\}}, \quad \hat{\xi}^{MLE} = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{J_t=4\}},$$

and $\delta_{\text{in}}, \delta_{\text{out}}$ can be obtained through solving

$$\sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n)/n}{i + \delta_{\text{in}}} - \frac{\frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{J_t \in \{3,4,5\}\}}}{\delta_{\text{in}}} - \frac{1}{n} \sum_{t=1}^{n} \frac{N(t)}{t + \delta_{\text{in}} N(t)} \mathbf{1}_{\{J_t \in \{1,2\}\}} = 0,$$

$$\sum_{j=0}^{\infty} \frac{N_{>j}^{\text{out}}(n)/n}{j + \delta_{\text{out}}} - \frac{\frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{J_t \in \{1,4,5\}\}}}{\delta_{\text{out}}} - \frac{1}{n} \sum_{t=1}^{n} \frac{N(t)}{t + \delta_{\text{out}} N(t)} \mathbf{1}_{\{J_t \in \{2,3\}\}} = 0.$$

We first naively fit the linear preferential attachment model to the full network using MLE. The MLE estimators are

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\xi}, \hat{\rho}, \hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}}) =$$
$$(3.08 \times 10^{-3}, 8.55 \times 10^{-1}, 1.39 \times 10^{-1}, 4.76 \times 10^{-5}, 3.06 \times 10^{-3}, 0.547, 0.134). \tag{6.1}$$
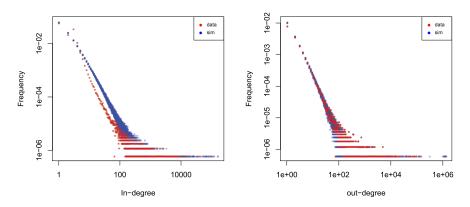
FIG 6.1. *Empirical in- and out-degree frequencies of the full Wiki talk network (red) and that from 20 realizations of the linear preferential attachment network with fitted parameter values (6.1) from MLE (blue). The scatter plots for the degree frequencies from the 20 simulations are overlaid together to form an informal confidence region for the degree distribution of the fitted model*

To evaluate the goodness-of-fit, 20 network realizations were simulated from the fitted model. We overlaid the empirical in- and out-degree frequencies of the original network with that of the simulations. If the model fits the data well, the degree frequencies of the data should lie within the range formed by that of the simulations, which gives an informal confidence region for the degree distributions. From Figure 6.1, we see that while the data roughly agrees with the simulations in the out-degree frequencies, the deviation in the in-degree frequencies is noticeable.

To better understand the discrepancy in the in-degree frequencies, we examined the link data and their timestamps and discovered bursts of messages originating from certain nodes over small time intervals. According to Wikipedia policy [23], certain administrating accounts are allowed to send group messages to multiple users simultaneously. These bursts presumably represent broadcast announcements generated from these accounts. These administrative broadcasts can also be detected if we apply the linear preferential attachment model to the network in local time intervals. We divided the total time frame down to subintervals of varying length each containing the formation of $10^4$ edges. The number $10^4$ is chosen to ensure good asymptotics as shown in Table 5.1. This process generated 155 networks,

$$G(n_{k-1}), \ldots, G(n_k - 1), \quad k = 1, \ldots, 155.$$

For each of the 155 datasets, we fit a preferential attachment model using MLE. The resulting estimates $(\hat{\delta}_{in}, \hat{\delta}_{out})$ are plotted against the corresponding timeline on the upper left panel of Figure 6.2. Notice that $\hat{\delta}_{in}$ exhibits large spikes at various times. Recall from (2.1), a large value of $\delta_{in}$ indicates that the probability of an existing node $v$ receiving a new message becomes less dependent on its in-degree, i.e., previous popularity. These spikes appear to be directly related to the
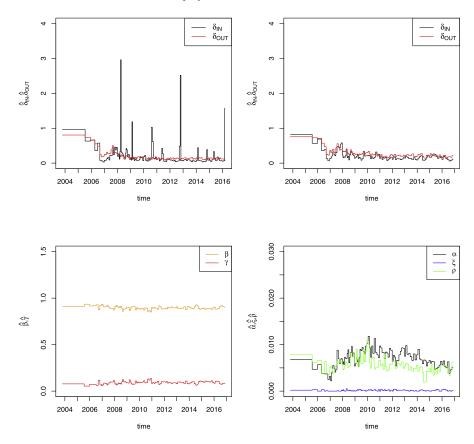
FIG 6.2. *Local parameter estimates of the linear preferential attachment model for the full and reduced Wiki talk network. Upper left: $(\hat{\delta}_{in}, \hat{\delta}_{out})$ for the full network. Upper right, lower left, lower right: $(\hat{\delta}_{in}, \hat{\delta}_{out})$, $(\hat{\beta}, \hat{\gamma})$, $(\hat{\alpha}, \hat{\xi}, \hat{\rho})$ for the reduced network, respectively.*

occurrences of group messages. This plot is truncated after the day 2016/3/16, on which a massive group message of size 48,957 was sent and the model can no longer be fit.

We identified 37 users who have sent, at least once, 40 or more consecutive messages in the message history. This is evidence that group messages were sent by this user. We presume these nodes are administrative accounts; they are responsible for about 30% of the total messages sent. Since their behavior cannot be regarded as normal social interaction, we excluded messages from these accounts from the dataset in our analysis. We then also removed nodes with zero in- and out-degrees.

The re-estimated parameters after the data cleaning are displayed in the other three panels of Figure 6.2. Here all parameter estimates are quite stable through time.

The reduced network now contains 112,919 nodes and 1,086,982 edges, to which we fit the linear preferential attachment model. The fitted parameters
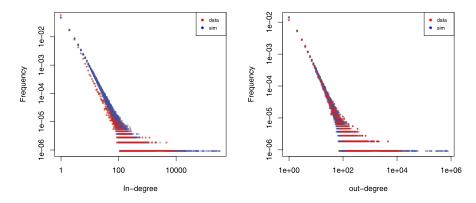
FIG 6.3. *Empirical in- and out-degree frequencies of the reduced Wiki talk network (red) and that from 20 realizations of the linear preferential attachment network with fitted parameter values* (6.2) *from MLE (blue).*

based on MLE for our reduced dataset are

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\xi}, \hat{\rho}, \hat{\delta}_{\mathrm{in}}, \hat{\delta}_{\mathrm{out}}) =$$
$$(6.95 \times 10^{-3}, 8.96 \times 10^{-1}, 9.10 \times 10^{-2}, 1.44 \times 10^{-4}, 5.61 \times 10^{-3}, 0.174, 0.257).$$
$$(6.2)$$

Again the degree distributions of the data and 20 simulations from the fitted model are displayed in Figure 6.3. The out-degree distribution of the data agrees reasonably well with the simulations. For the in-degree distribution, the fit is better than that for the entire dataset (Figure 6.1). However, for smaller in-degrees, the fitted model over-estimates the in-degree frequencies. We speculate that in many social networks, the out-degree is in line with that predicted by the preferential attachment model. An individual node would be more likely to reach out to others if having done so many times previously. For in-degrees, the situation is complicated and may depend on a multitude of factors. For instance, the choice of recipient may depend on the community that the sender is in, the topic being discussed in the message, etc. As an example, a group leader might send messages to his/her team on a regular basis. Such examples violate the base assumptions of the preferential attachment model and could result in the deviation between the data and the simulations.

Next we consider the estimation method of Section 4 applied to a single snapshot of the data. In order to implement this procedure, we donned blinders and assumed that our dataset consists only of the information of the wiki data at the last timestamp. That is, information about administrative broadcasts, and other aspects of the data learned by looking at the previous history of the data are unavailable. In particular, we would have no knowledge of the existence of the two additional scenarios corresponding to $J_n = 4, 5$. With this in mind, we fit the three scenario model using the methods in Section 4. The fitted parameters
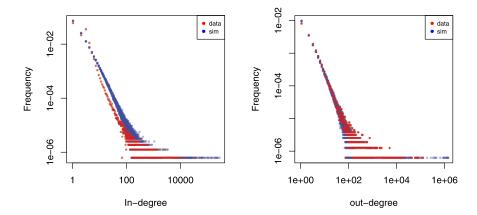
FIG 6.4. *Empirical in- and out-degree frequencies of the full Wiki talk network (red) and that from 20 realizations of the linear preferential attachment network with fitted parameter values* (6.3) *from the snapshot estimator (blue).*

are

$$(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}}) = (5.80 \times 10^{-4}, 8.55 \times 10^{-1}, 1.45 \times 10^{-1}, 0.199, 0.165). \quad (6.3)$$

The comparison of the degree distributions between the data and simulations from the fitted model is displayed in Figure 6.4 and is not too dissimilar to the plots in Figure 6.1 that are based on maximum likelihood estimation using the full network data. In particular, the out-degree distribution is matched reasonably well, but the fitted model does a poor job of capturing the in-degree distribution.

We see from this example that while the linear preferential attachment model is perhaps too simplistic for the Wiki talk network dataset, it has the ability to illuminate some gross features, such as the out-degrees, as well as to capture important structural changes such as the group message behavior. Consequently, despite its limitation, this model may be used as a building block for more flexible models. Modifications to the existing model formulation and more careful analysis of change points in parameters are directions for future research.

## Appendix A: Proofs

### A.1. For the proof of Theorem 3.2: Lemmas A.1 and A.2

**Lemma A.1.** *For $\lambda > 0$, the function $\psi(\lambda)$ in* (3.10) *has a unique zero at $\delta_{in}$ and, $\psi(\lambda) > 0$ when $\lambda < \delta_{in}$ and $\psi(\lambda) < 0$ when $\lambda > \delta_{in}$.*

*Proof.* The probabilities $\{p_i^{\text{in}}(\lambda)\}$ satisfy the recursions in $i$ (cf. [3]):

$$p_0^{\text{in}}(\lambda) \left( \lambda + \frac{1}{a_1(\lambda)} \right) = \frac{\alpha}{a_1(\lambda)}, \qquad (A.1a)$$

$$p_1^{\text{in}}(\lambda)\left(1+\lambda+\frac{1}{a_1(\lambda)}\right) = \lambda p_0^{\text{in}}(\lambda) + \frac{\gamma}{a_1(\lambda)},$$

$$p_2^{\text{in}}(\lambda)\left(2+\lambda+\frac{1}{a_1(\lambda)}\right) = (1+\lambda)p_1^{\text{in}}(\lambda),$$

$$\vdots$$

$$p_i^{\text{in}}(\lambda)\left(i+\lambda+\frac{1}{a_1(\lambda)}\right) = (i-1+\lambda)p_{i-1}^{\text{in}}(\lambda), \quad (i \geq 2),$$

where $a_1(\lambda) := (\alpha+\beta)/(1+\lambda(1-\beta))$. Summing the recursions in (A.1) from $0$ to $i$, we get (with the convention that $\sum_{i=0}^{-1} = 0$)

$$\sum_{k=0}^{i} p_k^{\text{in}}(\lambda)\left(k+\lambda+\frac{1}{a_1(\lambda)}\right) = \sum_{k=0}^{i-1}(k+\lambda)p_k^{\text{in}}(\lambda) + \frac{\alpha}{a_1(\lambda)} + \frac{\gamma}{a_1(\lambda)}\mathbf{1}_{\{i\geq1\}}, \quad i \geq 0,$$

which can be simplified to

$$\frac{1}{a_1(\lambda)}\sum_{k=0}^{i} p_k^{\text{in}}(\lambda) + (i+\lambda)p_i^{\text{in}}(\lambda) = \frac{1-\beta}{a_1(\lambda)} - \frac{\gamma}{a_1(\lambda)}\mathbf{1}_{\{i=0\}}, \quad i \geq 0. \qquad \text{(A.2)}$$

From (2.3),

$$\sum_{i=0}^{\infty} p_i^{\text{in}}(\lambda) = \sum_{i,j} p_{ij}(\lambda) = 1-\beta. \qquad \text{(A.3)}$$

Hence by rearranging (A.2), we have

$$(i+\lambda)p_i^{\text{in}}(\lambda) + \frac{\gamma}{a_1(\lambda)}\mathbf{1}_{\{i=0\}} = \frac{1}{a_1(\lambda)}\left(1-\beta-\sum_{k=0}^{i} p_k^{\text{in}}(\lambda)\right) = \frac{1}{a_1(\lambda)}p_{>i}^{\text{in}}(\lambda),$$

or equivalently,

$$p_{>i}^{\text{in}}(\lambda) = a_1(\lambda)(i+\lambda)p_i^{\text{in}}(\lambda) + \gamma\mathbf{1}_{\{i=0\}}. \qquad \text{(A.4)}$$

Now with the help of (A.3) and (A.4), we can rewrite $\psi(\lambda)$ in the following way:

$$\begin{aligned}
\psi(\lambda) &= \sum_{i=0}^{\infty} \frac{p_{>i}^{\text{in}}(\delta_{\text{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - (1-\beta)a_1(\lambda) \\
&= \sum_{i=0}^{\infty} \frac{p_{>i}^{\text{in}}(\delta_{\text{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})a_1(\lambda)(i+\lambda)}{i+\lambda} \\
&= \sum_{i=0}^{\infty} \frac{a_1(\delta_{\text{in}})(i+\delta_{\text{in}})p_i^{\text{in}}(\delta_{\text{in}}) + \gamma\mathbf{1}_{\{i=0\}}}{i+\lambda} - \frac{\gamma}{\lambda} - \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})a_1(\lambda)(i+\lambda)}{i+\lambda} \\
&= \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}\Big(a_1(\delta_{\text{in}})(i+\delta_{\text{in}}) - a_1(\lambda)(i+\lambda)\Big)
\end{aligned}$$

$$= \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda} \int_{\lambda}^{\delta_{\text{in}}} \frac{\partial}{\partial s} \Big( a_1(s)(i+s) \Big) ds$$

$$= \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda} \int_{\lambda}^{\delta_{\text{in}}} \frac{(\alpha+\beta)(1-i(1-\beta))}{(1+s(1-\beta))^2} ds$$

$$= \left( \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}(1-i(1-\beta)) \right) \int_{\lambda}^{\delta_{\text{in}}} \frac{\alpha+\beta}{(1+s(1-\beta))^2} ds$$

$$=: C(\lambda) \int_{\lambda}^{\delta_{\text{in}}} \frac{\alpha+\beta}{(1+s(1-\beta))^2} ds. \tag{A.5}$$

The series defining $C(\lambda)$ converges absolutely for any $\lambda > 0$ since

$$\sum_{i=0}^{\infty} \left| \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}(1-i(1-\beta)) \right| < \sum_{i=0}^{\infty} p_i^{\text{in}}(\delta_{\text{in}}) \left| \frac{i(1-\beta)}{i+\lambda} + \frac{1}{i+\lambda} \right|$$

$$< (1-\beta)(1-\beta+\frac{1}{\lambda}) < \infty.$$

Summing over $i$ in (A.4), we get by monotone convergence

$$\sum_{i=0}^{\infty} p_{>i}^{\text{in}}(\lambda) = \sum_{i=0}^{\infty} i p_i^{\text{in}}(\lambda) = a_1(\lambda) \sum_{i=0}^{\infty} i p_i^{\text{in}}(\lambda) + a_1(\lambda)\lambda \sum_{i=0}^{\infty} p_i^{\text{in}}(\lambda) + \gamma.$$

The infinite series converge because $p_i^{\text{in}}(\lambda)$ is a power law with index greater than 2; see (2.4) and (2.5). Solving for the infinite series we get

$$\sum_{i=0}^{\infty} i p_i^{\text{in}}(\lambda) = \frac{a_1(\lambda)\lambda}{1-a_1(\lambda)}(1-\beta) + \frac{\gamma}{1-a_1(\lambda)} = 1. \tag{A.6}$$

Hence we have

$$C(\lambda) = \sum_{i \le (1-\beta)^{-1}} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}(1-i(1-\beta)) - \sum_{i > (1-\beta)^{-1}} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}(i(1-\beta)-1)$$

$$> \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{(1-\beta)^{-1}+\lambda}(1-i(1-\beta))$$

$$= \frac{1}{(1-\beta)^{-1}+\lambda} \sum_{i=0}^{\infty} p_i^{\text{in}}(\delta_{\text{in}}) - \frac{1-\beta}{(1-\beta)^{-1}+\lambda} \sum_{i=0}^{\infty} i p_i^{\text{in}}(\delta_{\text{in}})$$

$$= \frac{1}{(1-\beta)^{-1}+\lambda}(1-\beta) - \frac{1-\beta}{(1-\beta)^{-1}+\lambda}1$$

$$= 0.$$

Now recall from (A.5) that $\psi(\lambda)$ is of the form

$$\psi(\lambda) = C(\lambda) \int_{\lambda}^{\delta_{\text{in}}} \frac{\alpha+\beta}{(1+s(1-\beta))^2} ds,$$

where $C(\lambda) > 0$ for all $\lambda > 0$. Therefore $\psi(\cdot)$ has a unique zero at $\delta_{\text{in}}$ and $\psi(\lambda) > 0$ when $\lambda < \delta_{\text{in}}$ and $\psi(\lambda) < 0$ when $\lambda > \delta_{\text{in}}$. $\qquad\square$

We show the uniform convergence of $\psi_n$ to $\psi$ in the next lemma.

**Lemma A.2.** *As $n \to \infty$, for any $\epsilon > 0$,*

$$\sup_{\lambda \geq \epsilon} |\psi_n(\lambda) - \psi(\lambda)| \xrightarrow{a.s.} 0.$$

*Proof.* By the definition of $\psi$, $p_{>i}^{\text{in}}(\delta_{\text{in}})$ is a function of $\delta_{\text{in}}$ and is a constant with respect to $\lambda$. Hence we suppress the dependence on $\delta_{\text{in}}$ and simply write it as $p_{>i}^{\text{in}}$ when considering the difference $\psi_n - \psi$ as a function of $\lambda$:

$$\psi_n(\lambda) - \psi(\lambda) = \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n)/n - p_{>i}^{\text{in}}}{i + \lambda} - \frac{1}{\lambda}\left(\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t=3\}} - (1 - \alpha - \beta)\right)$$
$$- \frac{1}{n}\sum_{t=1}^{n}\left(\frac{N(t-1)}{t - 1 + \lambda N(t-1)}\mathbf{1}_{\{J_t \in \{1,2\}\}} - \frac{(1-\beta)(\alpha+\beta)}{1 + \lambda(1-\beta)}\right).$$

Thus,

$$\sup_{\lambda \geq \epsilon} |\psi_n(\lambda) - \psi(\lambda)|$$
$$\leq \sup_{\lambda \geq \epsilon} \sum_{i=0}^{\infty} \frac{\left|N_{>i}^{\text{in}}(n)/n - p_{>i}^{\text{in}}\right|}{i + \lambda} + \sup_{\lambda \geq \epsilon} \frac{1}{\lambda}\left|\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t=3\}} - (1 - \alpha - \beta)\right|$$
$$+ \sup_{\lambda \geq \epsilon}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{N(t-1)}{t - 1 + \lambda N(t-1)}\mathbf{1}_{\{J_t \in \{1,2\}\}} - \frac{(1-\beta)(\alpha+\beta)}{1 + \lambda(1-\beta)}\right|. \qquad (A.7)$$

For the first term, note that for all $i \geq 0$,

$$iN_{>i}^{\text{in}}(n) = \sum_{k=i+1}^{\infty} N_k^{\text{in}}(n)i \leq \sum_{k=1}^{\infty} kN_k^{\text{in}}(n) = n,$$

since the assumption on initial conditions implies the sum of in-degrees at $n$ is $n$. Therefore $N_{>i}^{\text{in}}(n)/n \leq i^{-1}$ for $i \geq 1$, and it then follows that

$$\sum_{i=0}^{\infty} \frac{\left|N_{>i}^{\text{in}}(n)/n - p_{>i}^{\text{in}}\right|}{i + \lambda} \leq \sum_{i=0}^{M} \frac{\left|N_{>i}^{\text{in}}(n)/n - p_{>i}^{\text{in}}\right|}{i + \lambda} + \sum_{i=M+1}^{\infty} \frac{1/i}{i + \lambda} + \sum_{i=M+1}^{\infty} \frac{p_{>i}^{\text{in}}}{i + \lambda}.$$

Note that the last two terms on the right side can be made arbitrarily small uniformly on $[\epsilon, \infty)$ if we choose $M$ sufficiently large.

Recall the convergence of the degree distribution $\{N_{ij}(n)/N(n)\}$ to the probability distribution $\{f_{ij}\}$ in (2.3), we have

$$\frac{N_{>i}^{\text{in}}(n)}{n} = \frac{N(n)}{n}\frac{N_{>i}^{\text{in}}(n)}{N(n)} \xrightarrow{a.s.} (1 - \beta)\sum_{l \geq 0, k > i} f_{kl} = p_{>i}^{\text{in}}, \quad \forall i \geq 0.$$

Hence, for any fixed $M$,

$$\sum_{i=0}^{M} \frac{\left|N_{>i}^{\text{in}}(n)/n - p_{>i}^{\text{in}}\right|}{i+\epsilon} \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \to \infty.$$

which implies further that choosing $M$ arbitrarily large gives

$$\sup_{\lambda \geq \epsilon} \sum_{i=0}^{\infty} \frac{\left|N_{>i}^{\text{in}}(n)/n - p_{>i}^{\text{in}}\right|}{i+\lambda}$$
$$\leq \sum_{i=0}^{M} \frac{\left|N_{>i}^{\text{in}}(n)/n - p_{>i}^{\text{in}}\right|}{i+\epsilon} + \sum_{i=M+1}^{\infty} \frac{1/i}{i+\epsilon} + \sum_{i=M+1}^{\infty} \frac{p_{>i}^{\text{in}}}{i+\epsilon} \xrightarrow{\text{a.s.}} 0.$$

The second term in (A.7) converges to 0 almost surely by strong law of large numbers, and the third term in (A.7) can be written as

$$\left| \frac{1}{n} \sum_{t=1}^{n} \left( \frac{N(t-1)}{t-1+\lambda N(t-1)} - \frac{(1-\beta)}{1+\lambda(1-\beta)} \right) \mathbf{1}_{\{J_t \in \{1,2\}\}} \right.$$
$$\left. + \frac{1-\beta}{1+\lambda(1-\beta)} \frac{1}{n} \sum_{t=1}^{n} \left( \mathbf{1}_{\{J_t \in \{1,2\}\}} - (\alpha+\beta) \right) \right|,$$

which is bounded by

$$\left| \frac{1}{n} \sum_{t=1}^{n} \frac{N(t-1)}{t-1+\lambda N(t-1)} - \frac{(1-\beta)}{1+\lambda(1-\beta)} \right|$$
$$+ \frac{1-\beta}{1+\lambda(1-\beta)} \left| \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{J_t \in \{1,2\}\}} - (\alpha+\beta) \right|.$$

We have

$$\sup_{\lambda \geq \epsilon} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{N(t-1)}{t-1+\lambda N(t-1)} - \frac{(1-\beta)}{1+\lambda(1-\beta)} \right|$$
$$= \sup_{\lambda \geq \epsilon} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{N(t-1)/(t-1)-(1-\beta)}{(1+\lambda N(t-1)/(t-1))(1+\lambda(1-\beta))} \right|$$
$$\leq \frac{1}{n} \sum_{t=1}^{n} \left| \frac{N(t-1)/(t-1)-(1-\beta)}{(1+\epsilon N(t-1)/(t-1))(1+\epsilon(1-\beta))} \right|,$$

which converges to 0 almost surely by Cesàro convergence of random variables, since

$$\left| \frac{N(n)/n - (1-\beta)}{(1+\epsilon N(n)/n)(1+\epsilon(1-\beta))} \right| \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \to \infty.$$

Further, by the strong law of large numbers,

$$\sup_{\lambda \geq \epsilon} \frac{1-\beta}{1+\lambda(1-\beta)} \left| \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{J_t \in \{1,2\}\}} - (\alpha+\beta) \right|$$

$$\leq \frac{1-\beta}{1+\epsilon(1-\beta)} \left| \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{J_t \in \{1,2\}\}} - (\alpha+\beta) \right| \xrightarrow{\text{a.s.}} 0, \ \text{ as } n \to \infty.$$

Hence the third term of (A.7) also goes to 0 almost surely as $n \to \infty$. The result of the lemma follows. $\qquad\square$

### A.2. For the proof of Theorem 3.3: Lemmas A.3 and A.4

**Lemma A.3.** *As* $n \to \infty$,

$$n^{-1/2} \sum_{t=1}^{n} u_t(\delta_{in}) \xrightarrow{d} N(0, I_{in}). \tag{A.8}$$

*Proof.* Let $\mathcal{F}_n = \sigma(G(0), \ldots, G(n))$ be the $\sigma$-field generated by the information contained in the graphs. We first observe that $\{\sum_{t=1}^{n} u_t(\delta_{in}), \mathcal{F}_n, n \geq 1\}$ is a martingale. To see this, note from (3.15) that $|u_t(\delta)| \leq 2/\delta$ and

$$\mathbf{E}[u_t(\delta_{in})|\mathcal{F}_{t-1}]$$

$$= \mathbf{E}\left[ \frac{1}{D_{in}^{(t-1)}(v_t^{(2)}) + \delta_{in}} \mathbf{1}_{\{J_t \in \{1,2\}\}} \,\middle|\, \mathcal{F}_{t-1} \right]$$

$$\quad - \frac{N(t-1)}{t-1+\delta_{in}N(t-1)} \mathbf{E}[\mathbf{1}_{\{J_t \in \{1,2\}\}}|\mathcal{F}_{t-1}]$$

$$= \mathbf{E}\left[ \frac{1}{D_{in}^{(t-1)}(v_t^{(2)}) + \delta_{in}} \,\middle|\, J_t = 1, \mathcal{F}_{t-1} \right] \mathbf{P}[J_t = 1]$$

$$\quad + \mathbf{E}\left[ \frac{1}{D_{in}^{(t-1)}(v_t^{(2)}) + \delta_{in}} \,\middle|\, J_t = 2, \mathcal{F}_{t-1} \right] \mathbf{P}[J_t = 2]$$

$$\quad - (\alpha+\beta)\frac{N(t-1)}{t-1+\delta_{in}N(t-1)}$$

$$= (\alpha+\beta) \sum_{v \in V_{t-1}} \frac{1}{D_{in}^{(t-1)}(v) + \delta_{in}} \frac{D_{in}^{(t-1)}(v) + \delta_{in}}{t-1+\delta_{in}N(t-1)}$$

$$\quad - (\alpha+\beta)\frac{N(t-1)}{t-1+\delta_{in}N(t-1)}$$

$$= (\alpha+\beta) \left( \sum_{v \in V_{t-1}} \frac{1}{t-1+\delta_{in}N(t-1)} - \frac{N(t-1)}{t-1+\delta_{in}N(t-1)} \right)$$

$$= 0,$$

which satisfies the definition of a martingale difference. Hence

$$\left\{ n^{-1/2} \sum_{r=1}^{t} u_r(\delta_{in}) \right\}_{t=1,\ldots,n}$$

is a zero-mean, square-integrable martingale array. The convergence (A.8) follows from the martingale central limit theory (cf. Theorem 3.2 of [7]) if the following three conditions can be verified:

(a) $n^{-1/2} \max_t |u_t(\delta_{\text{in}})| \xrightarrow{p} 0$,
(b) $n^{-1} \sum_t u_t^2(\delta_{\text{in}}) \xrightarrow{p} I_{\text{in}}$,
(c) $\mathbf{E}\left(n^{-1} \max_t u_t^2(\delta_{\text{in}})\right)$ is bounded in $n$.

Since $|u_t(\delta_{\text{in}})| \leq 2/\delta_{\text{in}}$, we have

$$n^{-1/2} \max_t |u_t(\delta_{\text{in}})| \leq \frac{2}{n^{1/2}\delta_{\text{in}}} \to 0,$$

and

$$n^{-1} \max_t u_t^2 \leq \frac{4}{n\delta_{\text{in}}^2} \to 0.$$

Hence conditions (a) and (c) are straightforward.

To show (b), observe that

$$\frac{1}{n} \sum_{t=1}^n u_t^2(\delta_{\text{in}})$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbf{1}_{\{J_t \in \{1,2\}\}} \left( \frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} - \frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)} \right)^2$$

$$= \frac{1}{n} \sum_{t=1}^n \frac{\mathbf{1}_{\{J_t \in \{1,2\}\}}}{\left( D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}} \right)^2}$$

$$\quad - \frac{2}{n} \sum_{t=1}^n \frac{\mathbf{1}_{\{J_t \in \{1,2\}\}}}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} \frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)}$$

$$\quad + \frac{1}{n} \sum_{t=1}^n \mathbf{1}_{\{J_t \in \{1,2\}\}} \left( \frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)} \right)^2$$

$$=: T_1 - 2T_2 + T_3.$$

Following the calculations in the proof of Lemma A.2, we have for $T_1$,

$$T_1 = \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n)/n}{(i+\delta_{\text{in}})^2} - \frac{1}{\delta_{\text{in}}^2} \frac{1}{n} \sum_{t=1}^n \mathbf{1}_{\{J_t=3\}} \xrightarrow{p} \sum_{i=0}^{\infty} \frac{p_{>i}^{\text{in}}}{(i+\delta_{\text{in}})^2} - \frac{\gamma}{\delta_{\text{in}}^2}.$$

We then rewrite $T_2$ as

$$T_2 = \frac{1}{n} \sum_{t=1}^n \frac{\mathbf{1}_{\{J_t \in \{1,2\}\}}}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} \left( \frac{N(t-1)/(t-1)}{1+\delta_{\text{in}}N(t-1)/(t-1)} - \frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)} \right)$$

$$\quad + \frac{1}{n} \sum_{t=1}^n \frac{\mathbf{1}_{\{J_t \in \{1,2\}\}}}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} \frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)}$$

$$=: T_{21} + T_{22},$$

where

$$|T_{21}| \leq \frac{1}{n}\sum_{t=1}^{n}\frac{1}{\delta_{\mathrm{in}}}\left|\frac{N(t-1)/(t-1)}{1+\delta_{\mathrm{in}}N(t-1)/(t-1)} - \frac{1-\beta}{1+\delta_{\mathrm{in}}(1-\beta)}\right| \xrightarrow{p} 0$$

by Cesàro's convergence and

$$T_{22} = \frac{1-\beta}{1+\delta_{\mathrm{in}}(1-\beta)}\left(\sum_{i=0}^{\infty}\frac{N_{>i}^{\mathrm{in}}(n)/n}{i+\delta_{\mathrm{in}}} - \frac{1}{\delta_{\mathrm{in}}}\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t=3\}}\right)$$

$$\xrightarrow{p} \frac{1-\beta}{1+\delta_{\mathrm{in}}(1-\beta)}\left(\sum_{i=0}^{\infty}\frac{p_{>i}^{\mathrm{in}}}{i+\delta_{\mathrm{in}}} - \frac{\gamma}{\delta_{\mathrm{in}}}\right) = \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{\mathrm{in}}(1-\beta))^2},$$

where the equality follows from (A.4). For $T_3$, similar to $T_1$, we have

$$T_3 = \frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t\in\{1,2\}\}}\left(\left(\frac{N(t-1)/(t-1)}{1+\delta_{\mathrm{in}}N(t-1)/(t-1)}\right)^2 - \frac{(1-\beta)^2}{(1+\delta_{\mathrm{in}}(1-\beta))^2}\right)$$

$$+ \frac{(1-\beta)^2}{(1+\delta_{\mathrm{in}}(1-\beta))^2}\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t\in\{1,2\}\}} \xrightarrow{p} \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{\mathrm{in}}(1-\beta))^2}.$$

Combining these results together,

$$\frac{1}{n}\sum_{t=1}^{n}u_t^2(\delta_{\mathrm{in}}) = T_1 - 2(T_{21}+T_{22}) + T_3$$

$$\xrightarrow{p} \sum_{i=0}^{\infty}\frac{p_{>i}^{\mathrm{in}}}{(i+\delta_{\mathrm{in}})^2} - \frac{\gamma}{\delta_{\mathrm{in}}^2} - \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{\mathrm{in}}(1-\beta))^2} = I_{\mathrm{in}}. \qquad (\mathrm{A}.9)$$

This completes the proof.                                                     □

**Lemma A.4.** *As $n\to\infty$,*

$$\frac{1}{n}\sum_{t=1}^{n}\dot{u}_t(\hat{\delta}_{in}^*) \xrightarrow{p} -I_{in}.$$

*Proof.* The result of this lemma can be established by showing first

$$\frac{1}{n}\sum_{t=1}^{n}\dot{u}_t(\delta_{\mathrm{in}}) \xrightarrow{p} -I_{\mathrm{in}} \qquad (\mathrm{A}.10)$$

and then

$$\left|\frac{1}{n}\sum_{t=1}^{n}\dot{u}_t(\hat{\delta}_{\mathrm{in}}^*) - \frac{1}{n}\sum_{t=1}^{n}\dot{u}_t(\delta_{\mathrm{in}})\right| \xrightarrow{p} 0. \qquad (\mathrm{A}.11)$$

We first observe that

$$\dot{u}_t(\delta) = -\left(\frac{1}{D_{\mathrm{in}}^{(t-1)}(v_t^{(2)})+\delta}\right)^2\mathbf{1}_{\{J_t\in\{1,2\}\}}$$

$$+ \left( \frac{N(t-1)}{t-1+\delta N(t-1)} \right)^2 \mathbf{1}_{\{J_t \in \{1,2\}\}}$$

$$= -u_t^2(\delta) - 2u_t(\delta) \frac{N(t-1)}{t-1+\delta N(t-1)}.$$

Recall the definition and convergence result for $T_2$ and $T_3$ in Lemma A.3, we have

$$\frac{1}{n} \sum_{t=1}^{n} u_t(\delta_{\text{in}}) \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} = T_2 - T_3 \xrightarrow{p} 0.$$

Also from (A.9),

$$\frac{1}{n} \sum_{t=1}^{n} u_t^2(\delta_{\text{in}}) \xrightarrow{p} I_{\text{in}}.$$

Hence

$$\frac{1}{n} \sum_{t=1}^{n} \dot{u}_t(\delta_{\text{in}}) = -\frac{1}{n} \sum_{t=1}^{n} u_t^2(\delta_{\text{in}}) - \frac{2}{n} \sum_{t=1}^{n} u_t(\delta_{\text{in}}) \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \xrightarrow{p} -I_{\text{in}}$$

and (A.10) is established.

By construction and definition, we have $\hat{\delta}_{\text{in}}, \hat{\delta}_{\text{in}}^*, \delta_{\text{in}} > 0$. To prove (A.11), note that

$$|u_t(\hat{\delta}_{\text{in}}^*) - u_t(\delta_{\text{in}})|$$

$$\leq \mathbf{1}_{\{J_t \in \{1,2\}\}} \left| \frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \hat{\delta}_{\text{in}}^*} - \frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} \right|$$

$$+ \mathbf{1}_{\{J_t \in \{1,2\}\}} \left| \frac{N(t-1)}{t-1+\hat{\delta}_{\text{in}}^* N(t-1)} - \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \right|$$

$$= \mathbf{1}_{\{J_t \in \{1,2\}\}} \left| \frac{\delta_{\text{in}} - \hat{\delta}_{\text{in}}^*}{\left( D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \hat{\delta}_{\text{in}}^* \right) \left( D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}} \right)} \right|$$

$$+ \mathbf{1}_{\{J_t \in \{1,2\}\}} \left| \frac{(N(t-1))^2 (\delta_{\text{in}} - \hat{\delta}_{\text{in}}^*)}{\left( t-1+\hat{\delta}_{\text{in}}^* N(t-1) \right) (t-1+\delta_{\text{in}} N(t-1))} \right|$$

$$\leq \frac{2|\hat{\delta}_{\text{in}}^* - \delta_{\text{in}}|}{\hat{\delta}_{\text{in}}^* \delta_{\text{in}}}.$$

Then

$$|u_t^2(\hat{\delta}_{\text{in}}^*) - u_t^2(\delta_{\text{in}})| = \left| u_t(\hat{\delta}_{\text{in}}^*) - u_t(\delta_{\text{in}}) \right| \left| u_t(\hat{\delta}_{\text{in}}^*) + u_t(\delta_{\text{in}}) \right|$$

$$\leq \frac{2 \left| \hat{\delta}_{\text{in}}^* - \delta_{\text{in}} \right|}{\hat{\delta}_{\text{in}}^* \delta_{\text{in}}} \left( \frac{2}{\hat{\delta}_{\text{in}}^*} + \frac{2}{\delta_{\text{in}}} \right),$$

and

$$\left| u_t(\hat{\delta}_{\text{in}}^*) \, \frac{N(t-1)}{t-1+\hat{\delta}_{\text{in}}^* N(t-1)} - u_t(\delta_{\text{in}}) \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \right|$$

$$\leq \left| u_t(\hat{\delta}_{\text{in}}^*) - u_t(\delta_{\text{in}}) \right| \frac{\frac{N(t-1)}{t-1}}{1+\delta_{\text{in}} \frac{N(t-1)}{t-1}}$$

$$+ \left| u_t(\hat{\delta}_{\text{in}}^*) \right| \left| \frac{\frac{N(t-1)}{t-1}}{1+\hat{\delta}_{\text{in}}^* \frac{N(t-1)}{t-1}} - \frac{\frac{N(t-1)}{t-1}}{1+\delta_{\text{in}} \frac{N(t-1)}{t-1}} \right|$$

$$\leq \frac{2 \left| \hat{\delta}_{\text{in}}^* - \delta_{\text{in}} \right|}{\hat{\delta}_{\text{in}}^* \delta_{\text{in}}} \frac{1}{\delta_{\text{in}}} + \frac{2}{\hat{\delta}_{\text{in}}^*} \frac{\left| \hat{\delta}_{\text{in}}^* - \delta_{\text{in}} \right|}{\hat{\delta}_{\text{in}}^* \delta_{\text{in}}}.$$

From Theorem 3.2, $\hat{\delta}_{\text{in}}^{MLE}$ is consistent for $\delta_{\text{in}}$, hence

$$\left| \hat{\delta}_{\text{in}}^* - \delta_{\text{in}} \right| \leq \left| \hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}} \right| \xrightarrow{p} 0.$$

We have

$$\left| \frac{1}{n} \sum_{t=1}^n \dot{u}_t(\hat{\delta}_{\text{in}}^*) - \frac{1}{n} \sum_{t=1}^n \dot{u}_t(\delta_{\text{in}}) \right|$$

$$\leq \frac{1}{n} \sum_{t=1}^n \left| \dot{u}_t(\hat{\delta}_{\text{in}}^*) - \dot{u}_t(\delta_{\text{in}}) \right| \leq \frac{1}{n} \sum_{t=1}^n \left| u_t^2(\hat{\delta}_{\text{in}}^*) - u_t^2(\delta_{\text{in}}) \right|$$

$$+ \frac{2}{n} \sum_{t=1}^n \left| u_t(\hat{\delta}_{\text{in}}^*) \frac{N(t-1)}{t-1+\hat{\delta}_{\text{in}}^* N(t-1)} - u_t(\delta_{\text{in}}) \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \right|$$

$$\leq \frac{2 \left| \hat{\delta}_{\text{in}}^* - \delta_{\text{in}} \right|}{\hat{\delta}_{\text{in}}^* \delta_{\text{in}}} \left( \frac{2}{\hat{\delta}_{\text{in}}^*} + \frac{2}{\delta_{\text{in}}} \right) + \frac{4 \left| \hat{\delta}_{\text{in}}^* - \delta_{\text{in}} \right|}{\hat{\delta}_{\text{in}}^* \delta_{\text{in}}} \frac{1}{\delta_{\text{in}}} + \frac{4}{\hat{\delta}_{\text{in}}^*} \frac{\left| \hat{\delta}_{\text{in}}^* - \delta_{\text{in}} \right|}{\hat{\delta}_{\text{in}}^* \delta_{\text{in}}} \xrightarrow{p} 0.$$

This proves (A.11) and completes the proof of Lemma A.4.                    □

### A.3. Proof of Theorem 4.1

*Proof.* First observe that $\sum_i i N_i^{\text{in}}(n)$ sums up to the total number of edges $n$, so

$$\sum_{i=0}^\infty \frac{N_{>i}^{\text{in}}(n)}{n} = \sum_{i=0}^\infty \frac{i N_i^{\text{in}}(n)}{n} = 1.$$

We can re-write (4.4a) as

$$\alpha + \tilde{\beta} = \left( \frac{1}{\delta_{\text{in}}} - \sum_{i=0}^\infty \frac{N_{>i}^{\text{in}}(n)/n}{i+\delta_{\text{in}}} \right) \bigg/ \left( \frac{1}{\delta_{\text{in}}} - \frac{1-\tilde{\beta}}{1+\delta_{\text{in}}(1-\tilde{\beta})} \right)$$

$$= \left( \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n)/n}{\delta_{\text{in}}} - \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n)/n}{i + \delta_{\text{in}}} \right) \Big/ \left( \frac{1}{\delta_{\text{in}}(1 + \delta_{\text{in}}(1 - \tilde{\beta}))} \right)$$

$$= \sum_{i=1}^{\infty} \frac{N_{>i}^{\text{in}}(n)}{n} \frac{i}{i + \delta_{\text{in}}} \left( 1 + \delta_{\text{in}}(1 - \tilde{\beta}) \right) =: f_n(\delta_{\text{in}}), \qquad \text{(A.12)}$$

and (4.4b) as

$$\alpha + \tilde{\beta} = \left( \frac{N_0^{\text{in}}(n)}{n} + \tilde{\beta} \right) \Big/ \left( 1 - \frac{N_0^{\text{in}}(n)}{n} \frac{\delta_{\text{in}}}{1 + (1 - \tilde{\beta})\delta_{\text{in}}} \right) =: g_n(\delta_{\text{in}}).$$

Then $\tilde{\delta}_{\text{in}}$ can be obtained by solving

$$f_n(\delta) - g_n(\delta) = 0, \qquad \delta \in [\epsilon, K].$$

Similar to the proof of Theorem 3.2, we define the limit versions of $f_n$, and $g_n$ as follows:

$$f(\delta) := \sum_{i=1}^{\infty} p_{>i}^{\text{in}} \frac{i}{i + \delta} (1 + \delta(1 - \beta)),$$

$$g(\delta) := \left( p_0^{\text{in}} + \beta \right) \Big/ \left( 1 - p_0^{\text{in}} \frac{\delta}{1 + (1 - \beta)\delta} \right), \qquad \delta \in [\epsilon, K].$$

Now we apply the re-parametrization

$$\eta := \frac{\delta}{1 + \delta(1 - \beta)} \in \left[ \frac{1}{\epsilon^{-1} + 1 - \beta}, \frac{1}{K^{-1} + 1 - \beta} \right] =: \mathcal{I} \qquad \text{(A.13)}$$

to $f$ and $g$, such that

$$\tilde{f}(\eta) := f(\delta(\eta)) = \sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + (i^{-1} - (1 - \beta))\eta},$$

$$\tilde{g}(\eta) := g(\delta(\eta)) = \frac{p_0^{\text{in}} + \beta}{1 - \eta p_0^{\text{in}}}.$$

Note that for all $\eta \in \mathcal{I}$:

- Set $b_i(\eta) := (i^{-1} - (1 - \beta))\eta$, then $1 + b_i(\eta) > 0$ for all $i \geq 1$. So that $\tilde{f}(\eta) > 0$ on $\mathcal{I}$;
- $\tilde{f}(\eta) \leq \frac{1}{1 - (1 - \beta)\eta} \sum_{i=0}^{\infty} p_{>i}^{\text{in}} \leq 1 + (1 - \beta)K < \infty.$

Meanwhile, $\tilde{g}$ is also well defined and strictly positive for $\eta \in \mathcal{I}$ because

$$1/p_0^{\text{in}} > 1/(1 - \beta) > \eta. \qquad \text{(A.14)}$$

The first inequality holds since:

$$1/p_0^{\text{in}} > 1/(1-\beta) \Leftrightarrow p_0^{\text{in}} < 1 - \beta$$
$$\Leftrightarrow \frac{\alpha}{1 + \frac{(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}}} < 1 - \beta$$
$$\Leftrightarrow \alpha + \beta < 1 + \frac{(1-\beta)(\alpha+\beta)\delta_{\text{in}}}{1 + (1-\beta)\delta_{\text{in}}}$$
$$\Leftrightarrow \alpha + \beta < 1 + (1-\beta)\delta_{\text{in}}.$$

We know $\alpha + \beta < 1$ by our model assumption, thus verifying (A.14).

Define for $\eta \in \mathcal{I}$,

$$\tilde{h}(\eta) := \frac{1}{\tilde{f}(\eta)} - \frac{1}{\tilde{g}(\eta)} = \left( \sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + (i^{-1} - (1-\beta))\eta} \right)^{-1} - \frac{1 - \eta p_0^{\text{in}}}{p_0^{\text{in}} + \beta},$$

then it follows that

$$\tilde{h}(\eta) = 0 \quad \Leftrightarrow \quad \tilde{f}(\eta) = \tilde{g}(\eta), \qquad \eta \in \mathcal{I}.$$

We now show that $\tilde{h}$ is concave and $\tilde{h}(\eta) \to 0$ as $\eta \to 0$, then the uniqueness of the solution follows.

First observe that

$$\frac{\partial^2}{\partial \eta^2} \tilde{h}(\eta) = \frac{\partial^2}{\partial \eta^2} \left( \sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + (i^{-1} - (1-\beta))\eta} \right)^{-1}$$
$$= \frac{\partial^2}{\partial \eta^2} \left( \sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right)^{-1}$$
$$= 2 \left( \sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right)^{-3} \left[ \frac{\partial}{\partial \eta} \left( \sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right) \right]^2$$
$$- \left( \sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right)^{-2} \frac{\partial^2}{\partial \eta^2} \left( \sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right). \qquad (\text{A}.15)$$

We now claim that

$$\frac{\partial}{\partial \eta} \left( \sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right) = \sum_{i=1}^{\infty} \frac{\partial}{\partial \eta} \left( \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right) = - \sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}(i^{-1} - (1-\beta))}{(1 + b_i(\eta))^2},$$
$$(\text{A}.16)$$

$$\frac{\partial^2}{\partial \eta^2} \left( \sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right) = \sum_{i=1}^{\infty} \frac{\partial^2}{\partial \eta^2} \left( \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right) = 2 \sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}(i^{-1} - (1-\beta))^2}{(1 + b_i(\eta))^3}.$$
$$(\text{A}.17)$$

It suffices to check:

$$\sum_{i=1}^{\infty} \sup_{\eta \in \mathcal{I}} \left| \frac{\partial}{\partial \eta} \left( \frac{p_{>i}^{\mathrm{in}}}{1 + b_i(\eta)} \right) \right| < \infty, \qquad \sum_{i=1}^{\infty} \sup_{\eta \in \mathcal{I}} \left| \frac{\partial^2}{\partial \eta^2} \left( \frac{p_{>i}^{\mathrm{in}}}{1 + b_i(\eta)} \right) \right| < \infty.$$

Note that for $i \geq 1$,

$$\begin{aligned}
\sup_{\eta \in \mathcal{I}} \left| \frac{\partial}{\partial \eta} \left( \frac{p_{>i}^{\mathrm{in}}}{1 + b_i(\eta)} \right) \right| &= \sup_{\eta \in \mathcal{I}} \frac{p_{>i}^{\mathrm{in}} |i^{-1} - (1 - \beta)|}{(1 + b_i(\eta))^2} \\
&\leq (2 - \beta) \sup_{\eta \in \mathcal{I}} \frac{p_{>i}^{\mathrm{in}}}{(1 + b_i(\eta))^2} \\
&\leq (2 - \beta)(1 + (1 - \beta)K)^2 p_{>i}^{\mathrm{in}}.
\end{aligned}$$

Recall (A.6), we then have

$$\sum_{i=0}^{\infty} p_{>i}^{\mathrm{in}} = \sum_{i=0}^{\infty} \sum_{k>i} p_k^{\mathrm{in}} = \sum_{k=0}^{\infty} \sum_{i=0}^{k-1} p_k^{\mathrm{in}} = \sum_{k=0}^{\infty} k p_k^{\mathrm{in}} = 1.$$

Hence,

$$\begin{aligned}
\sum_{i=1}^{\infty} \sup_{\eta \in \mathcal{I}} \left| \frac{\partial}{\partial \eta} \left( \frac{p_{>i}^{\mathrm{in}}}{1 + b_i(\eta)} \right) \right| &\leq (2 - \beta)(1 + (1 - \beta)K)^2 \sum_{i=0}^{\infty} p_{>i}^{\mathrm{in}} \\
&= (2 - \beta)(1 + (1 - \beta)K)^2 < \infty,
\end{aligned}$$

which implies (A.16). Equation (A.17) then follows by a similar argument. Combining (A.15), (A.16) and (A.17) gives

$$\frac{\partial^2}{\partial \eta^2} \tilde{h}(\eta) = 2 \left( \sum_{i=1}^{\infty} \frac{p_{>i}^{\mathrm{in}}}{1 + b_i(\eta)} \right)^{-3}$$

$$\times \left[ \left( \sum_{i=1}^{\infty} \frac{p_{>i}^{\mathrm{in}}(i^{-1} - (1 - \beta))}{(1 + b_i(\eta))^2} \right)^2 - \left( \sum_{i=1}^{\infty} \frac{p_{>i}^{\mathrm{in}}}{1 + b_i(\eta)} \right) \left( \sum_{i=1}^{\infty} \frac{p_{>i}^{\mathrm{in}}(i^{-1} - 1 + \beta)^2}{(1 + b_i(\eta))^3} \right) \right]$$

$$< 0,$$

by the Cauchy-Schwarz inequality. Hence $\tilde{h}$ is concave on $\mathcal{I}$.

From Lemma A.1, $\psi(\delta_{\mathrm{in}}) = 0$ where $\psi(\cdot)$ is as defined in (3.10). Hence we have $f(\delta_{\mathrm{in}}) = \alpha + \beta$ in a similar derivation to that of (A.12). Also from (4.2), we have $g(\delta_{\mathrm{in}}) = \alpha + \beta$. Hence, $\delta_{\mathrm{in}}$ is a solution to $f(\delta) = g(\delta)$.

Under the $\delta \mapsto \eta$ reparametrization in (A.13), we have that $\tilde{f}(\eta_{\mathrm{in}}) = \tilde{g}(\eta_{\mathrm{in}})$ where $\eta_{\mathrm{in}} := \delta_{\mathrm{in}}/(1 + \delta_{\mathrm{in}}(1 - \beta))$, and also

$$\lim_{\eta \downarrow 0} \tilde{f}(\eta) = \sum_{i=1}^{\infty} p_{>i}^{\mathrm{in}} = 1 - p_{>0}^{\mathrm{in}} = \beta + p_0^{\mathrm{in}} = \lim_{\eta \downarrow 0} \tilde{g}(\eta).$$

This, along with the concavity of $\tilde{h}$, implies that $\eta_{\text{in}}$ is the unique solution to $\tilde{h}(\eta) = 0$, or equivalently, to $\tilde{f}(\eta) = \tilde{g}(\eta)$ on $\mathcal{I}$.

Let $\tilde{f}_n(\eta) := f_n(\delta(\eta))$, $\tilde{g}_n(\eta) := g_n(\delta(\eta))$. We can show in a similar fashion that $\tilde{\eta} := \tilde{\delta}_{\text{in}}/(1 - \tilde{\delta}_{\text{in}}(1 - \tilde{\beta}))$ is the unique solution to $\tilde{f}_n(\eta) = \tilde{g}_n(\eta)$. Using an analogue of the arguments in the proof of Theorem A.2, we have

$$\sup_{\eta \in \mathcal{I}} |\tilde{f}_n(\eta) - \tilde{f}(\eta)| \xrightarrow{\text{a.s.}} 0, \quad \sup_{\eta \in \mathcal{I}} |\tilde{g}_n(\eta) - \tilde{g}(\eta)| \xrightarrow{\text{a.s.}} 0,$$

and therefore $\tilde{\eta} \xrightarrow{\text{a.s.}} \eta_{\text{in}}$. Since $\delta \mapsto \eta$ is a one-to-one transformation from $[\epsilon, K]$ to $\mathcal{I}$, we have that $\tilde{\delta}_{\text{in}}$ is the unique solution to $f_n(\delta) = g_n(\delta)$ and that $\tilde{\delta}_{\text{in}} \xrightarrow{\text{a.s.}} \delta_{\text{in}}$. On the other hand, $\tilde{\alpha}$ can be solved uniquely by plugging $\tilde{\delta}_{\text{in}}$ into (A.12) and is also strongly consistent, which completes the proof. $\qquad\square$

## Acknowledgement

## References

[1] J. Atwood, B. Ribeiro, and D. Towsley. Efficient network generation under general preferential attachment. *Computational Social Networks*, 2(1):7, 2015.

[2] S. Bhamidi. Universal techniques to analyze preferential attachment trees: Global and local analysis. *Preprint, available: http://www.unc.edu/~bhamidi/preferent.pdf*, 2007.

[3] 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. MR1974912

[4] R. T. Durrett. *Random Graph Dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010. MR2656427

[5] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets*. Cambridge University Press, Cambridge, 2010. MR2677125

[6] F. Gao and A. van der Vaart. On the asymptotic normality of estimating the affine preferential attachment network models with random initial degrees. *Stochastic Processes and their Applications*, 2017.

[7] P. Hall and C. C. Heyde. *Martingale Limit Theory and its Application.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980. Probability and Mathematical Statistics. MR0624435

[8] J. H. Jones and M. S. Handcock. An assessment of preferential attachment as a mechanism for human sexual network formation. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1520):1123–1128, 2003.

[9] J. H. Jones and M. S. Handcock. Social networks (communication arising): Sexual contacts and epidemic thresholds. *Nature*, 423(6940):605–606, 2003.

[10] E. D. Kolaczyk and G. Csárdi. *Statistical Analysis of Network Data with R.* Use R! Springer, New York, 2014. MR3288852

[11] P. Krapivsky, G. Rodgers, and S. Redner. Degree distributions of growing networks. *Phys. Rev. Lett*, 86, 2001.

[12] P. L. Krapivsky and S. Redner. Organization of growing random networks. *Physical Review E*, 63(6):066123:1–14, 2001.

[13] R. A. Kronmal and A. V. Peterson Jr. On the alias method for generating random variables from a discrete distribution. *The American Statistician*, 33(4):214–218, 1979. MR0552783

[14] J. Kunegis. Konect: the Koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1343–1350. ACM, 2013.

[15] S. I. Resnick. *Heavy Tail Phenomena: Probabilistic and Statistical Modeling.* Springer Series in Operations Research and Financial Engineering. Springer-Verlag, New York, 2007. ISBN: 0-387-24272-4. MR2271424

[16] S. I. Resnick. *Extreme Values, Regular Variation and Point Processes.* Springer, New York, 2008. Reprint of the 1987 original. MR0900810

[17] 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. MR3377813

[18] A. Rinaldo, S. Petrović, and S. E. Fienberg. Maximum likelihood estimation in the $\beta$-model. *Ann. Statist.*, 41(3):1085–1110, 2013. MR3113804

[19] 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, March 2016. MR3540787

[20] R. Tonelli, G. Concas, and M. Locci. Three efficient algorithms for implementing the preferential attachment mechanism in Yule-Simon stochastic process. *WSEAS Trans. Inf. Sci. App*, 7, 2010.

[21] R. van der Hofstad. *Random Graphs and Complex Networks. Vol. 1.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2017. MR3617364

[22] 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.

[23] Wikipedia, mass message senders – wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Wikipedia:Mass_message_senders, 2016. Accessed: 2017-01-27.

[24] T. Yan, C. Leng, and J. Zhu. Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *Ann. Statist.*, 44(1):31–57, 2016. MR3449761