Fitting the Linear Preferential Attachment Model

Phyllis Wan[1], Tiandong Wang[2], Richard A. Davis[1], and Sidney I. Resnick[2]

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

[2]*School of Operations Research and Information Engineering, Cornell University , Ithaca, NY 14853*
e-mail: tw398@cornell.edu; 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. At the end of the paper, we also present a semi-parametric method to model heavy-tailed features of the degree distributions of the network using ideas from extreme value theory.

Keywords and phrases: power laws, multivariate heavy tail statistics, preferential attachment, regular variation, 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 [7] while more mathematical treatments are available in [2, 6, 21]. Also see [10] for a statistical survey of methods for network data.

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, 6, 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 [16, 17]) which is akin to a *joint power law*, was only recently established [18, 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 $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}})$. The first three parameters, α, β, γ, 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, δ_{in} and δ_{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 (6.5) 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 nth 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 $G(n_0), G(n_0+1), \dots, G(n_0+n)$. After deriving the likelihood function, we show that there exists a unique maximum at $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}})$ 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 $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$, $\hat{\delta}_{\text{in}}$, and $\hat{\delta}_{\text{out}}$ 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 are 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 when 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 stimulated 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 parameters) 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 inspecting the edge evolutions in separate time intervals.

Often it is difficult to believe in the existence of a true model, especially one whose parameters are unchanging over time. Perhaps a family of models containing the truth is not available to the modeler. Nevertheless, some models can capture certain salient features in the data, such as heavy-tailed properties of the degree distributions. In cases where a true model family is not apparent, maximum likelihood assuming a model may deliver misleading estimates of model parameters. An alternative to maximum likelihood is to estimate certain tail properties using semi-parametric methods from extreme value theory. For example, for a network that exhibits power-like tails, one can use the Hill estimator and familiar complementary cumulative log-log plots of the power-law indices directly. Further, applying extreme value methods, also referred to as the asymptotic method in the sequel, it is possible to estimate various dependence structures in the network data related to in- and out-degrees. We discuss this approach at the end of the paper as an alternative to maximum likelihood based on the fully evolved network and a single snapshot where an exact parametric model is assumed.

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 is available. We test these estimators against simulated data in Section 4 and then explore the Wiki talk network in Section 5. Lastly Section 6 addresses the issue of estimation using asymptotic methods and discusses a semi-parametric estimation method through the tail asymptotics of in- and out-degrees.

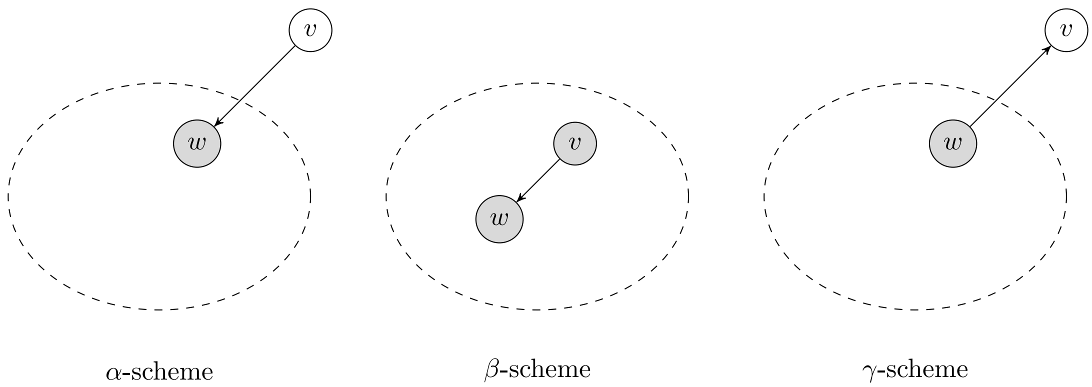
2. Model specification and simulation

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 nonnegative real numbers α, β, γ, δ_{in} and δ_{out}, where $\alpha+\beta+\gamma =$

1 and $\delta_{\text{in}}, \delta_{\text{out}} > 0$. To avoid degenerate situations, assume that each of the numbers α, β, γ 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 $u \in V(n)$, $D_{\text{in}}^{(n)}(u)$ and $D_{\text{out}}^{(n)}(u)$ denote the in- and out-degree of u respectively in $G(n)$. There are three scenarios that we call the α, β and γ-schemes, which are activated by flipping a 3-sided coin whose outcomes are $1, 2, 3$ with probabilities α, β, γ. 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 α, β, γ. Then the graph $G(n)$ is obtained from $G(n-1)$ as follows.



- If $J_n = 1$ (with probability α), 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 $G(n-1)$:

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

- If $J_n = 2$ (with probability β), 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

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

- If $J_n = 3$ (with probability γ), 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

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

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. Simulation algorithm

We describe an efficient simulation procedure for the preferential attachment network given the parameter values $(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{out})$, where $\alpha + \beta + \gamma = 1$. The simulation cost of the algorithm is linear in time. This algorithm was provided by Joyjit Roy during his graduate work at Cornell University.

Algorithm 1: Simulating a directed edge preferential attachment network

Algorithm

```
Input: α, β, δ_in, δ_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 ← n_0
while t < n do
    N(t) ← |V(t)|
    Generate U ~ Uniform(0, 1)
    if U < α then
        v^(1) ← N(t) + 1
        v^(2) ← Node_Sample(E(t), 2, δ_in)
        V(t) ← Append(V(t), N(t) + 1)
    else if α ≤ U < α + β then
        v^(1) ← Node_Sample(E(t), 1, δ_out)
        v^(2) ← Node_Sample(E(t), 2, δ_in)
    else if U ≥ α + β then
        v^(1) ← Node_Sample(E(t), 1, δ_out)
        v^(2) ← N(t) + 1
        V(t) ← Append(V(t), N(t) + 1)
    E(t + 1) ← Append(E(t), (v^(1), v^(2)))
    t ← 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; δ ∈ {δ_in, δ_out}, the offset parameter
Output: the sampled node, v
Generate W ~ Uniform(0, t + N(t)δ)
if W ≤ t then
    v ← v^(j)_⌈W⌉
else if W > t then
    v ← ⌈(W − t)/δ⌉
return v
```

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 nodes, 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. We describe a simulation algorithm in Algorithm 1 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 $W \sim \text{Uniform}(0, n-1+N(n-1)\delta_{\text{in}})$ and set

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

Then

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

which corresponds to the desired selection probability (2.1).

3. Parameter estimation

In this section, we estimate the preferential attachment parameter vector $(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$ 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. MLE based on the full network history

3.1.1. 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 $(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$, where $\delta_{\text{in}}, \delta_{\text{out}} > 0$ and α, β are non-negative probabilities. The γ 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 $\delta_{\text{in}}, \delta_{\text{out}}$ to be $[\epsilon, K]$, for some sufficiently small $\epsilon > 0$ and large K. In particular, the true value of $\delta_{\text{in}}, \delta_{\text{out}}$ is assumed to be contained in (ϵ, 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 $G(t)$;
- $D_{\text{in}}^{(t-1)}(v)$, $D_{\text{out}}^{(t-1)}(v)$, 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

$$\begin{aligned}
&L(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}} \mid G(n_0), (e_t)_{t=n_0+1}^n) \\
&= \prod_{t=n_0+1}^{n} \left(\alpha \frac{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}}{t-1+\delta_{\text{in}} N(t-1)} \right)^{\mathbf{1}_{\{J_t=1\}}}
\end{aligned}$$

$$\times \prod_{t=n_0+1}^{n} \left(\beta \left(\frac{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}}{t-1+\delta_{\text{in}}N(t-1)}\right)\left(\frac{D_{\text{out}}^{(t-1)}(v_t^{(1)}) + \delta_{\text{out}}}{t-1+\delta_{\text{out}}N(t-1)}\right)\right)^{\mathbf{1}_{\{J_t=2\}}}$$
$$\times \prod_{t=n_0+1}^{n} \left((1-\alpha-\beta)\frac{D_{\text{out}}^{(t-1)}(v_t^{(1)}) + \delta_{\text{out}}}{t-1+\delta_{\text{out}}N(t-1)}\right)^{\mathbf{1}_{\{J_t=3\}}} \tag{3.1}$$

and the log likelihood function is

$$\begin{aligned}
&\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}}|\ G(n_0),(e_t)_{t=n_0+1}^n) \qquad (3.2)\\
&= \log\alpha \sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=1\}} + \log\beta \sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=2\}} + \log(1-\alpha-\beta)\sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=3\}}\\
&\quad + \sum_{t=n_0+1}^{n}\log\left(D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta_{\text{in}}\right)\mathbf{1}_{\{J_t\in\{1,2\}\}} + \sum_{t=n_0+1}^{n}\log\left(D_{\text{out}}^{(t-1)}(v_t^{(1)})+\delta_{\text{out}}\right)\mathbf{1}_{\{J_t\in\{2,3\}\}}\\
&\quad - \sum_{t=n_0+1}^{n}\log(t-1+\delta_{\text{in}}N(t-1))\mathbf{1}_{\{J_t\in\{1,2\}\}} - \sum_{t=n_0+1}^{n}\log(t-1+\delta_{\text{out}}N(t-1))\mathbf{1}_{\{J_t\in\{2,3\}\}}.
\end{aligned}$$

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

$$\frac{\partial}{\partial\alpha}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}}|\ G(n_0),(e_t)_{t=n_0+1}^n) = \frac{1}{\alpha}\sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=1\}} - \frac{1}{1-\alpha-\beta}\sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=3\}}, \tag{3.3}$$
$$\frac{\partial}{\partial\beta}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}}|\ G(n_0),(e_t)_{t=n_0+1}^n) = \frac{1}{\beta}\sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=2\}} - \frac{1}{1-\alpha-\beta}\sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=3\}}, \tag{3.4}$$
$$\begin{aligned}
&\frac{\partial}{\partial\delta_{\text{in}}}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}}|\ G(n_0),(e_t)_{t=n_0+1}^n) \qquad (3.5)\\
&= \sum_{t=n_0+1}^{n}\frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta_{\text{in}}}\mathbf{1}_{\{J_t\in\{1,2\}\}} - \sum_{t=n_0+1}^{n}\frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)}\mathbf{1}_{\{J_t\in\{1,2\}\}},\\
&\frac{\partial}{\partial\delta_{\text{out}}}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}}|\ G(n_0),(e_t)_{t=n_0+1}^n)\\
&= \sum_{t=n_0+1}^{n}\frac{1}{D_{\text{out}}^{(t-1)}(v_t^{(1)})+\delta_{\text{out}}}\mathbf{1}_{\{J_t\in\{2,3\}\}} - \sum_{t=n_0+1}^{n}\frac{N(t-1)}{t-1+\delta_{\text{out}}N(t-1)}\mathbf{1}_{\{J_t\in\{2,3\}\}}.
\end{aligned}$$

Note that the score functions (3.3), (3.4) for α and β do not depend on δ_{in} and δ_{out}. One can show that the Hessian matrix of the log-likelihood for (α,β) is positive definite. Thus setting (3.3) and (3.4) to zero gives the unique MLE estimates for α and β.

$$\hat{\alpha}^{MLE} = \frac{1}{n-n_0}\sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=1\}}, \tag{3.6}$$
$$\hat{\beta}^{MLE} = \frac{1}{n-n_0}\sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=2\}}. \tag{3.7}$$

These estimates are strongly consistent by applying the strong law of large numbers for the $\{J_t\}$ sequence.

Next, consider the first term of the score function for δ_{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 $\left\{D_{\mathrm{in}}^{(t-1)}(v_t^{(2)}) = i, J_t \in \{1,2\}\right\}$ 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 $N_i^{\mathrm{in}}(n)$ and $N_{>i}^{\mathrm{in}}(n)$ to be the number of nodes with in-degree equal to i and greater than i, respectively, so that,

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

Then

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

On the other hand, when $i = 0$, $\left\{D_{\mathrm{in}}^{(t-1)}(v_t^{(2)}) = 0, J_t \in \{1,2\}\right\}$ 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 γ-scheme (otherwise it would have been born with in-degree 1).

This implies:

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

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

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

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

$$\begin{aligned}&\frac{1}{n-n_0} \sum_{i=0}^{\infty} \frac{N_{>i}^{\mathrm{in}}(n) - N_{>i}^{\mathrm{in}}(n_0)}{i+\delta_{\mathrm{in}}} \\ &\quad - \frac{1}{\delta_{\mathrm{in}}(n-n_0)} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}} - \frac{1}{n-n_0} \sum_{t=n_0+1}^{n} \frac{N(t-1)}{t-1+\delta_{\mathrm{in}} N(t-1)} \mathbf{1}_{\{J_t \in \{1,2\}\}} = 0, \end{aligned} \tag{3.9}$$

where the only unknown parameter is δ_{in}. In Section 3.1.2, we show that the solution to (3.9) actually maximizes the likelihood function in δ_{in}. Similarly, the MLE for δ_{out} can be solved from

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

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

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

$$
\begin{aligned}
&L(\alpha,\beta,\delta_{in},\delta_{out} \mid G(n_0),(e_t)_{t=n_0+1}^n) \\
&= \alpha^{\sum_{t=n_0+1}^n \boldsymbol{1}_{\{J_t=1\}}}\ \beta^{\sum_{t=n_0+1}^n \boldsymbol{1}_{\{J_t=2\}}}\ (1-\alpha-\beta)^{\sum_{t=n_0+1}^n \boldsymbol{1}_{\{J_t=3\}}} \\
&\quad \times \prod_{t=n_0+1}^n (t-1+\delta_{in}N(t-1))^{-\boldsymbol{1}_{\{J_t\in\{1,2\}\}}}\ (t-1+\delta_{out}N(t-1))^{-\boldsymbol{1}_{\{J_t\in\{2,3\}\}}} \\
&\quad \times \prod_{t=n_0+1}^n \left[\prod_{i=0}^\infty (i+\delta_{in})^{\boldsymbol{1}_{\left\{D_{in}^{(t-1)}(v_t^{(2)})=i,\,J_t\in\{1,2\}\right\}}} \prod_{j=0}^\infty (j+\delta_{out})^{\boldsymbol{1}_{\left\{D_{out}^{(t-1)}(v_t^{(1)})=j,\,J_t\in\{2,3\}\right\}}}\right] \\
&= \alpha^{\sum_{t=n_0+1}^n \boldsymbol{1}_{\{J_t=1\}}}\ \beta^{\sum_{t=n_0+1}^n \boldsymbol{1}_{\{J_t=2\}}}\ (1-\alpha-\beta)^{\sum_{t=n_0+1}^n \boldsymbol{1}_{\{J_t=3\}}} \\
&\quad \times \prod_{t=n_0+1}^n (t-1+\delta_{in}N(t-1))^{-\boldsymbol{1}_{\{J_t\in\{1,2\}\}}}\ (t-1+\delta_{out}N(t-1))^{-\boldsymbol{1}_{\{J_t\in\{2,3\}\}}}\ \delta_{in}^{-\boldsymbol{1}_{\{J_t=3\}}}\ \delta_{out}^{-\boldsymbol{1}_{\{J_t=1\}}} \\
&\quad \times \prod_{i=0}^\infty (i+\delta_{in})^{N_{>i}^{in}(n)-N_{>i}^{in}(n_0)} \prod_{j=0}^\infty (j+\delta_{out})^{N_{>j}^{out}(n)-N_{>j}^{out}(n_0)}.
\end{aligned}
$$

Hence by the factorization theorem, $N(n_0)$, $(J_t)_{t=n_0+1}^n$, $(N_{>i}^{in}(n)-N_{>i}^{in}(n_0))_{i\ge0}$, $(N_{>j}^{out}(n)-N_{>j}^{out}(n_0))_{j\ge0}$ *are sufficient statistics for* $(\alpha,\beta,\delta_{in},\delta_{out})$.

3.1.2. Consistency of MLE

We remarked after (3.6) and (3.7) that $\hat\alpha^{MLE}$ and $\hat\beta^{MLE}$ converge almost surely to α and β. We now prove that the MLE of $(\delta_{\rm in},\delta_{\rm out})$ 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 $i,j\ge0$,

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

and $(n-n_0)/n \to 1$. In other words, n_0, $N_{>i}^{\rm in}(n_0)$, $N_{>j}^{\rm out}(n_0)$ 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 $N(0)\ge1$. 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 $\delta_{\rm in}$ and $\delta_{\rm out}$ respectively, normalized by $1/n$, i.e.

$$
\begin{aligned}
\Psi_n(\lambda) &:= \sum_{i=0}^\infty \frac{N_{>i}^{\rm in}(n)}{n}\log(i+\lambda) - \frac{\log\lambda}{n}\sum_{t=1}^n \mathbf{1}_{\{J_t=3\}} - \frac{1}{n}\sum_{t=1}^n \log\left(t-1+\lambda N(t-1)\right)\mathbf{1}_{\{J_t\in\{1,2\}\}}, \\
\Phi_n(\mu) &:= \sum_{j=0}^\infty \frac{N_{>j}^{\rm out}(n)}{n}\log(j+\mu) - \frac{\log\mu}{n}\sum_{t=1}^n \mathbf{1}_{\{J_t=1\}} - \frac{1}{n}\sum_{t=1}^n \log\left(t-1+\mu N(t-1)\right)\mathbf{1}_{\{J_t\in\{2,3\}\}}.
\end{aligned}
$$

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

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

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

These are the MLE estimators of δ_{in},δ_{out} *and they are strongly consistent; that is, as* $n\to\infty$*,*

$$
\hat\delta_{in}^{MLE} \xrightarrow{a.s.} \delta_{in}, \qquad \hat\delta_{out}^{MLE} \xrightarrow{a.s.} \delta_{out}.
$$

We only verify the consistency of $\hat{\delta}_{\text{in}}^{MLE}$ and similar arguments apply to $\hat{\delta}_{\text{out}}^{MLE}$. Before the proof, we give some preliminaries. Define

$$\psi_n(\lambda) := \Psi_n'(\lambda) = \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n)/n}{i+\lambda} - \frac{\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=3\}}}{\lambda} - \frac{1}{n}\sum_{t=1}^{n}\left(\frac{N(t-1)}{t-1+\lambda N(t-1)}\mathbf{1}_{\{J_t\in\{1,2\}\}}.\right. \tag{3.10}$$

From [3], there exists a proper probability distribution $\{f_{ij}\}$ such that almost surely

$$\frac{N_{ij}(n)}{N(n)} \to f_{ij} =: \frac{p_{ij}}{1-\beta}, \quad (n\to\infty). \tag{3.11}$$

Define

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

and denote $p_i^{\text{in}} := \sum_j p_{ij}$. From [3, Equation (3.10)],

$$\begin{aligned} p_0^{\text{in}} &= \frac{\alpha}{1+a_1(\delta_{\text{in}})\delta_{\text{in}}}, \\ p_i^{\text{in}} &= \frac{\Gamma(i+\delta_{\text{in}})\Gamma(1+\delta_{\text{in}}+a_1(\delta_{\text{in}})^{-1})}{\Gamma(i+1+\delta_{\text{in}}+a_1(\delta_{\text{in}})^{-1})\Gamma(1+\delta_{\text{in}})}\left(\frac{\alpha\delta_{\text{in}}}{1+a_1(\delta_{\text{in}})\delta_{\text{in}}}+\frac{\gamma}{a_1(\delta_{\text{in}})}\right)\left(\quad i\ge 1.\right. \end{aligned}$$

We write $p_i^{\text{in}}(\delta_{\text{in}})$ to emphasize the dependence on δ_{in}. Define $p_{>i}^{\text{in}}(\delta_{\text{in}}) := \sum_{k>i} p_k^{\text{in}}(\delta_{\text{in}})$ and

$$\psi(\lambda) := \sum_{i=0}^{\infty}\left(\frac{p_{>i}^{\text{in}}(\delta_{\text{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - (1-\beta)a_1(\lambda).\right. \tag{3.12}$$

Here is an outline of the proof. We think of ψ in (3.12) as a limit version of ψ_n given in (3.10). Lemma 3.3 shows that $\psi(\cdot)$ has a unique zero at δ_{in}, and $\psi(\lambda)$ is positive to the left of δ_{in} and negative to the right of δ_{in}. Then Lemma 3.4 shows that $\sup_{\lambda\ge\epsilon}|\psi_n(\lambda)-\psi(\lambda)|\to 0$ almost surely. The result of the theorem follows by a straightforward argument.

Lemma 3.3. *For $\lambda > 0$, the function $\psi(\lambda)$ in (3.12) has a unique zero at δ_{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]):

$$p_0^{\text{in}}(\lambda)\left(\lambda+\frac{1}{a_1(\lambda)}\right)\left(= \frac{\alpha}{a_1(\lambda)},\right. \tag{3.13a}$$

$$p_1^{\text{in}}(\lambda)\left(1+\lambda+\frac{1}{a_1(\lambda)}\right)\left(= \lambda p_0^{\text{in}}(\lambda)+\frac{\gamma}{a_1(\lambda)},\right. \tag{3.13b}$$

$$p_2^{\text{in}}(\lambda)\left(2+\lambda+\frac{1}{a_1(\lambda)}\right)\left(= (1+\lambda)p_1^{\text{in}}(\lambda),\right. \tag{3.13c}$$

$$\vdots$$

$$p_i^{\text{in}}(\lambda)\left(i+\lambda+\frac{1}{a_1(\lambda)}\right)\left(= (i-1+\lambda)p_{i-1}^{\text{in}}(\lambda), \quad (i\ge 2).\right. \tag{3.13d}$$

Summing the recursions in (3.13) from 0 to i, we get (with the convention that $\sum_{i=0}^{-1} = 0$)

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

which can be simplified to

$$\frac{1}{a_1(\lambda)}\sum_{k=0}^{i} p_k^{\text{in}}(\lambda) + (i+\lambda)p_i^{\text{in}}(\lambda) = \frac{1-\beta}{a_1(\lambda)} - \frac{\gamma}{a_1(\lambda)}\mathbf{1}_{\{i=0\}}, \quad i \geq 0. \tag{3.14}$$

From (3.11),

$$\sum_{i=0}^{\infty} p_i^{\text{in}}(\lambda) = \sum_{i,j} p_{ij}(\lambda) = 1-\beta. \tag{3.15}$$

Hence by rearranging (3.14), 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,

$$p_{>i}^{\text{in}}(\lambda) = a_1(\lambda)(i+\lambda)p_i^{\text{in}}(\lambda) + \gamma\mathbf{1}_{\{i=0\}}. \tag{3.16}$$

Now with the help of (3.15) and (3.16), we can rewrite $\psi(\lambda)$ in the following way:

$$\begin{aligned}
\psi(\lambda) &= \sum_{i=0}^{\infty}\frac{p_{>i}^{\text{in}}(\delta_{\text{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - (1-\beta)a_1(\lambda) \\
&= \sum_{i=0}^{\infty}\frac{p_{>i}^{\text{in}}(\delta_{\text{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - \sum_{i=0}^{\infty}\frac{p_i^{\text{in}}(\delta_{\text{in}})a_1(\lambda)(i+\lambda)}{i+\lambda} \\
&= \sum_{i=0}^{\infty}\frac{a_1(\delta_{\text{in}})(i+\delta_{\text{in}})p_i^{\text{in}}(\delta_{\text{in}}) + \gamma\mathbf{1}_{\{i=0\}}}{i+\lambda} - \frac{\gamma}{\lambda} - \sum_{i=0}^{\infty}\frac{p_i^{\text{in}}(\delta_{\text{in}})a_1(\lambda)(i+\lambda)}{i+\lambda} \\
&= \sum_{i=0}^{\infty}\frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}\Big(a_1(\delta_{\text{in}})(i+\delta_{\text{in}}) - a_1(\lambda)(i+\lambda)\Big) \\
&= \sum_{i=0}^{\infty}\frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}\int_{\lambda}^{\delta_{\text{in}}}\frac{\partial}{\partial s}\Big(a_1(s)(i+s)\Big)ds \\
&= \sum_{i=0}^{\infty}\frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}\int_{\lambda}^{\delta_{\text{in}}}\frac{(\alpha+\beta)(1-i(1-\beta))}{(1+s(1-\beta))^2}ds \\
&= \left(\sum_{i=0}^{\infty}\frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}(1-i(1-\beta))\right)\int_{\lambda}^{\delta_{\text{in}}}\frac{\alpha+\beta}{(1+s(1-\beta))^2}ds \\
&=: C(\lambda)\int_{\lambda}^{\delta_{\text{in}}}\frac{\alpha+\beta}{(1+s(1-\beta))^2}ds.
\end{aligned} \tag{3.17}$$

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 (3.16), we get by monotone convergence

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

The infinite series converge because $p_i^{\text{in}}(\lambda)$ is a power law with index greater than 2; see (6.3). Solving for the infinite series we get

$$\sum_{i=0}^{\infty} i p_i^{\text{in}}(\lambda) = \frac{a_1(\lambda)\lambda}{1-a_1(\lambda)}(1-\beta) + \frac{\gamma}{1-a_1(\lambda)} = 1. \tag{3.18}$$

Hence we have

$$\begin{aligned} C(\lambda) &= \sum_{i\le(1-\beta)^{-1}} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}(1-i(1-\beta)) - \sum_{i>(1-\beta)^{-1}} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}(i(1-\beta)-1) \\ &> \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{(1-\beta)^{-1}+\lambda}(1-i(1-\beta)) \\ &= \frac{1}{(1-\beta)^{-1}+\lambda}\sum_{i=0}^{\infty} p_i^{\text{in}}(\delta_{\text{in}}) - \frac{1-\beta}{(1-\beta)^{-1}+\lambda}\sum_{i=0}^{\infty} i p_i^{\text{in}}(\delta_{\text{in}}) \\ &= \frac{1}{(1-\beta)^{-1}+\lambda}(1-\beta) - \frac{1-\beta}{(1-\beta)^{-1}+\lambda}1 \\ &= 0. \end{aligned}$$

Now recall from (3.17) 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 δ_{in} and $\psi(\lambda) > 0$ when $\lambda < \delta_{\text{in}}$ and $\psi(\lambda) < 0$ when $\lambda > \delta_{\text{in}}$. □

We show the uniform convergence of ψ_n to ψ in the next lemma.

Lemma 3.4. *As $n \to \infty$, for any $\epsilon > 0$,*

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

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

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

Thus,

$$\begin{aligned} \sup_{\lambda\ge\epsilon}|\psi_n(\lambda)-\psi(\lambda)| \le &\sup_{\lambda\ge\epsilon}\sum_{i=0}^{\infty}\frac{\left|N_{>i}^{\text{in}}(n)/n - p_{>i}^{\text{in}}\right|}{i+\lambda} + \sup_{\lambda\ge\epsilon}\frac{1}{\lambda}\left|\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t=3\}} - (1-\alpha-\beta)\right| \\ &+ \sup_{\lambda\ge\epsilon}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{N(t-1)}{t-1+\lambda N(t-1)}\mathbf{1}_{\{J_t\in\{1,2\}\}} - \frac{(1-\beta)(\alpha+\beta)}{1+\lambda(1-\beta)}\right|. \end{aligned} \tag{3.19}$$

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

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

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

$$\sum_{i=0}^{\infty} \frac{\left|N^{\text{in}}_{>i}(n)/n - p^{\text{in}}_{>i}\right|}{i+\lambda} \le \sum_{i=0}^{M} \frac{\left|N^{\text{in}}_{>i}(n)/n - p^{\text{in}}_{>i}\right|}{i+\lambda} + \sum_{i=M+1}^{\infty} \frac{1/i}{i+\lambda} + \sum_{i=M+1}^{\infty} \frac{p^{\text{in}}_{>i}}{i+\lambda}.$$

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 (3.11), we have

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

Hence, for any fixed M,

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

which implies further that choosing M arbitrarily large gives

$$\sup_{\lambda\ge\epsilon} \sum_{i=0}^{\infty} \frac{\left|N^{\text{in}}_{>i}(n)/n - p^{\text{in}}_{>i}\right|}{i+\lambda} \le \sum_{i=0}^{M} \frac{\left|N^{\text{in}}_{>i}(n)/n - p^{\text{in}}_{>i}\right|}{i+\epsilon} + \sum_{i=M+1}^{\infty} \frac{1/i}{i+\epsilon} + \sum_{i=M+1}^{\infty} \frac{p^{\text{in}}_{>i}}{i+\epsilon} \xrightarrow{\text{a.s.}} 0.$$

The second term in (3.19) converges to 0 almost surely by strong law of large numbers, and the third term in (3.19) can be written as

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

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

$$\begin{aligned}\sup_{\lambda\ge\epsilon} &\left| \frac{1}{n} \sum_{t=1}^{n} \frac{N(t-1)}{t-1+\lambda N(t-1)} - \frac{(1-\beta)}{1+\lambda(1-\beta)} \right| \\ &= \sup_{\lambda\ge\epsilon} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{N(t-1)/(t-1) - (1-\beta)}{(1+\lambda N(t-1)/(t-1))(1+\lambda(1-\beta))} \right| \\ &\le \frac{1}{n} \sum_{t=1}^{n} \left| \frac{N(t-1)/(t-1) - (1-\beta)}{(1+\epsilon N(t-1)/(t-1))(1+\epsilon(1-\beta))} \right|,\end{aligned}$$

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,

$$\begin{aligned}
&\sup_{\lambda\ge\epsilon}\frac{1-\beta}{1+\lambda(1-\beta)}\left|\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t\in\{1,2\}\}}-(\alpha+\beta)\right| \\
&\quad\le \frac{1-\beta}{1+\epsilon(1-\beta)}\left|\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t\in\{1,2\}\}}-(\alpha+\beta)\right| \xrightarrow{\text{a.s.}} 0, \quad \text{as } n\to\infty.
\end{aligned}$$

Hence the third term of (3.19) also goes to 0 almost surely as $n\to\infty$. The result of the lemma follows. □

We are now ready to establish the consistency for $\hat{\delta}_{\text{in}}^{MLE}$.

Proof of Theorem 3.2. From Lemma 3.3 and the fact that ψ 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 Lemma 3.4,

$$\mathbf{P}\left(\exists N_\kappa \ s.t. \sup_{n>N_\kappa}\sup_{\lambda\in[\epsilon,K]}|\psi_n(\lambda)-\psi(\lambda)|<\varepsilon_\kappa/2\right)=1. \tag{3.20}$$

Note that Lemma 3.4 implies $\sup_{\lambda\in[\epsilon,K]}|\psi_n(\lambda)-\psi(\lambda)|<\varepsilon_\kappa/2$, which further gives

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

and

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

These jointly indicate that $\delta_{\text{in}}-\kappa\le\hat{\delta}_{\text{in}}^{MLE}\le\delta_{\text{in}}+\kappa$. Hence (3.20) implies

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

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

3.1.3. 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 3.5. *Let $\hat{\boldsymbol{\theta}}_n^{MLE}$ be the MLE estimator for $\boldsymbol{\theta}$, the parameter vector of the preferential attachment model. Then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{MLE}-\boldsymbol{\theta})\xrightarrow{d}N\left(\mathbf{0},\Sigma(\boldsymbol{\theta})\right), \tag{3.21}$$

where

$$\Sigma^{-1}(\boldsymbol{\theta})=I(\boldsymbol{\theta})=\begin{bmatrix}\frac{1-\beta}{\alpha(1-\alpha-\beta)} & \frac{1}{1-\alpha-\beta} & 0 & 0\\ \frac{1}{1-\alpha-\beta} & \frac{1-\alpha}{\beta(1-\alpha-\beta)} & 0 & 0\\ 0 & 0 & I_{in} & 0\\ 0 & 0 & 0 & I_{out}\end{bmatrix}, \tag{3.22}$$

with

$$I_{in}=\sum_{i=0}^{\infty}\frac{p_{>i}^{in}}{(i+\delta_{in})^2}-\frac{\gamma}{\delta_{in}^2}-\frac{(\alpha+\beta)(1-\beta)^2}{\left(1+\delta_{in}(1-\beta)\right)^2}, \tag{3.23}$$

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

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

Remark 3.6. *From Theorem 3.5, the estimators $(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE})$, $\hat{\delta}_{in}^{MLE}$, and $\hat{\delta}_{out}^{MLE}$ are asymptotically independent.*

Proof. We first show the limiting distributions for $(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE})$, $\hat{\delta}_{\text{in}}^{MLE}$, and $\hat{\delta}_{\text{out}}^{MLE}$. The joint asymptotic normality is illustrated at the end of the proof.

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 $\{J_t\}$ 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.

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), the score function for δ_{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

$$u_t(\delta) := \frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta} \mathbf{1}_{\{J_t \in \{1,2\}\}} - \frac{N(t-1)}{t-1+\delta N(t-1)} \mathbf{1}_{\{J_t \in \{1,2\}\}}. \tag{3.24}$$

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 $\sum_{t=1}^{n} u_t(\delta)$,

$$0 = \sum_{t=1}^{n} u_t(\hat{\delta}_{\text{in}}^{MLE}) = \sum_{t=1}^{n} u_t(\delta_{\text{in}}) + (\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}}) \sum_{t=1}^{n} \dot{u}_t(\hat{\delta}_{\text{in}}^{*}), \tag{3.25}$$

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 (3.25) 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_{in} is as defined in (3.23), 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 3.7 and 3.8, respectively.

Lemma 3.7. *As $n \to \infty$,*

$$n^{-1/2} \sum_{t=1}^{n} u_t(\delta_{in}) \xrightarrow{d} N(0, I_{in}). \tag{3.26}$$

Proof. Let $\mathcal{F}_n = \sigma(G(0), \dots, G(n))$ be the σ-field generated by the information contained in the graphs. We first observe that $\{\sum_{t=1}^n u_t(\delta_{\text{in}}), \mathcal{F}_n, n \geq 1\}$ is a martingale. To see this, note from (3.24) that $|u_t(\delta)| \leq 2/\delta$ and

$$\begin{aligned}
&\mathbf{E}[u_t(\delta_{\text{in}})|\mathcal{F}_{t-1}] \\
&= \mathbf{E}\left[\frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} \mathbf{1}_{\{J_t \in \{1,2\}\}} \middle| \mathcal{F}_{t-1} \right] - \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \mathbf{E}[\mathbf{1}_{\{J_t \in \{1,2\}\}} | \mathcal{F}_{t-1}] \\
&= \mathbf{E}\left[\frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} \middle| J_t = 1, \mathcal{F}_{t-1} \right] \mathbf{P}[J_t = 1] \\
&\quad + \mathbf{E}\left[\frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} \middle| J_t = 2, \mathcal{F}_{t-1} \right] \mathbf{P}[J_t = 2] - (\alpha+\beta) \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \\
&= (\alpha+\beta) \sum_{v \in V_{t-1}} \frac{1}{D_{\text{in}}^{(t-1)}(v) + \delta_{\text{in}}} \frac{D_{\text{in}}^{(t-1)}(v) + \delta_{\text{in}}}{t-1+\delta_{\text{in}} N(t-1)} - (\alpha+\beta) \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \\
&= (\alpha+\beta) \left(\sum_{v \in V_{t-1}} \frac{1}{t-1+\delta_{\text{in}} N(t-1)} - \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \right) \\
&= 0,
\end{aligned}$$

which satisfies the definition of a martingale difference. Hence $\left\{ n^{-1/2} \sum_{r=1}^t u_r(\delta_{\text{in}}) \right\}_{t=1,\dots,n}$ is a zero-mean, square-integrable martingale array. The convergence (3.26) follows from the martingale central limit theory (cf. Theorem 3.2 of [8]) 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

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^n u_t^2(\delta_{\text{in}}) &= \frac{1}{n} \sum_{t=1}^n \mathbf{1}_{\{J_t \in \{1,2\}\}} \left(\frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} - \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \right)^2 \\
&= \frac{1}{n} \sum_{t=1}^n \frac{\mathbf{1}_{\{J_t \in \{1,2\}\}}}{\left(D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}} \right)^2} - \frac{2}{n} \sum_{t=1}^n \frac{\mathbf{1}_{\{J_t \in \{1,2\}\}}}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \\
&\quad + \frac{1}{n} \sum_{t=1}^n \mathbf{1}_{\{J_t \in \{1,2\}\}} \left(\frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \right)^2 \\
&=: T_1 - 2T_2 + T_3.
\end{aligned}$$

Following the calculations in the proof of Lemma 3.4, we have for T_1,

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

We then rewrite T_2 as

$$\begin{aligned} T_2 &= \frac{1}{n}\sum_{t=1}^{n} \frac{\mathbf{1}_{\{J_t\in\{1,2\}\}}}{D_{\mathrm{in}}^{(t-1)}(v_t^{(2)})+\delta_{\mathrm{in}}} \left(\frac{N(t-1)/(t-1)}{1+\delta_{\mathrm{in}}N(t-1)/(t-1)} - \frac{1-\beta}{1+\delta_{\mathrm{in}}(1-\beta)} \right) \\ &\quad + \frac{1}{n}\sum_{t=1}^{n} \frac{\mathbf{1}_{\{J_t\in\{1,2\}\}}}{D_{\mathrm{in}}^{(t-1)}(v_t^{(2)})+\delta_{\mathrm{in}}} \frac{1-\beta}{1+\delta_{\mathrm{in}}(1-\beta)} \\ &=:\ T_{21}+T_{22}, \end{aligned}$$

where

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

by Cesàro's convergence and

$$\begin{aligned} T_{22} &= \frac{1-\beta}{1+\delta_{\mathrm{in}}(1-\beta)} \left(\sum_{i=0}^{\infty} \frac{N_{>i}^{\mathrm{in}}(n)/n}{i+\delta_{\mathrm{in}}} - \frac{1}{\delta_{\mathrm{in}}}\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t=3\}} \right) \\ &\xrightarrow{p} \frac{1-\beta}{1+\delta_{\mathrm{in}}(1-\beta)} \left(\sum_{i=0}^{\infty} \frac{p_{>i}^{\mathrm{in}}}{i+\delta_{\mathrm{in}}} - \frac{\gamma}{\delta_{\mathrm{in}}} \right) = \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{\mathrm{in}}(1-\beta))^2}, \end{aligned}$$

where the equality follows from (3.16). For T_3, similar to T_1, we have

$$\begin{aligned} T_3 =& \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t\in\{1,2\}\}} \left(\left(\frac{N(t-1)/(t-1)}{1+\delta_{\mathrm{in}}N(t-1)/(t-1)} \right)^2 - \frac{(1-\beta)^2}{(1+\delta_{\mathrm{in}}(1-\beta))^2} \right) \\ &+ \frac{(1-\beta)^2}{(1+\delta_{\mathrm{in}}(1-\beta))^2}\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t\in\{1,2\}\}} \xrightarrow{p} \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{\mathrm{in}}(1-\beta))^2}. \end{aligned}$$

Combining these results together,

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n} u_t^2(\delta_{\mathrm{in}}) &= T_1 - 2(T_{21}+T_{22}) + T_3 \\ &\xrightarrow{p} \sum_{i=0}^{\infty} \frac{p_{>i}^{\mathrm{in}}}{(i+\delta_{\mathrm{in}})^2} - \frac{\gamma}{\delta_{\mathrm{in}}^2} - \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{\mathrm{in}}(1-\beta))^2} = \ I_{\mathrm{in}}. \end{aligned} \tag{3.27}$$

This completes the proof. □

Lemma 3.8. *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

$$\frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\delta_{\mathrm{in}}) \xrightarrow{p} -I_{\mathrm{in}} \tag{3.28}$$

and then

$$\left|\frac{1}{n}\sum_{t=1}^{n}\dot{u}_t(\hat{\delta}^*_{\text{in}}) - \frac{1}{n}\sum_{t=1}^{n}\dot{u}_t(\delta_{\text{in}})\right| \xrightarrow{p} 0. \tag{3.29}$$

We first observe that

$$\begin{aligned}\dot{u}_t(\delta) &= -\left(\frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta}\right)^2 \mathbf{1}_{\{J_t\in\{1,2\}\}} + \left(\frac{N(t-1)}{t-1+\delta N(t-1)}\right)^2 \mathbf{1}_{\{J_t\in\{1,2\}\}} \\ &= -u_t^2(\delta) - 2u_t(\delta)\frac{N(t-1)}{t-1+\delta N(t-1)}.\end{aligned}$$

Recall the definition and convergence result for T_2 and T_3 in Lemma 3.7, 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 (3.27),

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

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 (3.28) is established.

By construction and definition, we have $\hat{\delta}_{\text{in}}, \hat{\delta}^*_{\text{in}}, \delta_{\text{in}} > 0$. To prove (3.29), note that

$$\begin{aligned}|u_t(\hat{\delta}^*_{\text{in}}) - u_t(\delta_{\text{in}})| &\le \mathbf{1}_{\{J_t\in\{1,2\}\}}\left|\frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)})+\hat{\delta}^*_{\text{in}}} - \frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta_{\text{in}}}\right| \\ &\quad + \mathbf{1}_{\{J_t\in\{1,2\}\}}\left|\frac{N(t-1)}{t-1+\hat{\delta}^*_{\text{in}}N(t-1)} - \frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)}\right| \\ &= \mathbf{1}_{\{J_t\in\{1,2\}\}}\left|\frac{\delta_{\text{in}}-\hat{\delta}^*_{\text{in}}}{\left(D_{\text{in}}^{(t-1)}(v_t^{(2)})+\hat{\delta}^*_{\text{in}}\right)\left(D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta_{\text{in}}\right)}\right| \\ &\quad + \mathbf{1}_{\{J_t\in\{1,2\}\}}\left|\frac{(N(t-1))^2(\delta_{\text{in}}-\hat{\delta}^*_{\text{in}})}{\left(t-1+\hat{\delta}^*_{\text{in}}N(t-1)\right)(t-1+\delta_{\text{in}}N(t-1))}\right| \\ &\le \frac{2|\hat{\delta}^*_{\text{in}}-\delta_{\text{in}}|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}}.\end{aligned}$$

Then

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

and

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

$$
\begin{aligned}
&\leq \left|u_t(\hat{\delta}_{\mathrm{in}}^*) - u_t(\delta_{\mathrm{in}})\right| \frac{\frac{N(t-1)}{t-1}}{1+\delta_{\mathrm{in}}\frac{N(t-1)}{t-1}} + \left|u_t(\hat{\delta}_{\mathrm{in}}^*)\right| \left| \frac{\frac{N(t-1)}{t-1}}{1+\hat{\delta}_{\mathrm{in}}^*\frac{N(t-1)}{t-1}} - \frac{\frac{N(t-1)}{t-1}}{1+\delta_{\mathrm{in}}\frac{N(t-1)}{t-1}} \right| \\
&\leq \frac{2\left|\hat{\delta}_{\mathrm{in}}^* - \delta_{\mathrm{in}}\right|}{\hat{\delta}_{\mathrm{in}}^*\delta_{\mathrm{in}}} \frac{1}{\delta_{\mathrm{in}}} + \frac{2}{\hat{\delta}_{\mathrm{in}}^*} \frac{\left|\hat{\delta}_{\mathrm{in}}^* - \delta_{\mathrm{in}}\right|}{\hat{\delta}_{\mathrm{in}}^*\delta_{\mathrm{in}}}.
\end{aligned}
$$

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

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

We have

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\hat{\delta}_{\mathrm{in}}^*) - \frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\delta_{\mathrm{in}})\right| \\
&\quad \leq \frac{1}{n}\sum_{t=1}^{n}\left|\dot{u}_t(\hat{\delta}_{\mathrm{in}}^*) - \dot{u}_t(\delta_{\mathrm{in}})\right| \leq \frac{1}{n}\sum_{t=1}^{n}\left|u_t^2(\hat{\delta}_{\mathrm{in}}^*) - u_t^2(\delta_{\mathrm{in}})\right| \\
&\qquad + \frac{2}{n}\sum_{t=1}^{n}\left|u_t(\hat{\delta}_{\mathrm{in}}^*)\frac{N(t-1)}{t-1+\hat{\delta}_{\mathrm{in}}^* N(t-1)} - u_t(\delta_{\mathrm{in}})\frac{N(t-1)}{t-1+\delta_{\mathrm{in}} N(t-1)}\right| \\
&\quad \leq \frac{2\left|\hat{\delta}_{\mathrm{in}}^* - \delta_{\mathrm{in}}\right|}{\hat{\delta}_{\mathrm{in}}^*\delta_{\mathrm{in}}}\left(\frac{2}{\hat{\delta}_{\mathrm{in}}^*} + \frac{2}{\delta_{\mathrm{in}}}\right) + \frac{4\left|\hat{\delta}_{\mathrm{in}}^* - \delta_{\mathrm{in}}\right|}{\hat{\delta}_{\mathrm{in}}^*\delta_{\mathrm{in}}}\frac{1}{\delta_{\mathrm{in}}} + \frac{4}{\hat{\delta}_{\mathrm{in}}^*}\frac{\left|\hat{\delta}_{\mathrm{in}}^* - \delta_{\mathrm{in}}\right|}{\hat{\delta}_{\mathrm{in}}^*\delta_{\mathrm{in}}} \xrightarrow{p} 0.
\end{aligned}
$$

This proves (3.29) and completes the proof of Lemma 3.8. □

We now establish the joint asymptotic normality of the MLE estimator $\hat{\boldsymbol{\theta}}_n^{MLE}$. Denote the joint score function vector for $\boldsymbol{\theta}$ by

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

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

$$
\mathbf{0} = \mathbf{S}_n\left(\hat{\boldsymbol{\theta}}_n^{MLE}\right) = \mathbf{S}_n(\boldsymbol{\theta}) + \dot{\mathbf{S}}_n\left(\hat{\boldsymbol{\theta}}_n^{*}\right)\left(\hat{\boldsymbol{\theta}}_n^{MLE} - \boldsymbol{\theta}\right), \tag{3.30}
$$

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

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

Hence

$$
\frac{1}{n}\dot{\mathbf{S}}_n(\hat{\boldsymbol{\theta}}_n^{*}) = \begin{bmatrix}
\frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial\alpha^2} & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial\alpha\partial\beta} & 0 & 0 \\
\frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial\beta\partial\alpha} & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial\beta^2} & 0 & 0 \\
0 & 0 & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial\delta_{\mathrm{in}}^2} & 0 \\
0 & 0 & 0 & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial\delta_{\mathrm{out}}^2}
\end{bmatrix} \xrightarrow{p} I(\boldsymbol{\theta}) \tag{3.31}
$$

as implied in the previous part of the proof, where $I(\boldsymbol{\theta})$ is as defined in (3.22) and 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 (3.30) and (3.31). □

3.2. Estimation based on one snapshot

Based only on the single snapshot $G(n)$, we propose a parameter estimation procedure. Since no information on the initial graph $G(n_0)$ is available, we assume that n_0 and $N(n_0)$ are negligible compared to n and $N(n)$.

Among the sufficient statistics for $(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$ derived in Remark 3.1, $\left(N^{\text{in}}_{>i}(n)\right)_{i\geq 0}$, $\left(N^{\text{out}}_{>j}(n)\right)_{j\geq 0}$ 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 3.4:

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

and

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

Replacing them in (3.9), we estimate δ_{in} in terms of α and β by solving

$$\sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{i+\delta_{\text{in}}} - \frac{1-\alpha-\beta}{\delta_{\text{in}}} - \frac{(\alpha+\beta)(1-\beta)}{1+(1-\beta)\delta_{\text{in}}} = 0. \tag{3.32}$$

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

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

To obtain an estimate for α, we make use of the recursive formula for $\{p^{\text{in}}_i\}$ in (3.13a):

$$\left(1 + \frac{(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}}\right) p^{\text{in}}_0 = \alpha, \tag{3.33}$$

and replace p^{in}_0 by $N^{\text{in}}_0(n)/n$ for large n,

$$\left(1 + \frac{(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}}\right) \frac{N^{\text{in}}_0(n)}{n} = \alpha. \tag{3.34}$$

Plug the strongly consistent estimator $\tilde{\beta}$ into (3.32) and (3.34), and we now show that solving the system of equations:

$$\sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n)/n}{i+\delta_{\text{in}}} - \frac{1-\alpha-\tilde{\beta}}{\delta_{\text{in}}} - \frac{(\alpha+\tilde{\beta})(1-\tilde{\beta})}{1+(1-\tilde{\beta})\delta_{\text{in}}} = 0, \tag{3.35a}$$

$$\left(1+\frac{(\alpha+\tilde{\beta})\delta_{\text{in}}}{1+(1-\tilde{\beta})\delta_{\text{in}}}\right)\frac{N_0^{\text{in}}(n)}{n} = \alpha, \tag{3.35b}$$

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

Theorem 3.9. *The solution* $(\tilde{\alpha}, \tilde{\delta}_{in})$ *to the system of equations in* (3.35) *is unique and strongly consistent for* (α, δ_{in})*, i.e.*

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

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 (3.35a) as

$$\begin{aligned}
\alpha+\tilde{\beta} &= \left(\frac{1}{\delta_{\text{in}}} - \sum_{i=0}^{\infty}\frac{N_{>i}^{\text{in}}(n)/n}{i+\delta_{\text{in}}}\right) \Big/ \left(\frac{1}{\delta_{\text{in}}} - \frac{1-\tilde{\beta}}{1+\delta_{\text{in}}(1-\tilde{\beta})}\right) \\
&= \left(\sum_{i=0}^{\infty}\frac{N_{>i}^{\text{in}}(n)/n}{\delta_{\text{in}}} - \sum_{i=0}^{\infty}\frac{N_{>i}^{\text{in}}(n)/n}{i+\delta_{\text{in}}}\right) \Big/ \left(\frac{1}{\delta_{\text{in}}(1+\delta_{\text{in}}(1-\tilde{\beta}))}\right) \\
&= \sum_{i=1}^{\infty}\frac{N_{>i}^{\text{in}}(n)}{n}\frac{i}{i+\delta_{\text{in}}}\left(1+\delta_{\text{in}}(1-\tilde{\beta})\right) =: f_n(\delta_{\text{in}}),
\end{aligned} \tag{3.36}$$

and (3.35b) 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:

$$\begin{aligned}
f(\delta) &:= \sum_{i=1}^{\infty} p_{>i}^{\text{in}} \frac{i}{i+\delta}(1+\delta(1-\beta)), \\
g(\delta) &:= (p_0^{\text{in}}+\beta) \Big/ \left(1-p_0^{\text{in}}\frac{\delta}{1+(1-\beta)\delta}\right), \qquad \delta \in [\epsilon, K].
\end{aligned}$$

Now we apply the re-parametrization

$$\eta := \frac{\delta}{1+\delta(1-\beta)} \in \left[\frac{1}{\epsilon^{-1}+1-\beta}, \frac{1}{K^{-1}+1-\beta}\right] =: \mathcal{I} \tag{3.37}$$

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 $\tilde{f}(\eta) > 0$ 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

$$1/p_0^{\text{in}} > 1/(1-\beta) > \eta. \tag{3.38}$$

The first inequality holds since:

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

We know $\alpha + \beta < 1$ by our model assumption, thus verifying (3.38).

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

$$\begin{aligned}
\frac{\partial^2}{\partial \eta^2} \tilde{h}(\eta) &= \frac{\partial^2}{\partial \eta^2} \left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + (i^{-1} - (1-\beta))\eta} \right)^{-1} = \frac{\partial^2}{\partial \eta^2} \left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right)^{-1} \\
&= 2 \left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right)^{-3} \left[\frac{\partial}{\partial \eta} \left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right) \right]^2 - \left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right)^{-2} \frac{\partial^2}{\partial \eta^2} \left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right)
\end{aligned} \tag{3.39}$$

We now claim that

$$\frac{\partial}{\partial \eta} \left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right) = \sum_{i=1}^{\infty} \frac{\partial}{\partial \eta} \left(\frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right) = -\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}(i^{-1} - (1-\beta))}{(1 + b_i(\eta))^2}, \tag{3.40}$$

$$\frac{\partial^2}{\partial \eta^2} \left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right) = \sum_{i=1}^{\infty} \frac{\partial^2}{\partial \eta^2} \left(\frac{p_{>i}^{\text{in}}}{1 + b_i(\eta)} \right) = 2\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}(i^{-1} - (1-\beta))^2}{(1 + b_i(\eta))^3}. \tag{3.41}$$

It suffices to check:

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

Note that for $i \geq 1$,

$$\sup_{\eta \in \mathcal{I}} \left| \frac{\partial}{\partial \eta} \left(\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)} \right) \right| = \sup_{\eta \in \mathcal{I}} \frac{p^{\text{in}}_{>i} |i^{-1} - (1-\beta)|}{(1+b_i(\eta))^2}$$
$$\leq (2-\beta) \sup_{\eta \in \mathcal{I}} \frac{p^{\text{in}}_{>i}}{(1+b_i(\eta))^2} \leq (2-\beta)(1+(1-\beta)K)^2 p^{\text{in}}_{>i}.$$

Recall (3.18), we then have

$$\sum_{i=0}^{\infty} p^{\text{in}}_{>i} = \sum_{i=0}^{\infty} \sum_{k>i} p^{\text{in}}_k = \sum_{k=0}^{\infty} \sum_{i=0}^{k-1} p^{\text{in}}_k = \sum_{k=0}^{\infty} k p^{\text{in}}_k = 1.$$

Hence,

$$\sum_{i=1}^{\infty} \sup_{\eta \in \mathcal{I}} \left| \frac{\partial}{\partial \eta} \left(\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)} \right) \right| \leq (2-\beta)(1+(1-\beta)K)^2 \sum_{i=0}^{\infty} p^{\text{in}}_{>i}$$
$$= (2-\beta)(1+(1-\beta)K)^2 < \infty,$$

which implies (3.40). Equation (3.41) then follows by a similar argument. Combining (3.39), (3.40) and (3.41) gives

$$\frac{\partial^2}{\partial \eta^2} \tilde{h}(\eta) = 2 \left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1+b_i(\eta)} \right)^{-3}$$
$$\times \left[\left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}(i^{-1}-(1-\beta))}{(1+b_i(\eta))^2} \right)^2 - \left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1+b_i(\eta)} \right) \left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}(i^{-1}-(1-\beta))^2}{(1+b_i(\eta))^3} \right) \right] < 0,$$

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

From Lemma 3.3, $\psi(\delta_{\text{in}}) = 0$ where $\psi(\cdot)$ is as defined in (3.12). Hence we have $f(\delta_{\text{in}}) = \alpha + \beta$ in a similar derivation to that of (3.36). Also from (3.33), we have $g(\delta_{\text{in}}) = \alpha + \beta$. Hence, δ_{in} is a solution to $f(\delta) = g(\delta)$.

Under the $\delta \mapsto \eta$ reparametrization in (3.37), we have that $\tilde{f}(\eta_{\text{in}}) = \tilde{g}(\eta_{\text{in}})$ where $\eta_{\text{in}} := \delta_{\text{in}}/(1+\delta_{\text{in}}(1-\beta))$, and also

$$\lim_{\eta \downarrow 0} \tilde{f}(\eta) = \sum_{i=1}^{\infty} p^{\text{in}}_{>i} = 1 - p^{\text{in}}_{>0} = \beta + p^{\text{in}}_0 = \lim_{\eta \downarrow 0} \tilde{g}(\eta).$$

This, along with the concavity of $\tilde{h}$, implies that η_{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 3.4, 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 (3.36) and is also strongly consistent, which completes the proof.

□

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 β by $\tilde{\beta} = 1 - N(n)/n$.

2. Obtain $\tilde{\delta}_{\text{in}}$ by solving

$$\sum_{i=1}^{\infty} \frac{N^{\text{in}}_{>i}(n)}{n} \frac{i}{i+\delta_{\text{in}}}(1+\delta_{\text{in}}(1-\tilde{\beta})) = \frac{\frac{N^{\text{in}}_0(n)}{n}+\tilde{\beta}}{1-\frac{N^{\text{in}}_0(n)}{n}\frac{\delta_{\text{in}}}{1+(1-\tilde{\beta})\delta_{\text{in}}}}.$$

3. Estimate α by

$$\tilde{\alpha} = \frac{\frac{N^{\text{in}}_0(n)}{n}+\tilde{\beta}}{1-\frac{N^{\text{in}}_0(n)}{n}\frac{\tilde{\delta}_{\text{in}}}{1+(1-\tilde{\beta})\tilde{\delta}_{\text{in}}}} - \tilde{\beta}.$$

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

$$\sum_{j=1}^{\infty} \frac{N^{\text{out}}_{>j}(n)}{n} \frac{j}{j+\delta_{\text{out}}}(1+\delta_{\text{out}}(1-\tilde{\beta})) = \frac{\frac{N^{\text{out}}_0(n)}{n}+\tilde{\beta}}{1-\frac{N^{\text{out}}_0(n)}{n}\frac{\delta_{\text{out}}}{1+(1-\tilde{\beta})\delta_{\text{out}}}}.$$

5. Estimate γ by

$$\tilde{\gamma} = \frac{\frac{N^{\text{out}}_0(n)}{n}+\tilde{\beta}}{1-\frac{N^{\text{out}}_0(n)}{n}\frac{\delta_{\text{out}}}{1+(1-\tilde{\beta})\delta_{\text{out}}}} - \tilde{\beta}.$$

Note that though this procedure does not guarantee $\tilde{\alpha}+\tilde{\beta}+\tilde{\gamma}=1$ precisely, all three estimators $\tilde{\alpha},\tilde{\beta},\tilde{\gamma}$ are strongly consistent, so $\tilde{\alpha}+\tilde{\beta}+\tilde{\gamma} \xrightarrow{\text{a.s.}} 1$.

4. Simulation study

We now apply the estimation procedures described in Sections 3.1 and 3.2 to simulated data, which allows us to compare the estimation results using the full evolution of the network with that using just one snapshot. Algorithm 1 is used to simulate realizations of the preferential attachment tnetwork.

4.1. MLE

For the first scenario of having full evolution of the network, we simulate 5000 independent replications of the preferential attachment network with 10^5 edges under the true parameter

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

For each realization, the MLE estimate of the parameters is computed and standardized according to its asymptotic limit (3.21), i.e., by

$$\frac{(\hat{\boldsymbol{\theta}}_n)_i - (\boldsymbol{\theta})_i}{\sqrt{\sigma^2_{ii}/n}}, \tag{4.1}$$

where $(\hat{\boldsymbol{\theta}}_n)_i$ and $(\boldsymbol{\theta})_i$ denote the i-th components of $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}$ respectively, and σ^2_{ii} is the i-th diagonal component of $\Sigma(\boldsymbol{\theta})$.

The QQ plots of the normalized estimates are shown in Figure 4.1, all of which line up quite well with the $y=x$ line (the red line). This reaffirms the asymptotic theory in Theorem 3.5. We can obtain confidence intervals for $\boldsymbol{\theta}$ by replacing the variance with estimated asymptotic variance of $\hat{\boldsymbol{\theta}}_n$ given by

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

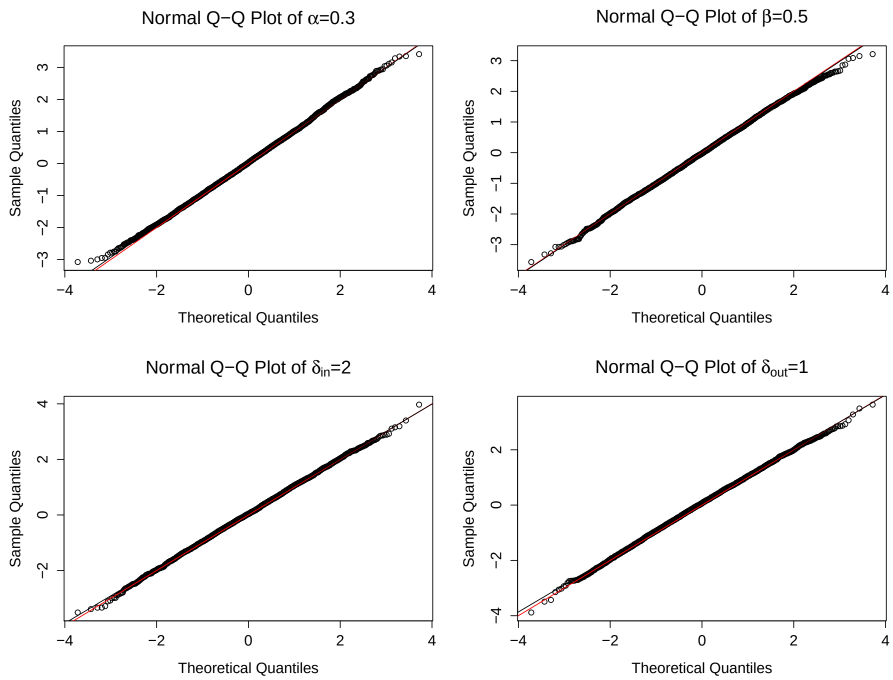


Fig 4.1. *Normal QQ-plots for normalized estimates in* (4.1) *under* 5000 *replications of a preferential attachment network with* 10^5 *edges and* $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$. *The fitted lines in black are the traditional qq-lines used to check normality of the estimates. The red line is the* $y = x$ *line in all plots.*

where

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

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

Given a single realization, the $(1 - \varepsilon)$ confidence interval for $(\boldsymbol{\theta})_i$ can be written as

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

where $z_{\varepsilon/2}$ is the upper $\varepsilon/2$ quantile of $N(0, 1)$ and $\hat{\sigma}^2_{ii}$ is the i-th diagonal component of $\hat{\Sigma}$.

4.2. One snapshot

We use the same simulated data as in Section 4.1, and obtain parameter estimates $\tilde{\boldsymbol{\theta}}_n := (\tilde{\alpha}, \tilde{\beta}, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}})$ using only the final snapshot, following the procedure described at the end of Section 3.2. The same normalization is applied by the true mean and true variance of the full MLE:

$$\frac{(\tilde{\boldsymbol{\theta}}_n)_i - (\boldsymbol{\theta})_i}{\sqrt{\sigma^2_{ii}/n}}, \quad i = 1, \ldots, 4, \tag{4.2}$$

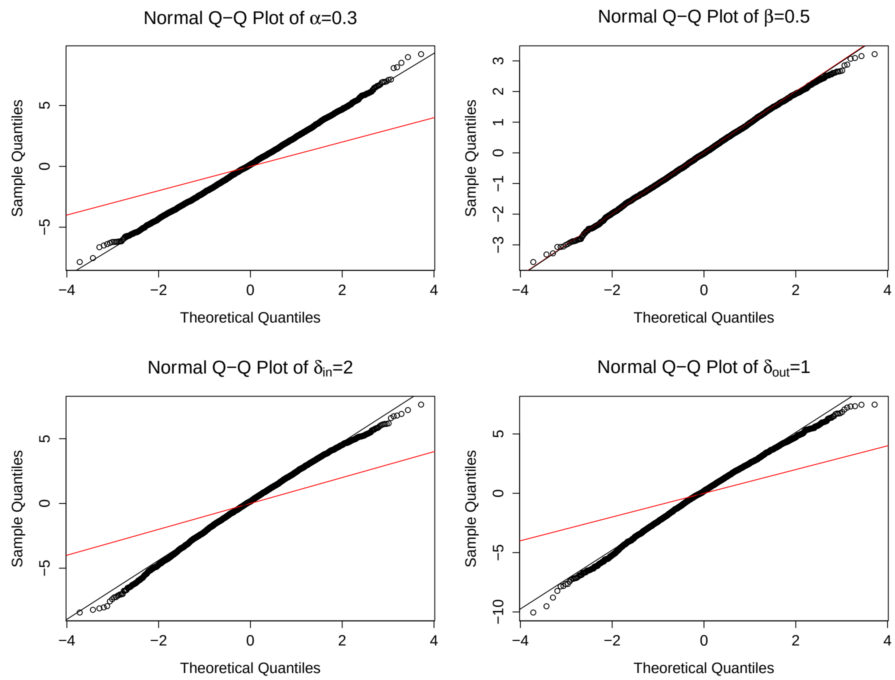


FIG 4.2. *Normal QQ-plots for the normalized estimates in* (4.2) *under* 5000 *replications of a preferential attachment network with* 10^5 *edges and* $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$. *The fitted lines in black are the traditional qq-lines used to check normality of the estimates. The red line is the* $y = x$ *line in all plots.*

where $(\tilde{\boldsymbol{\theta}}_n)_i$ denotes the i-th components of $\tilde{\boldsymbol{\theta}}_n$.

Figure 4.2 gives QQ-plots for the normalized estimates from the snapshots. Again, the fitted lines in black are the traditional qq-lines and the red lines are the $y = x$ line. The estimates of all four parameters look normal, but the slope of the QQ-lines for $\tilde{\alpha}, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}}$ are much steeper than the diagonal line, which gives an indication of the relative efficiency of $\tilde{\boldsymbol{\theta}}_n$ compared to $\hat{\boldsymbol{\theta}}_n$. The asymptotic relative efficiencies (ARE) are

$$\begin{aligned} ARE(\tilde{\alpha}) &= \lim_{n\to\infty} \frac{\alpha(1-\alpha)}{n\mathrm{Var}(\tilde{\alpha})} \approx \frac{\alpha(1-\alpha)}{n\widehat{\mathrm{Var}}(\tilde{\alpha})} \approx 0.2026, \\ ARE(\tilde{\delta}_{\text{in}}) &= \lim_{n\to\infty} \frac{I_{\text{in}}^{-1}}{n\mathrm{Var}(\tilde{\delta}_{\text{in}})} \approx \frac{\tilde{I}_{\text{in}}^{-1}}{n\widehat{\mathrm{Var}}(\tilde{\delta}_{\text{in}})} \approx 0.1922, \\ ARE(\tilde{\delta}_{\text{out}}) &= \lim_{n\to\infty} \frac{I_{\text{out}}^{-1}}{n\mathrm{Var}(\tilde{\delta}_{\text{out}})} \approx \frac{\tilde{I}_{\text{out}}^{-1}}{n\widehat{\mathrm{Var}}(\tilde{\delta}_{\text{out}})} \approx 0.1693. \end{aligned} \tag{4.3}$$

Comparing Figure 4.2 to Figure 4.1, we see that though using one snapshot gives consistent estimation, it inflates the estimator variance for all parameters except for β, where the true MLE (3.7) can be estimated directly from $G(n)$. This is as expected since knowing only the final snapshot provides far less information than the whole network history.

Given a single realization, the variance of the estimates can be estimated through resampling as follows. Using the estimated parameter $\tilde{\boldsymbol{\theta}}$, we simulate 10^4 independent replications of the network with 10^5 edges. Next, the model is fitted to each simulated network and the resulting parameter estimates, denoted by

$$\hat{\hat{\boldsymbol{\theta}}}_n := \left(\hat{\hat{\alpha}},\, \hat{\hat{\beta}},\, \hat{\hat{\delta}}_{\text{in}},\, \hat{\hat{\delta}}_{\text{out}}\right)\Bigg($$

are collected. The sample variance of $\hat{\tilde{\boldsymbol{\theta}}}_n$ can then be used as an approximation for the variance of $\tilde{\boldsymbol{\theta}}$. The ARE can be estimated by:

$$\begin{aligned} ARE(\tilde{\alpha}) &\approx \frac{\tilde{\alpha}(1-\tilde{\alpha})}{n\widehat{\mathrm{Var}}(\hat{\tilde{\alpha}})} \approx 0.2021, \\ ARE(\tilde{\delta}_{\mathrm{in}}) &\approx \frac{\tilde{I}_{\mathrm{in}}^{-1}}{n\widehat{\mathrm{Var}}(\hat{\tilde{\delta}}_{\mathrm{in}})} \approx 0.1887, \\ ARE(\tilde{\delta}_{\mathrm{out}}) &\approx \frac{\tilde{I}_{\mathrm{out}}^{-1}}{n\widehat{\mathrm{Var}}(\hat{\tilde{\delta}}_{\mathrm{out}})} \approx 0.1711, \end{aligned}$$

all of which correspond well with (4.3). Hence the $(1-\varepsilon)$-confidence interval for $\boldsymbol{\theta}$, assuming asymptotic normality, can be approximated by

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

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

5. Real network example

In this section, we explore fitting a preferential attachment model to a social network. As illustration, we chose the Dutch Wiki talk network dataset, available on KONECT (`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 appeared 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 ξ), 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 ρ), append to $G(n-1)$ a new node $v \in V(n) \setminus V(n-1)$ and self loop (v, v) onto itself.

These scenarios have been observed in other social network data, such as the Facebook wall post network (`http://konect.uni-koblenz.de/networks/facebook-wosn-wall`), etc. They occur in small proportions and can be easily accommodated by a slight modification in the model fitting procedure. The new model has parameters $(\alpha, \beta, \gamma, \xi, \delta_{\mathrm{in}}, \delta_{\mathrm{out}})$, and ρ is implicitly defined through $\rho = 1 - (\alpha + \beta + \gamma + \xi)$. Similar to the derivations in Section 3.1, the MLE estimators for $\alpha, \beta, \gamma, \xi$ are

$$\begin{aligned} \hat{\alpha}^{MLE} &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=1\}}, \quad \hat{\beta}^{MLE} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=2\}}, \\ \hat{\gamma}^{MLE} &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=3\}}, \quad \hat{\xi}^{MLE} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=4\}}, \end{aligned}$$

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

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

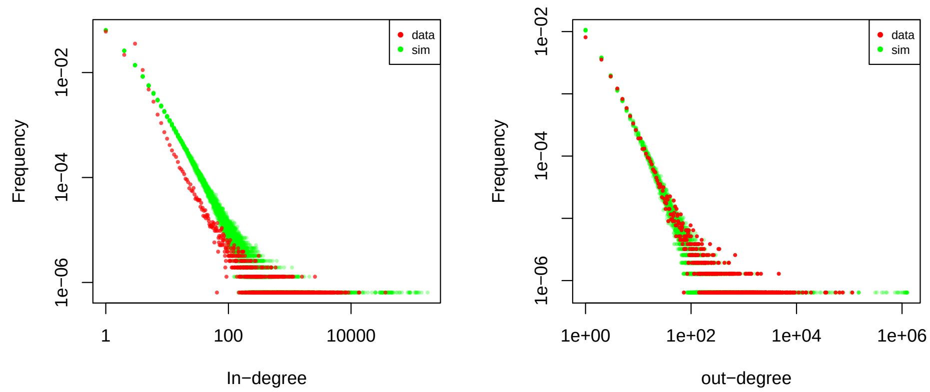


FIG 5.1. *In- and out-degree frequencies of the full Wiki talk network (red) and 20 simulated fitted linear preferential attachment networks with constant parameters (green).*

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

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

$$(\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\xi},\hat{\rho},\hat{\delta}_{\text{in}},\hat{\delta}_{\text{out}}) = (3.08\times10^{-3}, 8.55\times10^{-1}, 1.39\times10^{-1}, 4.76\times10^{-5}, 3.06\times10^{-3}, 0.547, 0.134).$$

To evaluate the goodness-of-fit, 20 network realizations are simulated from the fitted model. We overlaid the in- and out-degree frequencies of the original network with that of the simulations. If the model fits the data well, the degrees of the data should lie within the range formed by that of the simulations. From Figure 5.1, 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 time stamps and discovered bursts of messages originating from certain nodes over small time intervals. According to Wikipedia policy [23], certain administrating accounts are allowed to sent group messages to multiple users simultaneously. These bursts presumably represent broadcast announcements generated from these accounts. These administrative broadcasts can also be detected when we apply the linear preferential attachment model to the network in local time intervals. We divided the total time frame into sub-intervals each containing the formation of 10^4 edges. This generated 20 data sets

$$\Big(\{G(n_{k-1}),\dots,G(n_k-1)\}, k=1,\dots,20\Big).$$

For each of the 20 data sets, we fit a preferential attachment model using MLE. The resulting estimates $(\hat{\delta}_{\text{in}},\hat{\delta}_{\text{out}})$ are plotted against the corresponding timeline on the upper left panel of Figure 5.2. Notice that $\hat{\delta}_{\text{in}}$ exhibits large spikes at various times. Recall from (2.1), a large value of δ_{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 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.

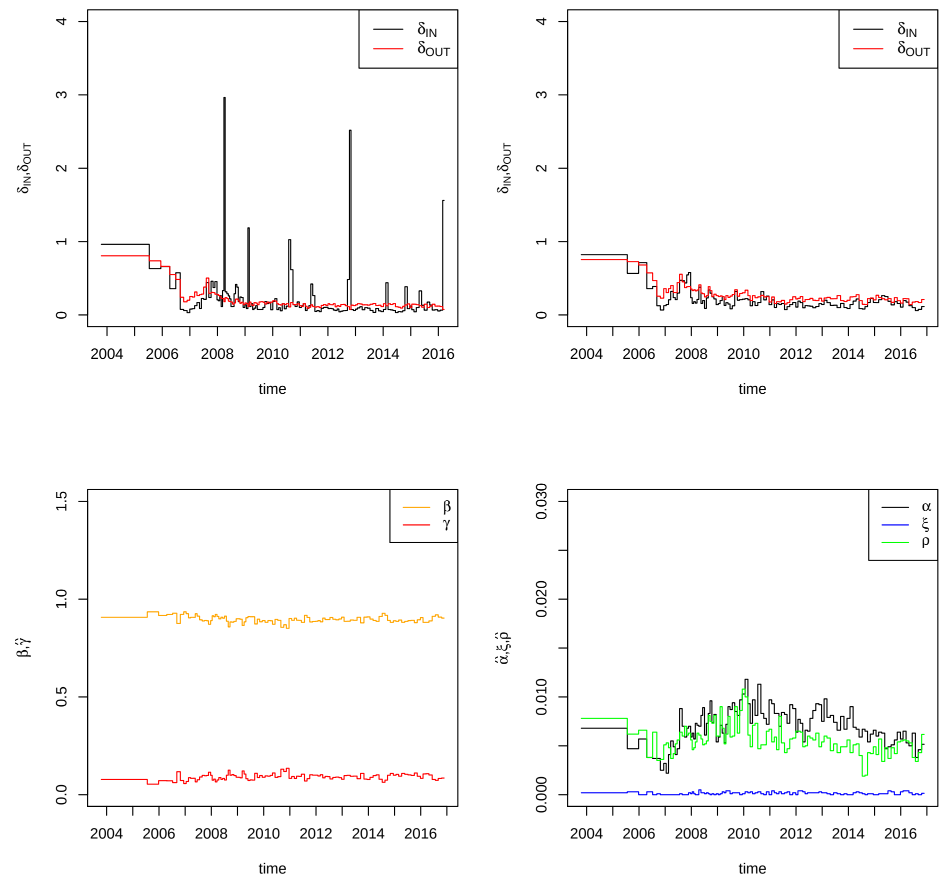


FIG 5.2. *Local parameter estimates of the linear preferential attachment model for the full and reduced Wiki talk network. Upper left:* $(\hat{\delta}_{in}, \hat{\delta}_{out})$ *for the full network. Upper right, lower left, lower right:* $(\hat{\delta}_{in}, \hat{\delta}_{out})$, $(\hat{\beta}, \hat{\gamma})$, $(\hat{\alpha}, \hat{\xi}, \hat{\rho})$ *for the reduced network, respectively.*

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 exclude the messages from these accounts from the dataset in our analysis. We 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 5.2. Here all parameter estimates are quite stable through time. This suggests that the network is less likely to contain large-scale group messages which stands out among normal individual interactions.

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 based on MLE for our reduced dataset are

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

Again the degree distributions of the data and 20 simulations from the fitted model are displayed in Figure

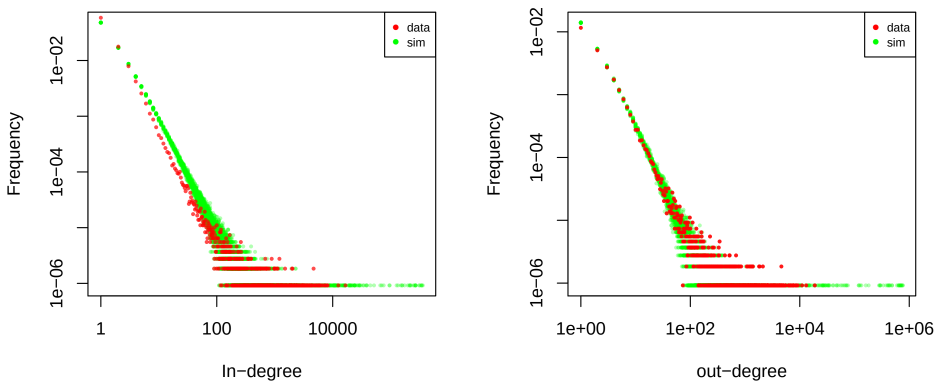


FIG 5.3. *In- and out-degree distributions of the reduced Wiki talk network (red) and 20 simulated fitted linear preferential attachment networks with constant parameters (green).*

5.3. 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 5.1). 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 inlined 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.

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. As a result, it may be used as a building block for more flexible models. Modification to the existing model formulation and more careful analysis of changepoints in parameters is a direction for future research.

6. Parameter estimation using asymptotics

So far we have seen procedures to fit the linear preferential attachment model under two data scenarios. Both procedures produce consistent and asymptotically normal estimators, but are heavily dependent on the correctness of the model. For real social network data, this adherence to the model assumptions is hard to guarantee and the issue becomes how much does one trust the correctness of the model.

In this section, we propose a semi-parametric asymptotic method through estimating the one-dimensional tail indices of both in- and out-degrees using extreme value analysis. This estimation procedure uses only the portion of the nodes with large degrees in the network data. For simulated data from a known model, these estimates do not perform as well as MLE; yet when the underlying model is misspecified, this procedure should be useful.

6.1. Methodology

We first give an overview of a method based on extreme value theory, and then apply it to the linear preferential attachment setting in Section 6.2 as a benchmark.

Let $\mathbb{M}(\mathbb{R}_+^2 \setminus \{\mathbf{0}\})$ be the set of Borel measures on $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$ that are finite on sets bounded away from the origin. A random vector $(X, Y) \geq 0$ is *non-standard regularly varying* on $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$, if there exists regularly varying functions b_1, b_2 with positive indices, called the scaling functions, and a measure $\nu(\cdot) \in \mathbb{M}(\mathbb{R}_+^2 \setminus \{\mathbf{0}\})$, called the limit or tail measure, such that as $t \to \infty$,

$$t\mathbf{P}\left[\left(\frac{X}{b_1(t)}, \frac{Y}{b_2(t)}\right) \in \cdot\right] \to \nu(\cdot), \quad \text{in } \mathbb{M}(\mathbb{R}_+^2 \setminus \{\mathbf{0}\}), \tag{6.1}$$

in the sense of $\mathbb{M}$-convergence, see [5, 9, 15] for details. Without loss of generality, we assume all scaling functions are continuous and strictly increasing. The phrasing in (6.1) implies the marginal distributions have regularly varying tails.

When the scaling functions are power functions, i.e., $b_i(t) = t^{\iota_i}$, $\iota_i > 0$, $i = 1, 2$, the vector can be transformed using a power function so that it is standard regularly varying. For instance with $a = \iota_2/\iota_1$,

$$t\mathbf{P}\left[\left(\frac{X^a}{t^{\iota_2}}, \frac{Y}{t^{\iota_2}}\right) \in \cdot\right] \to \tilde{\nu}(\cdot), \quad \text{in } \mathbb{M}(\mathbb{R}_+^2 \setminus \{\mathbf{0}\}), \tag{6.2}$$

where $\tilde{\nu} = \nu \circ T^{-1}$ with $T(x, y) = (x^a, y)$. The advantage of the standard form is that one can estimate the angular component of the transformed measure [16]. Of course, in order to apply the transformation T, one needs to know the tail indices ι_1 and ι_2. From the data, we estimate these quantities using a method proposed in [4] that we refer to as the 'minimum distance method'. We now give a brief summary of this method.

Given a sample of n iid observations, $Z_1, \ldots, Z_n$ from a power law distribiution, the minimum distance method suggests using the thresholded data consisting of the k upper-order statistics, $Z_{(1)} \geq \ldots \geq Z_{(k)}$, for estimating ι. The tail index can be estimated based on these order statistics by the Hill estimator defined by

$$\hat{\iota}(k) := \left(\frac{1}{k-1} \sum_{i=1}^{k-1} \log \frac{Z_{(i)}}{Z_{(k)}}\right)^{-1}, \quad k \geq 2.$$

To select k, we first compute the Kolmogorov-Smirnov (KS) distance between the empirical distribution of the upper k observations and the power-law distribution with index $\hat{\iota}(k)$:

$$D_k := \sup_{y \geq 1} \left| \frac{1}{k} \sum_{i=1}^{n} \epsilon_{Z_i/Z_{(k)}}(y, \infty] - y^{-\hat{\iota}(k)} \right|, \quad 1 \leq k \leq n.$$

Then the optimal k^* is the one that minimizes the KS distance

$$k^* := \operatorname*{argmin}_{1 \leq k \leq n} D_k,$$

and we estimate the tail index and threshold by $\hat{\iota}(k^*)$ and $Z_{(k^*)}$ respectively. This estimator performs well if the thresholded portion comes from a Pareto tail and also seems effective in a variety of non iid scenarios.

Returning to data satisfying (6.1), after determining marginal tail indices and standardizing the data according to (6.2), we apply a polar coordinate transform to the limit measure $\tilde{\nu}$ that factorizes into a product of a Pareto measure and an angular measure. The angular measure here is one way to describe the asymptotic dependence structure of the standardized (X, Y) (cf. [16, Section 6.1.4]). For models alternative to the one given in Section 2, applying parametric or non-parametric estimators to the angular measure may give estimates for other model parameters. Section 6.2 provides an example where we can compute the asymptotic angular density in closed form and use it as the true density for computing the MLE.

This semi-parametric asymptotic method may be more robust against modeling error and gives more flexibility in the choice of models. In contrast, the full MLE and one-snapshot methods rely heavily on the precise model assumptions and may not be accurate when the model is not correctly specified.

6.2. Applications to the preferential attachment model

For the preferential attachment model specified in Section 2.1, the limiting degree distribution (p_{ij}) is jointly regularly varying in i and j [19] with power law marginals given by

$$p_i^{\text{in}} := \sum_{j=0}^{\infty} p_{ij} \sim C_{\text{in}} i^{-\iota_{\text{in}}} \text{ as } i \to \infty, \quad \text{as long as } \alpha\delta_{\text{in}} + \gamma > 0, \tag{6.3}$$

$$p_j^{\text{out}} := \sum_{i=0}^{\infty} p_{ij} \sim C_{\text{out}} i^{-\iota_{\text{out}}} \text{ as } j \to \infty, \quad \text{as long as } \gamma\delta_{\text{out}} + \alpha > 0, \tag{6.4}$$

for some finite positive constants C_{in} and C_{out}, where

$$\iota_{\text{in}} = 1 + \frac{1 + \delta_{\text{in}}(\alpha+\gamma)}{\alpha+\beta} =: 1 + \frac{1}{c_1}, \quad \text{and} \quad \iota_{\text{out}} = 1 + \frac{1 + \delta_{\text{out}}(\alpha+\gamma)}{\beta+\gamma} =: 1 + \frac{1}{c_2}. \tag{6.5}$$

Therefore instead of estimating $(\hat{\alpha}, \hat{\beta}, \hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}})$, one may calculate $(\hat{\alpha}, \hat{\beta}, \hat{\iota}_{\text{in}}, \hat{\iota}_{\text{out}})$, where $\hat{\iota}_{\text{in}}, \hat{\iota}_{\text{out}}$ are estimates for the tail indices of the in- and out-degrees that can be obtained by applying the minimum distance method. In addition, similar to the one-snapshot case, β is estimated by $\hat{\beta} = 1 - N(n)/n$.

With $\hat{\beta}, \hat{\iota}_{\text{in}}, \hat{\iota}_{\text{out}}$ determined, the only parameter remaining to be estimated is α and we obtain $\hat{\alpha}$ from the asymptotic angular density as mentioned in Section 6.1. Let (I, O) be the pair of random variables that follows the limiting degree distribution p_{ij} defined in (3.11) and recall the results from [19]: as $t \to \infty$,

$$t\mathbf{P}\left[\left(\frac{I}{t^{1/(\iota_{\text{in}}-1)}}, \frac{O}{t^{1/(\iota_{\text{out}}-1)}}\right) \in \cdot\right] \to \nu(\cdot), \quad \text{in } \mathbb{M}(\mathbb{R}_+^2 \setminus \{\mathbf{0}\}).$$

Applying the power function followed by a switch to polar coordinates, i.e., $(I^a, O) \mapsto (\arctan(O/I^a), \sqrt{I^{2a} + O^2})$, where $a := c_2/c_1 = (\iota_{\text{in}} - 1)/(\iota_{\text{out}} - 1)$, the distribution of $\arctan(O/I^a)$ given $I^{2a} + O^2 > r_0$ converges to a random variable Θ as $r_0 \to \infty$. By [19, Section 4.1.2], the pdf of Θ is given by

$$\begin{aligned} f(\theta) \quad \propto \quad & \frac{\gamma}{\delta_{\text{in}}} (\cos\theta)^{\frac{\delta_{\text{in}}+1}{a} - 1} (\sin\theta)^{\delta_{\text{out}} - 1} \int_0^\infty t^{c_1^{-1} + \delta_{\text{in}} + a\delta_{\text{out}}} e^{-t(\cos\theta)^{1/a} - t^a \sin\theta} \mathrm{d}t \\ & + \frac{\alpha}{\delta_{\text{out}}} (\cos\theta)^{\frac{\delta_{\text{in}}}{a} - 1} (\sin\theta)^{\delta_{\text{out}}} \int_0^\infty t^{a - 1 + c_1^{-1} + \delta_{\text{in}} + a\delta_{\text{out}}} e^{-t(\cos\theta)^{1/a} - t^a \sin\theta} \mathrm{d}t. \end{aligned} \tag{6.6}$$

Given the degree counts for all nodes (I_i, O_i), we estimate $\hat{\alpha}$ by maximizing the likelihood of $f(\theta)$ based on the observations (I_i, O_i) for which $I_i^{2a} + O_i^2 > r_0$ for a large threshold r_0.

There are two issues when applying the minimum distance method to network data. First, the data is node-based and not like those collected from independent repeated sampling. Secondly, the degree counts are discrete and do not exactly comply with the Pareto assumption made in the minimum distance method. Our analysis shows that even if we ignore these two issues, the tail estimates obtained still perform reasonably well. However, if the model is correct, the asymptotic based estimates are dominated by MLE methods.

6.3. Estimation results

6.3.1. Tail estimates

We start with simulating one preferential attachment network with 10^6 edges under the true parameter $(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}) = (0.3, 0.5, 2, 1)$, so the theoretical values of the tail indices are $(\iota_{\text{in}}, \iota_{\text{out}}) = (1 + 1/c_1, 1 + 1/c_2) = (3.5, 3.14)$. The Hill plots correspond to the first 2000 upper order statistics of in- and out-degrees are included in Figure 6.1. Let i^* and j^* be the number of order statistics used for estimating ι_{in} and ι_{out}. The minimum distance method suggests choosing $(i^*, j^*) = (386, 213)$, which corresponds to using the observations such that $I > I^*$ and $O > O^*$ where $(\hat{I}^*, \hat{O}^*) = (52, 85)$. The tail estimates are $(\hat{\iota}_{\text{in}}, \hat{\iota}_{\text{out}}) = (3.43, 3.28)$.

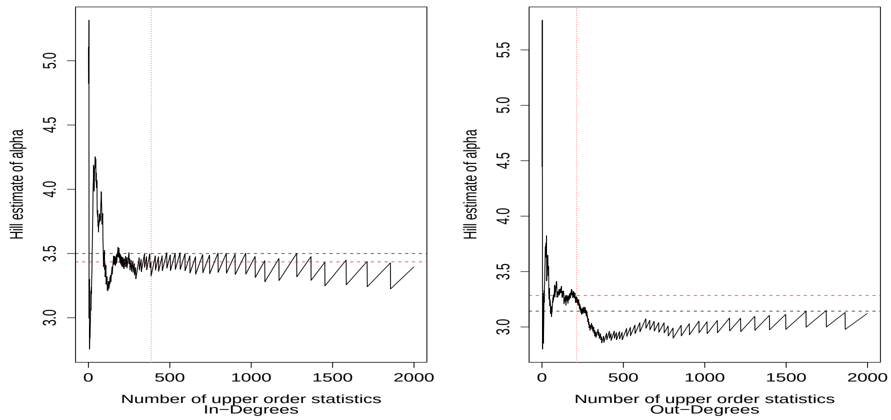


FIG 6.1. *Hill plots for in- (left) and out-degrees (right). The black dashed lines represent the theoretical values of the tail indices, and the red dashed lines reflect the estimation results using the minimum distance method.*

6.3.2. *Asymptotic method*

Table 6.1 presents some numerical results using the asymptotic method on simulated data. For each set of parameter values $(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$, a network with $n = 10^6$ edges is simulated and the true value of $(\iota_{\text{in}}, \iota_{\text{out}})$ is computed by (6.5). Then we estimate $(\iota_{\text{in}}, \iota_{\text{out}})$ by both the minimum distance method, denoted by $(\hat{\iota}_{\text{in}}, \hat{\iota}_{\text{out}})$, and from the one-snapshot method applied to the parametric model (cf. Section 4.2), denoted by $(\tilde{\iota}_{\text{in}}, \tilde{\iota}_{\text{out}})$.

With $(\hat{\iota}_{\text{in}}, \hat{\iota}_{\text{out}})$, $\hat{\alpha}$ is then estimated from the angular density. There are two underlying assumptions made here. First, given degree counts (I_i, O_i) for each node, we are assuming that n is large enough that the joint distribution of in- and out-degrees is close to that of the limit pair (I, O) which follows the mass function p_{ij}. Moreover, after choosing a large r_0, we also assume that the distribution of $\arctan(O/I^a)$, conditioned on $I^{2a} + O^2 > r_0^2$, has converged to the limit distribution of the angular component Θ with density as in (6.6).

In our experiment, r_0 is chosen to be the upper 99.9%-quantile of $(I_i^{2a}+O_i^2)$ (which includes approximately 600 observations here, according to our choice of β), and we then fit the limit density $f(\theta)$ to the transformed data $\arctan(O_i/I_i^a)$ for which $I_i^{2a} + O_i^2 > r_0^2$.

n	$(\alpha, \beta, \iota_{\text{in}}, \iota_{\text{out}})$	$(\hat{\alpha}, \hat{\beta}, \hat{\iota}_{\text{in}}, \hat{\iota}_{\text{out}})$	$(\tilde{\alpha}, \tilde{\beta}, \tilde{\iota}_{\text{in}}, \tilde{\iota}_{\text{out}})$
10^6	(0.3, 0.4, 2.857, 2.857)	(0.266, 0.400, 2.837, 2.729)	(0.300, 0.400, 2.858, 2.859)
10^6	(0.3, 0.4, 3.286, 3.286)	(0.302, 0.399, 3.258, 3.261)	(0.302, 0.399, 3.290, 3.295)
10^6	(0.3, 0.4, 5, 5)	(0.300, 0.400, 4.486, 4.852)	(0.302, 0.400, 5.051, 5.013)
10^6	(0.3, 0.4, 3.286, 4.143)	(0.328, 0.400, 3.365, 4.516)	(0.300, 0.400, 3.283, 4.156)
10^6	(0.1, 0.4, 4.2, 2.778)	(0.050, 0.400, 4.304, 2.719)	(0.100, 0.400, 4.215, 2.778)
10^6	(0.4, 0.4, 3, 3.667)	(0.431, 0.400, 3.021, 3.518)	(0.400, 0.400, 3.005, 3.668)

TABLE 6.1
Estimates for $(\iota_{in}, \iota_{out})$ using both minimum distance and one-snapshot methods.

For simulated data where we know the model is correct, one-snapshot parametric estimation described in Section 3.2 gives more accurate estimates in all cases than the asymptotic methods which may suffer from three possible sources of errors. The first is that both (6.3) and (6.4) hold only asymptotically for large in-/out-degrees, so the proportion of observations that exceeds the estimated threshold might be small. For example, in the case where $(\alpha, \beta, \iota_{\text{in}}, \iota_{\text{out}}) = (0.3, 0.4, 5, 5)$, the largest in-(out-)degree is 114(210), respectively, and we are using the largest 134(39) observations to estimate ι_{in} (ι_{out}). Under such circumstances, the tail estimates $(\hat{\iota}_{\text{in}}, \hat{\iota}_{\text{out}})$ will be more variable across simulation runs, leading to inaccuracy in the estimation of a before even attempting to calculate α by MLE from the angular density.

Secondly, as noted before, estimating $\hat{\alpha}$ from (6.6) not only depends on the accuracy of $\hat{\iota}_{\text{in}}$ and $\hat{\iota}_{\text{out}}$, but also on whether n is large enough that (6.2) thought of as an approximation is valid. Has the network evolved long enough? Thirdly, the polar-transformed data $\arctan(O_i/I_i^a)$ for which $I_i^{2a} + O_i^2 > r_0^2$ is not generated from the angular density $f(\theta)$ directly. This double limit approximation introduces yet another layer of uncertainty and Table 6.1 reveals that $\hat{\alpha}$ is very inaccurate when the true value of α is relatively small (e.g., the fifth row where $\alpha = 0.1$).

7. Acknowledgement

Research of the four authors was partially supported by Army MURI grant W911NF-12-1-0385. Don Towsley, University of Massachusetts, introduced us to the model and within his group, James Atwood graciously supplied us with a simulation algorithm designed for a class of growth models broader than the one specified in Section 2.1. Joyjit Roy, formerly of Cornell, created an efficient algorithm designed to capitalize on the linear growth structure.

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. *available:* `http://www.unc.edu/~bhamidi/preferent.pdf`, 2007. To be submitted to Probability Surveys.

[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.

[4] A. Clauset, C.R. Shalizi, and M.E.J. Newman. Power-law distributions in empirical data. *SIAM Rev.*, 51(4):661–703, 2009.

[5] B. Das, A. Mitra, and S. Resnick. Living on the multi-dimensional edge: Seeking hidden risks using regular variation. *Advances in Applied Probability*, 45(1):139–163, 2013.

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

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

[8] 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.

[9] H. Hult and F. Lindskog. Regular variation for measures on metric spaces. *Publ. Inst. Math. (Beograd) (N.S.)*, 80(94):121–140, 2006.

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

[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.

[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] F. Lindskog, S.I. Resnick, and J. Roy. Regularly varying measures on metric spaces: Hidden regular variation and hidden jumps. *Probab. Surv.*, 11:270–314, 2014.

[16] 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.

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

[18] 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.

[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.

[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, 2014. Available at `http://www.win.tue.nl/~rhofstad/`; 300 pages; to appear: Cambridge Series in Statistical and Probabilistic Mathematics.

[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.