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 [?] while more mathematical treatments are available in [? ? ?]. Also see [?] for a statistical survey of methods for network data and [?] for consideration of statistics of an undirected network and [?] for asymptotics of a directed exponential random graph models. Limit theory for estimates of an undirected preferential attachment model were considered in [?].

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 [? ? ? ? ?]. While the marginal degree power laws in a simple linear preferential attachment model were established in [? ? ?], the joint regular variation (see [? ?]) which is akin to a *joint power law*, was only recently established [? ?]. In addition, it was shown in [?] 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 (2.6) and (2.7) 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 $G(n_0), G(n_0+1), \ldots, G(n_0+n)$. 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 inspired by [**?**] and is similar to that of [**?**].

Our fitting methods are implemented in a real data setting using the Dutch Wiki talk network [**?**]. 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 fitting the model over separate time intervals.

Often it is difficult to believe in the existence of a true model, especially one whose parameters remain constant over time. Allowing, as we do, a preferential attachment model with only a few parameters may seem simplistic and unrealistic for social network data. Of course, preferential attachment is only one mechanism for network formation and evidence for its use in fields outside data networks is mixed [**? ?**] and we restrict attention to linear preferential attachment. Even imperfect models have the potential to capture salient properties in the data, such as heavy-tailedness of the in-degree and out-degree distributions, and to identify departures from model assumptions. While maximum likelihood estimation is essentially the gold standard for cases when the underlying model is a good representation of the data, it may perform poorly in case the model is far from being appropriate. In forthcoming work, we consider a semi-parametric estimation approach for network models that exhibit heavy-tailed degree distributions. This alternative estimation methodology borrows ideas from extreme value theory.

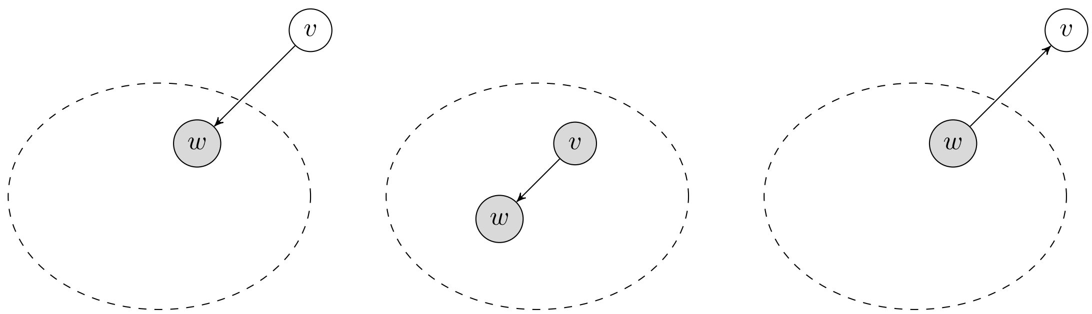
The rest of the paper is structured as follows. In Section 2, we formulate the linear preferential attachment network model and present an efficient simulation method for the network. Section 3 gives parameter estimators when either the full history is known or when only a single snapshot in time is available. We test these estimators against simulated data in Section 5 and then explore the Wiki talk network in Section 6.

2. Model specification and simulation

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

2.1. The linear preferential attachment model

The directed edge preferential attachment model [? ?] 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.



α-scheme β-scheme γ-scheme

- 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

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

- 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

$$\mathbf{P}[\text{choose } v \in V(n-1)] = \frac{D_{\text{out}}^{(n-1)}(v) + \delta_{out}}{n - 1 + \delta_{out} N(n-1)} . \tag{2.3}$$

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

2.2. Power law of degree distributions

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

$$\frac{N_{ij}(n)}{N(n)} \to f_{ij} =: \frac{p_{ij}}{1-\beta}, \quad (n \to \infty). \tag{2.4}$$

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

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

Moreover, p_i^{in} satisfies

$$p_i^{\text{in}} := \sum_{j=0}^{\infty} p_{ij} \sim C_{\text{in}} i^{-\iota_{\text{in}}} \text{ as } i \to \infty, \quad \text{as long as } \alpha\delta_{\text{in}} + \gamma > 0, \tag{2.5}$$

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

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

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

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

for $C_{\text{out}} \in (0, \infty)$ and

$$\iota_{\text{out}} = 1 + \frac{1 + \delta_{\text{out}}(\alpha+\gamma)}{\beta+\gamma}. \tag{2.7}$$

2.3. Simulation algorithm

We describe an efficient simulation procedure for the preferential attachment network given the parameter values $(\alpha, \beta, \gamma, \delta_{in}, \delta_{out})$, where $\alpha + \beta + \gamma = 1$. The simulation cost of the algorithm is linear in time. This algorithm, which was provided by Joyjit Roy during his graduate work at Cornell University, is presented below for completeness. Note that this simulation algorithm is specifically designed for the case where the preferential attachment probabilities (2.1)–(2.3) are linear in the degrees. A similar idea for the simulation of the Yule-Simon process appeared in [?]. Efficient simulation methods for the case where the preferential attachment probabilities are non-linear are studied in [?], where their algorithm trades some efficiency for the flexibility to model non-linear preferential attachment.

Using the notation from the introduction, at time $t = 0$, we initiate with an arbitrary graph $G(n_0) = (V(n_0), E(n_0))$ of n_0 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

Algorithm 1: Simulating a directed edge preferential attachment network

Algorithm

Input: $\alpha, \beta, \delta_{\text{in}}, \delta_{out}$, 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; $\delta \in \{\delta_{\text{in}}, \delta_{out}\}$, 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
```

$(v_n^{(1)}, v_n^{(2)})$. 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 [**?**] 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

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

Then

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

which corresponds to the desired selection probability (2.1).

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

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. 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 $e_t=(v^{(1)}_t,v^{(2)}_t)$ 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 $t=n_0+1,\dots,n$, the values of the following variables are known:

- $N(t)$, the number of nodes in graph $G(t)$;
- $D^{(t-1)}_{\text{in}}(v)$, $D^{(t-1)}_{\text{out}}(v)$, 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

$$\begin{aligned}
&L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}}|\ G(n_0),(e_t)_{t=n_0+1}^{n})\\
&=\prod_{t=n_0+1}^{n}\left(\alpha\frac{D^{(t-1)}_{\text{in}}(v^{(2)}_t)+\delta_{\text{in}}}{t-1+\delta_{\text{in}}N(t-1)}\right)^{\mathbf{1}_{\{J_t=1\}}}\\
&\quad\times\prod_{t=n_0+1}^{n}\left(\beta\Big(\frac{D^{(t-1)}_{\text{in}}(v^{(2)}_t)+\delta_{\text{in}}}{t-1+\delta_{\text{in}}N(t-1)}\Big)\Big(\frac{D^{(t-1)}_{\text{out}}(v^{(1)}_t)+\delta_{\text{out}}}{t-1+\delta_{\text{out}}N(t-1)}\Big)\right)^{\mathbf{1}_{\{J_t=2\}}}\\
&\quad\times\prod_{t=n_0+1}^{n}\left((1-\alpha-\beta)\frac{D^{(t-1)}_{\text{out}}(v^{(1)}_t)+\delta_{\text{out}}}{t-1+\delta_{\text{out}}N(t-1)}\right)^{\mathbf{1}_{\{J_t=3\}}}
\end{aligned}\tag{3.1}$$

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:

$$\frac{\partial}{\partial\alpha}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}}|\ G(n_0),(e_t)_{t=n_0+1}^n)=\frac{1}{\alpha}\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=1\}}-\frac{1}{1-\alpha-\beta}\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}}, \qquad (3.3)$$

$$\frac{\partial}{\partial\beta}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}}|\ G(n_0),(e_t)_{t=n_0+1}^n)=\frac{1}{\beta}\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=2\}}-\frac{1}{1-\alpha-\beta}\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}}, \qquad (3.4)$$

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

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.

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

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

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

to be the number of nodes with in-degree equal to i and greater than i, respectively, 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 3.2, 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)*:*

$$\begin{aligned}&L(\alpha, \beta, \delta_{in}, \delta_{out} |\, G(n_0), (e_t)_{t=n_0+1}^{n}) \\ &= \alpha^{\sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=1\}}}\, \beta^{\sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=2\}}}\, (1-\alpha-\beta)^{\sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}}}\end{aligned}$$

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

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

3.2. Consistency of MLE

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

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

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

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

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

The following theorem gives the consistency of the MLE of δ_{in} and δ_{out}.

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

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

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

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

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

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

Let us consider a limit version of ψ_n:

$$\psi(\lambda) := \sum_{i=0}^{\infty} \frac{p^{\mathrm{in}}_{>i}(\delta_{\mathrm{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - (1-\beta)a_1(\lambda), \tag{3.11}$$

where $p^{\mathrm{in}}_{>i}(\delta_{\mathrm{in}}) := \sum_{k>i} p^{\mathrm{in}}_k(\delta_{\mathrm{in}})$ with $p^{\mathrm{in}}_k(\delta_{\mathrm{in}}) := p^{\mathrm{in}}_k$ is as defined in (2.5), and

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

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

$$\sup_{\lambda \geq \epsilon} |\psi_n(\lambda) - \psi(\lambda)| \to 0. \tag{3.12}$$

Since ψ is continuous, for any $\kappa > 0$ arbitrarily small, there exists $\varepsilon_\kappa > 0$ such that $\psi(\lambda) > \varepsilon_\kappa$ for $\lambda \in [\epsilon, \delta_{\mathrm{in}} - \kappa]$ and $\psi(\lambda) < -\varepsilon_\kappa$ for $\lambda \in [\delta_{\mathrm{in}} + \kappa, K]$. From (3.12),

$$\mathbf{P}\left(\exists N_\kappa \ s.t. \sup_{n > N_\kappa} \sup_{\lambda \in [\epsilon, K]} |\psi_n(\lambda) - \psi(\lambda)| < \varepsilon_\kappa/2\right) = 1. \tag{3.13}$$

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

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

and

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

These jointly indicate that $\delta_{\mathrm{in}} - \kappa \leq \hat{\delta}^{MLE}_{\mathrm{in}} \leq \delta_{\mathrm{in}} + \kappa$. Hence (3.13) implies

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

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

3.3. Asymptotic normality of MLE

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

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

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

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^{MLE}_n - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \Sigma(\boldsymbol{\theta})), \tag{3.14}$$

where

$$\Sigma^{-1}(\boldsymbol{\theta}) = I(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1-\beta}{\alpha(1-\alpha-\beta)} & \frac{1}{1-\alpha-\beta} & 0 & 0 \\ \frac{1}{1-\alpha-\beta} & \frac{1-\alpha}{\beta(1-\alpha-\beta)} & 0 & 0 \\ 0 & 0 & I_{in} & 0 \\ 0 & 0 & 0 & I_{out} \end{bmatrix}, \tag{3.15}$$

with

$$I_{in} = \sum_{i=0}^{\infty} \frac{p^{in}_{>i}}{(i+\delta_{in})^2} - \frac{\gamma}{\delta_{in}^2} - \frac{(\alpha+\beta)(1-\beta)^2}{\left(1+\delta_{in}(1-\beta)\right)^2}, \tag{3.16}$$
$$I_{out} = \sum_{j=0}^{\infty} \frac{p^{out}_{>j}}{(j+\delta_{out})^2} - \frac{\alpha}{\delta_{out}^2} - \frac{(\gamma+\beta)(1-\beta)^2}{\left(1+\delta_{out}(1-\beta)\right)^2}.$$

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

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

Proof of Theorem 3.3. We first show the limiting distributions for $(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE})$, $\hat{\delta}^{MLE}_{\text{in}}$, and $\hat{\delta}^{MLE}_{\text{out}}$. 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 for sums of independent random variables.

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

$$u_t(\delta) := \frac{1}{D^{(t-1)}_{\text{in}}(v^{(2)}_t)+\delta}\mathbf{1}_{\{J_t\in\{1,2\}\}} - \frac{N(t-1)}{t-1+\delta N(t-1)}\mathbf{1}_{\{J_t\in\{1,2\}\}}. \tag{3.17}$$

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

$$0 = \sum_{t=1}^{n} u_t(\hat{\delta}^{MLE}_{\text{in}}) = \sum_{t=1}^{n} u_t(\delta_{\text{in}}) + (\hat{\delta}^{MLE}_{\text{in}} - \delta_{\text{in}}) \sum_{t=1}^{n} \dot{u}_t(\hat{\delta}^{*}_{\text{in}}), \tag{3.18}$$

where $\dot{u}_t$ denotes the derivative of u_t and $\hat{\delta}^{*}_{\text{in}} = \delta_{\text{in}} + \xi(\hat{\delta}^{MLE}_{\text{in}} - \delta_{\text{in}})$ for some $\xi \in [0,1]$. An elementary transformation of (3.18) gives

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

To establish

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

where I_{in} is as defined in (3.15), it suffices to show the following two results:

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

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

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

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

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

$$\mathbf{0} = \mathbf{S}_n\left(\hat{\boldsymbol{\theta}}_n^{MLE}\right) = \mathbf{S}_n(\boldsymbol{\theta}) + \dot{\mathbf{S}}_n\left(\hat{\boldsymbol{\theta}}_n^{*}\right)\left(\hat{\boldsymbol{\theta}}_n^{MLE} - \boldsymbol{\theta}\right), \tag{3.19}$$

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

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

Hence

$$\frac{1}{n}\dot{\mathbf{S}}_n(\hat{\boldsymbol{\theta}}_n^{*}) = \begin{bmatrix} \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial \alpha^2} & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial \alpha \partial \beta} & 0 & 0 \\ \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial \beta \partial \alpha} & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial \beta^2} & 0 & 0 \\ 0 & 0 & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial \delta_{\text{in}}^2} & 0 \\ 0 & 0 & 0 & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial \delta_{\text{out}}^2} \end{bmatrix} \xrightarrow{p} I(\boldsymbol{\theta}) \tag{3.20}$$

as implied in the previous part of the proof, where $I(\boldsymbol{\theta})$ is as defined in (3.15) 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 (3.19) and (3.20). □

4. Parameter estimation based on one snapshot

Based only on the single snapshot $G(n)$, we propose a parameter estimation procedure. It is worth noting that the choice of the snapshot should not not depend on any endogenous information related to the network, i.e., it is just a fixed time point. Since no information on the initial graph $G(n_0)$ is available, we merely assume n_0 and $N(n_0)$ are fixed and $n \to \infty$.

Among the sufficient statistics for $(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}})$ 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 A.2:

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

and

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

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

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

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 (A.1a):

$$\left(1+\frac{(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}}\right)p^{\text{in}}_0=\alpha, \tag{4.2}$$

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

$$\left(1+\frac{(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}}\right)\frac{N^{\text{in}}_0(n)}{n}=\alpha. \tag{4.3}$$

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

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

$$\left(1+\frac{(\alpha+\tilde{\beta})\delta_{\text{in}}}{1+(1-\tilde{\beta})\delta_{\text{in}}}\right)\frac{N^{\text{in}}_0(n)}{n}=\alpha, \tag{4.4b}$$

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

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

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

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

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

1. Estimate 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

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

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

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

5. Estimate γ by

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

Note that though this procedure does not imply

$$\tilde{\alpha} + \tilde{\beta} + \tilde{\gamma} = 1. \tag{4.5}$$

However, since all three estimators $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}$ are strongly consistent, (4.5) does hold in the limit, i.e., $\tilde{\alpha}+\tilde{\beta}+\tilde{\gamma} \xrightarrow{\text{a.s.}} 1$. One way to overcome this defect is to re-normalize the probabilities

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

such that (4.5) is satisfied. Then $\tilde{\delta}_{\text{in}}$ can be solved through (4.4a) by plugging the new value of $\tilde{\alpha}$ and similarly $\tilde{\delta}_{\text{out}}$ through a symmetric argument.

5. Simulation study

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

5.1. MLE

For the 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

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

For each realization, the MLE estimate of the parameters is computed and standardized as below:

$$\frac{(\hat{\boldsymbol{\theta}}_n^{MLE})_i - (\boldsymbol{\theta})_i}{\sqrt{\hat{\sigma}_{ii}^2/n}}, \tag{5.2}$$

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

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

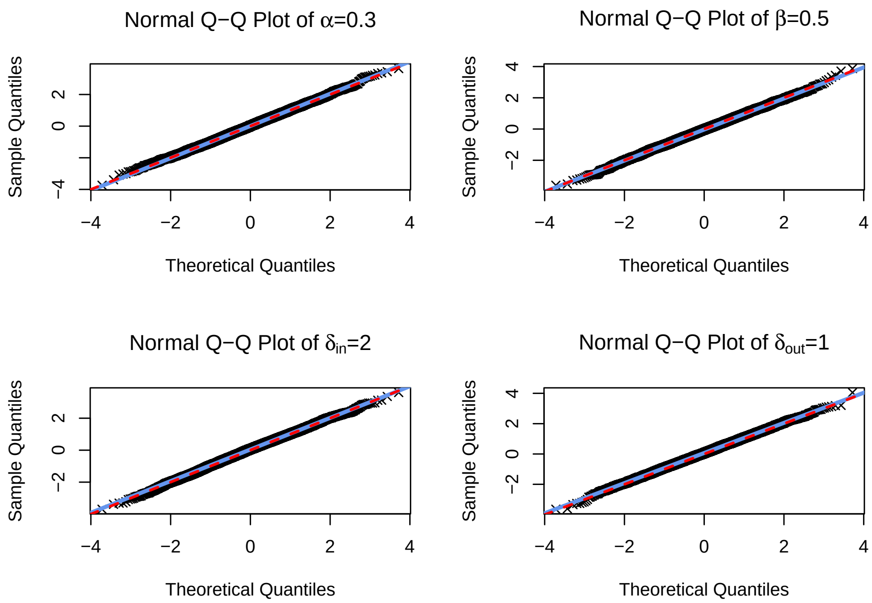


FIG 5.1. *Normal QQ-plots for normalized estimates in* (5.2) *under* 5000 *replications of a preferential attachment network with* 10^5 *edges and* $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$. *The fitted lines in green are the traditional qq-lines used to check normality of the estimates. The red dashed line is the* $y = x$ *line in all plots.*

where

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

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

By Theorem 3.3, we will expect the normalized MLE estimates follows a standard normal distribution so that the slope of the fitted QQ-lines must be close to 1.

The QQ plots of the normalized estimates are shown in Figure 5.1, all of which line up quite well with the $y = x$ line (the red dashed line). This therefore reaffirms the asymptotic theory in Theorem 3.3. We can now obtain confidence intervals for $\boldsymbol{\theta}$. Given a single realization, the $(1 - \varepsilon)$ confidence interval for $(\boldsymbol{\theta})_i$ can be written as

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

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

5.2. One snapshot

We use the same simulated data as in Section 5.1, 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 4. The same normalization

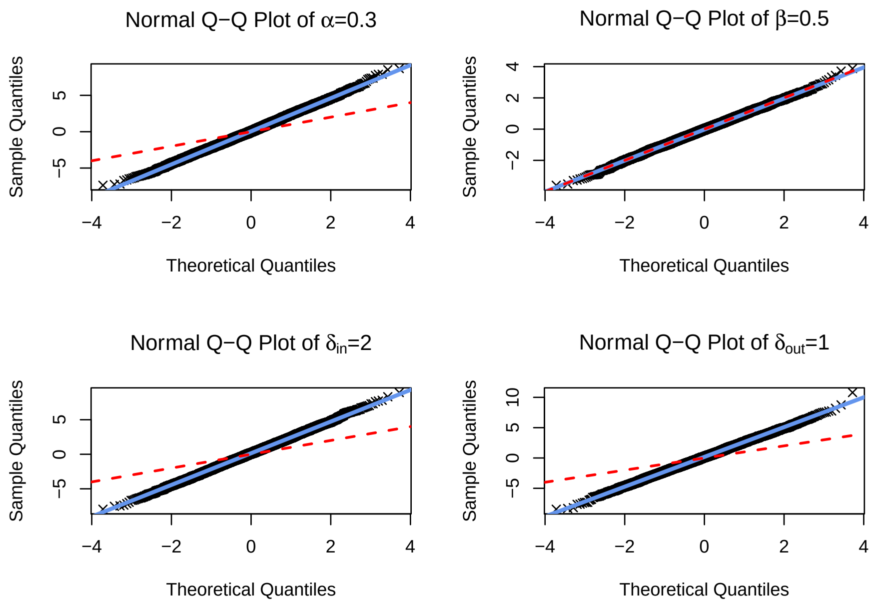


FIG 5.2. *Normal QQ-plots for the normalized estimates in* (5.3) *under* 5000 *replications of a preferential attachment network with* 10^5 *edges and* $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$. *The fitted lines in green are the traditional qq-lines used to check normality of the estimates. The red dashed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

is applied:

$$\frac{(\tilde{\boldsymbol{\theta}}_n)_i - (\boldsymbol{\theta})_i}{\sqrt{\hat{\sigma}_{ii}^2/n}}, \quad i = 1, \dots, 4, \tag{5.3}$$

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

Figure 5.2 gives QQ-plots for the normalized estimates from the snapshots. Again, the fitted lines in green are the traditional qq-lines and the red dashed lines are the $y = x$ line. The QQ-plot for $\tilde{\beta}$ exhibits a same shape as for $\hat{\beta}$. This is because $\tilde{\beta}$ is equal to the full MLE $\hat{\beta}$.

On the other hand, although the estimates of all four parameters look normal, the slopes of the QQ-lines for $\tilde{\alpha}, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}}$ are much steeper than the diagonal line, which indicates a loss of efficiency for $\tilde{\boldsymbol{\theta}}_n$ compared with $\hat{\boldsymbol{\theta}}_n$. Recall that for a consistent estimator T_n of a one-dimensional parameter θ, constructed from a random sample of size n, the asymptotic relative efficiencies (ARE) of T_n is defined by

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

where T_n^* denotes the asymptotically efficient estimator.

Hence in our cases, the ARE's are

$$ARE(\tilde{\alpha}) = \lim_{n\to\infty} \frac{n\text{Var}(\hat{\alpha}^{MLE})}{n\text{Var}(\tilde{\alpha})} \approx \frac{\widehat{\text{Var}}(\hat{\alpha}^{MLE})}{\widehat{\text{Var}}(\tilde{\alpha})} \approx 0.1971, \tag{5.4a}$$

$$ARE(\tilde{\delta}_{\text{in}}) = \lim_{n\to\infty} \frac{n\text{Var}(\hat{\delta}_{\text{in}}^{MLE})}{n\text{Var}(\tilde{\delta}_{\text{in}})} \approx \frac{\widehat{\text{Var}}(\hat{\delta}_{\text{in}}^{MLE})}{\widehat{\text{Var}}(\tilde{\delta}_{\text{in}})} \approx 0.1859, \tag{5.4b}$$

$$ARE(\tilde{\delta}_{\text{out}}) = \lim_{n\to\infty} \frac{n\text{Var}(\hat{\delta}_{\text{out}}^{MLE})}{n\text{Var}(\tilde{\delta}_{\text{out}})} \approx \frac{\widehat{\text{Var}}(\hat{\delta}_{\text{out}}^{MLE})}{\widehat{\text{Var}}(\tilde{\delta}_{\text{out}})} \approx 0.1695. \tag{5.4c}$$

Comparing Figure 5.2 to Figure 5.1, 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 $\tilde{\boldsymbol{\theta}}_n$, we simulate $b = 10^4$ independent bootstrap replicates of the network with 10^5 edges. Next, the model is fitted to each simulated network and the resulting parameter estimates, denoted by

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

are collected. The sample variance of $\tilde{\boldsymbol{\theta}}_n^b$ can then be used as an approximation for the variance of $\tilde{\boldsymbol{\theta}}_n$. Hence the $(1-\varepsilon)$-confidence interval for $\boldsymbol{\theta}$, assuming asymptotic normality, can be approximated by

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

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

5.3. Sensitivity test

Now we consider the sensitivity our estimates while values of the parameters $(n, \alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$ are allowed to vary. We first consider the impact of n, the number of edges in the network, holding the other parameters fixed with values given by (5.1). Here we take $n = 500, 1000, 5000, 10^4, 10^5$ and QQ-plots (not presented) for standardized estimates using both full MLE and one-snapshot methods were produced to check the asymptotic normality. It turns out that when n is relatively small, e.g. in our case $n = 500, 1000$, diagnostics reveal departures from normality for either MLE or snapshot estimates. However, after increasing n to 10^4, estimates obtained from both approaches appear normally distributed as expected.

n	$\hat{\boldsymbol{\theta}}_n^{MLE}$	$\tilde{\boldsymbol{\theta}}_n$	$ARE(\tilde{\boldsymbol{\theta}}_n)$
500	(0.301, 0.499, 2.122, 1.092)	(0.312, 0.501, 2.703, 1.354)	(0.187, 1.000, 0.088, 0.110)
1000	(0.301, 0.500, 2.074, 1.048)	(0.307, 0.501, 2.335, 1.175)	(0.185, 1.000, 0.139, 0.143)
5000	(0.300, 0.500, 2.020, 1.012)	(0.302, 0.500, 2.078, 1.040)	(0.198, 1.000, 0.176, 0.163)
10^4	(0.300, 0.500, 2.011, 1.006)	(0.301, 0.500, 2.042, 1.020)	(0.194, 1.000, 0.181, 0.166)
10^5	(0.300, 0.500, 2.002, 1.001)	(0.300, 0.500, 2.005, 1.002)	(0.195, 1.000, 0.178, 0.171)

Table 5.1

Full MLE and one-snapshot estimates, as well as ARE's of $\tilde{\boldsymbol{\theta}}_n$, for $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$ under different choices of n.

Numerical results are summarized in Table 5.1. For each value of n, we generate 5000 replicates of the network with n edges and parameters $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$. The means of MLE's $\hat{\boldsymbol{\theta}}_n^{MLE}$ (second column) are computed using the full history of these 5000 realizations, while the averages of one-snapshot estimates $\tilde{\boldsymbol{\theta}}_n$ (third column) are obtained using the snapshot method proposed in Section 4, pretending that we only have the last snapshot $G(n)$ available.

We then compute the variances of $\hat{\boldsymbol{\theta}}_n^{MLE}$ and $\tilde{\boldsymbol{\theta}}_n$. The ARE's of these estimates are approximated in the same way as (5.4):

$$ARE\Big(\big(\tilde{\boldsymbol{\theta}}_n\big)_i\Big) \approx \frac{\widehat{\text{Var}}\Big(\big(\hat{\boldsymbol{\theta}}_n^{MLE}\big)_i\Big)}{\widehat{\text{Var}}\Big(\big(\tilde{\boldsymbol{\theta}}_n\big)_i\Big)}, \qquad i = 1, 2, 3, 4. \tag{5.5}$$

We can see from Table 5.1 that both the accuracy and relative efficiency of $\tilde{\boldsymbol{\theta}}_n$ have been improved with the increase in n. Note that the ARE of $\tilde{\beta}$ is always 1 since $\tilde{\beta}$ is equal to $\hat{\beta}^{MLE}$.

(α,β)	$\hat{\boldsymbol{\theta}}_n^{MLE}$	$\tilde{\boldsymbol{\theta}}_n$	$ARE(\tilde{\boldsymbol{\theta}}_n)$
(0.001, 0.99)	(0.001, 0.990, 2.036, 1.016)	(0.001, 0.990, 2.095, 1.120)	(0.284, 1.000, 0.135, 0.049)
(0.01, 0.9)	(0.010, 0.900, 2.005, 1.002)	(0.010, 0.900, 2.012, 1.012)	(0.336, 1.000, 0.186, 0.062)
(0.1, 0.8)	(0.100, 0.800, 2.003, 1.001)	(0.100, 0.800, 2.007, 1.005)	(0.223, 1.000, 0.156, 0.128)
(0.2, 0.6)	(0.200, 0.600, 2.002, 1.001)	(0.200, 0.600, 2.011, 1.005)	(0.239, 1.000, 0.195, 0.143)
(0.5, 0.3)	(0.500, 0.300, 2.001, 1.001)	(0.500, 0.300, 2.010, 1.004)	(0.130, 1.000, 0.177, 0.206)
(0.7, 0.2)	(0.700, 0.200, 2.001, 1.002)	(0.701, 0.200, 2.010, 1.006)	(0.068, 1.000, 0.158, 0.316)
(0.9, 0.01)	(0.900, 0.010, 2.001, 1.006)	(0.901, 0.010, 2.010, 1.017)	(0.022, 1.000, 0.148, 0.223)

TABLE 5.2

For different values of (α,β), mean values of both MLE and one-snapshot estimates are computed, based on 5000 independent replicates. The ARE of the one-snapshot estimates are calculated according to (5.5). Other parameters are kept unchanged, i.e. $(n,\delta_{in},\delta_{out}) = (10^5, 2, 1)$.

Next we experimented with different values of (α,β), holding $(n,\delta_{\text{in}},\delta_{\text{out}}) = (10^5, 2, 1)$. Again, similar to Table 5.1, for each choice of (α,β), we generate 5000 independent realizations of the network. The second column summarizes the mean values of the full MLE's $\hat{\boldsymbol{\theta}}_n^{MLE}$ from the 5000 replicates, while those of one-snapshot estimates $\tilde{\boldsymbol{\theta}}_n$ are given in the third column. The ARE's of $\tilde{\boldsymbol{\theta}}_n$ are calculated as in (5.5). Corresponding results are presented in Table 5.2. Overall, the full MLE's are more accurate than the snapshot estimates, even if the values of (α,β) are relatively small. From the last column in Table 5.2, we see that as α increases, the snapshot estimate $\tilde{\alpha}$ becomes less efficient. However, the efficiency of $(\tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}})$ improves.

6. 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 $(\alpha, \beta, \gamma, \xi, \delta_{\text{in}}, \delta_{\text{out}})$, and ρ is implicitly defined through $\rho = 1 - (\alpha + \beta + \gamma + \xi)$. Similar to the derivations in Section 3, the MLE estimators for $\alpha, \beta, \gamma, \xi$ are

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

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

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

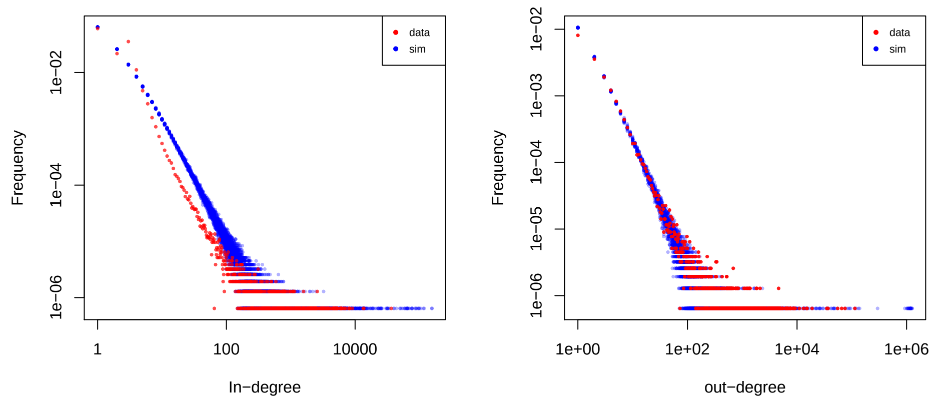


FIG 6.1. *Empirical in- and out-degree frequencies of the full Wiki talk network (red) and that from 20 realizations of the linear preferential attachment network with fitted parameter values (6.1) from MLE (blue). The scatter plots for the degree frequencies from the 20 simulations are overlaid together to form a 'plausible' range for the degree distribution of the fitted model*

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

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

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\xi}, \hat{\rho}, \hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}}) = \\ (3.08 \times 10^{-3}, 8.55 \times 10^{-1}, 1.39 \times 10^{-1}, 4.76 \times 10^{-5}, 3.06 \times 10^{-3}, 0.547, 0.134). \tag{6.1}$$

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 6.1, we see that while the data roughly agrees with the simulations in the out-degree frequencies, the deviation in the in-degree frequencies is noticeable.

To better understand the discrepancy in the in-degree frequencies, we examined the link data and their timestamps and discovered bursts of messages originating from certain nodes over small time intervals. According to Wikipedia policy [**?**], certain administrating accounts are allowed to send group messages to multiple users simultaneously. These bursts presumably represent broadcast announcements generated from these accounts. These administrative broadcasts can also be detected 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

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

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 6.2. 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.

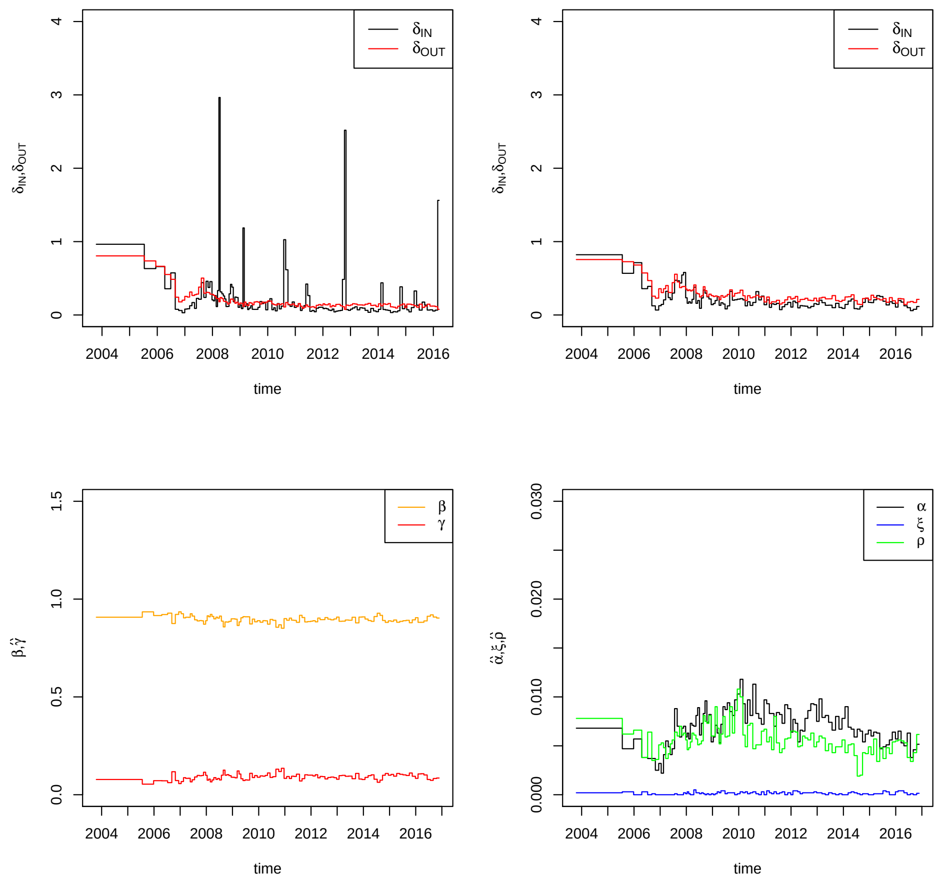


FIG 6.2. *Local parameter estimates of the linear preferential attachment model for the full and reduced Wiki talk network. Upper left:* $(\hat{\delta}_{in}, \hat{\delta}_{out})$ *for the full network. Upper right, lower left, lower right:* $(\hat{\delta}_{in}, \hat{\delta}_{out})$, $(\hat{\beta}, \hat{\gamma})$, $(\hat{\alpha}, \hat{\xi}, \hat{\rho})$ *for the reduced network, respectively.*

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 6.2. 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

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

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

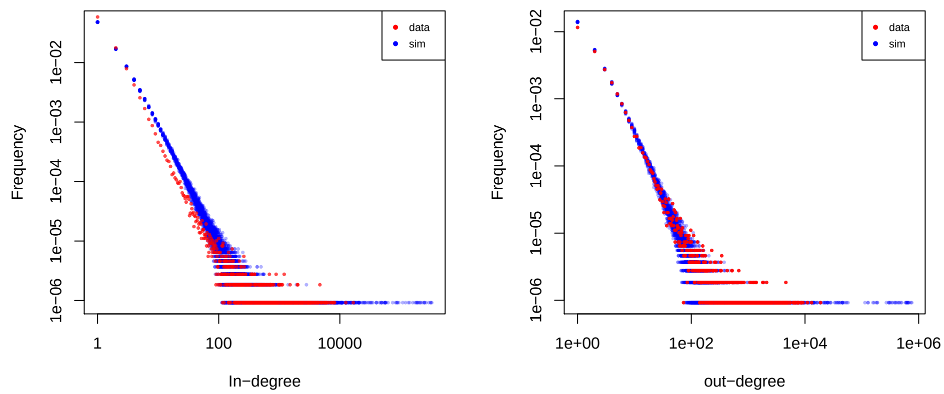


FIG 6.3. *Empirical in- and out-degree frequencies of the reduced Wiki talk network (red) and that from 20 realizations of the linear preferential attachment network with fitted parameter values* (6.2) *from MLE (blue).*

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

Next we consider the estimation method of Section 4 applied to a single snapshot of the data. In order to implement this procedure, we don blinders and assume that our data set consists only of the information of the wiki data at the last timestamp. That is, information about administrative broadcasts, and other aspects of the data learned by looking at the previous history of the data are unavailable. In particular, we would have no knowledge of the existence of the two additional scenarios interaction schemes corresponding to $J_n = 4, 5$. With this in mind, we fit the three scenario model using the methods in Section 4. The fitted parameters are

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

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

We see from this example that while the linear preferential attachment model is perhaps too simplistic for the Wiki talk network dataset, it has the ability to illuminate some gross features, such as the out-degrees, as well as to capture important structural changes such as the group message behavior. 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 change points in parameters is a direction for future research.

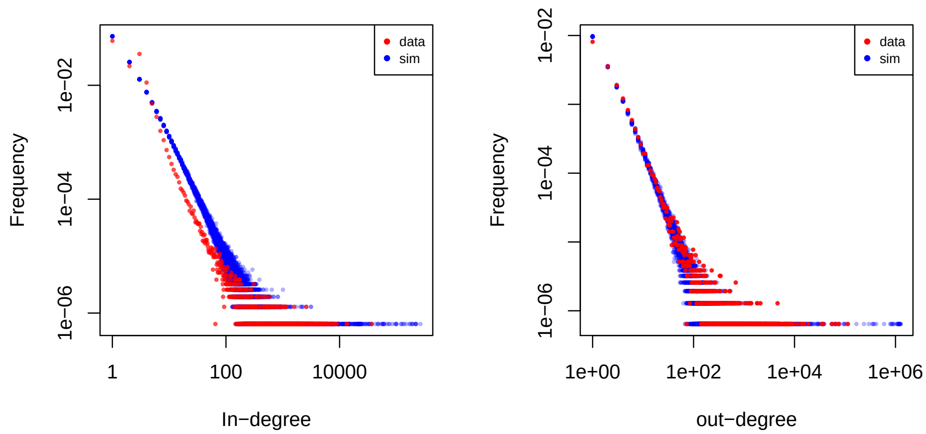


Fig 6.4. *Empirical in- and out-degree frequencies of the full Wiki talk network (red) and that from 20 realizations of the linear preferential attachment network with fitted parameter values* (6.3) *from the snapshot estimator (blue).*

7. Acknowledgement

Research of the four authors was partially supported by Army MURI grant W911NF-12-1-0385. Don Towsley from 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. We appreciate the many helpful and sensible comments of the referees and editors.

Appendix A: Proofs

A.1. Lemmas A.1 and A.2

Lemma A.1. *For $\lambda > 0$, the function $\psi(\lambda)$ in* (3.11) *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. [**?**]):

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

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

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

$$\vdots$$

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

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

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

which can be simplified to

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

From (2.4),

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

Hence by rearranging (A.2), we have

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

or equivalently,

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

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

$$\begin{aligned}
\psi(\lambda) &= \sum_{i=0}^{\infty}\frac{p_{>i}^{\mathrm{in}}(\delta_{\mathrm{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - (1-\beta)a_1(\lambda) \\
&= \sum_{i=0}^{\infty}\frac{p_{>i}^{\mathrm{in}}(\delta_{\mathrm{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - \sum_{i=0}^{\infty}\frac{p_i^{\mathrm{in}}(\delta_{\mathrm{in}})a_1(\lambda)(i+\lambda)}{i+\lambda} \\
&= \sum_{i=0}^{\infty}\frac{a_1(\delta_{\mathrm{in}})(i+\delta_{\mathrm{in}})p_i^{\mathrm{in}}(\delta_{\mathrm{in}}) + \gamma\mathbf{1}_{\{i=0\}}}{i+\lambda} - \frac{\gamma}{\lambda} - \sum_{i=0}^{\infty}\frac{p_i^{\mathrm{in}}(\delta_{\mathrm{in}})a_1(\lambda)(i+\lambda)}{i+\lambda} \\
&= \sum_{i=0}^{\infty}\frac{p_i^{\mathrm{in}}(\delta_{\mathrm{in}})}{i+\lambda}\Big(a_1(\delta_{\mathrm{in}})(i+\delta_{\mathrm{in}}) - a_1(\lambda)(i+\lambda)\Big) \\
&= \sum_{i=0}^{\infty}\frac{p_i^{\mathrm{in}}(\delta_{\mathrm{in}})}{i+\lambda}\int_{\lambda}^{\delta_{\mathrm{in}}}\frac{\partial}{\partial s}\Big(a_1(s)(i+s)\Big)ds \\
&= \sum_{i=0}^{\infty}\frac{p_i^{\mathrm{in}}(\delta_{\mathrm{in}})}{i+\lambda}\int_{\lambda}^{\delta_{\mathrm{in}}}\frac{(\alpha+\beta)(1-i(1-\beta))}{(1+s(1-\beta))^2}ds \\
&= \left(\sum_{i=0}^{\infty}\frac{p_i^{\mathrm{in}}(\delta_{\mathrm{in}})}{i+\lambda}(1-i(1-\beta))\right)\int_{\lambda}^{\delta_{\mathrm{in}}}\frac{\alpha+\beta}{(1+s(1-\beta))^2}ds \\
&=: C(\lambda)\int_{\lambda}^{\delta_{\mathrm{in}}}\frac{\alpha+\beta}{(1+s(1-\beta))^2}ds.
\end{aligned} \tag{A.5}$$

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

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

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

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

The infinite series converge because $p_i^{\text{in}}(\lambda)$ is a power law with index greater than 2; see (2.5) and (2.6). Solving for the infinite series we get

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

Hence we have

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

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

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

where $C(\lambda) > 0$ for all $\lambda > 0$. Therefore $\psi(\cdot)$ has a unique zero at 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 A.2. *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,

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

$$+\sup_{\lambda\geq\epsilon}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{N(t-1)}{t-1+\lambda N(t-1)}\mathbf{1}_{\{J_t\in\{1,2\}\}}-\frac{(1-\beta)(\alpha+\beta)}{1+\lambda(1-\beta)}\right|. \tag{A.7}$$

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

$$iN^{\text{in}}_{>i}(n)=\sum_{k=i+1}^{\infty}N^{\text{in}}_{k}(n)i\leq\sum_{k=1}^{\infty}kN^{\text{in}}_{k}(n)=n,$$

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

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

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 (2.4), we have

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

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

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

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

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

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

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

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

Further, by the strong law of large numbers,

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

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

A.2. Lemmas A.3 and A.4

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

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

Proof. Let $\mathcal{F}_n=\sigma(G(0),\ldots,G(n))$ be the 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 (3.17) 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 $\left\{n^{-1/2}\sum_{r=1}^{t}u_r(\delta_{\text{in}})\right\}_{t=1,\ldots,n}$ is a zero-mean, square-integrable martingale array. The convergence (A.8) follows from the martingale central limit theory (cf. Theorem 3.2 of [**?**]) 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 $|u_t(\delta_{\text{in}})| \le 2/\delta_{\text{in}}$, we have

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

and

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

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 A.2, we have for T_1,

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

We then rewrite T_2 as

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

where

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

by Cesàro's convergence and

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

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

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

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

Combining these results together,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^n u_t^2(\delta_{\rm in}) &= T_1 - 2(T_{21}+T_{22}) + T_3 \\ &\xrightarrow{p} \sum_{i=0}^\infty \frac{p^{\rm in}_{>i}}{(i+\delta_{\rm in})^2} - \frac{\gamma}{\delta_{\rm in}^2} - \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{\rm in}(1-\beta))^2} = I_{\rm in}. \end{aligned} \tag{A.9}$$

This completes the proof. □

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

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

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

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

and then

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

We first observe that

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

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

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

Also from (A.9),

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

Hence

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

and (A.10) is established.

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

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

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

Then

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

and

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

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

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

We have

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

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

A.3. Proof of Theorem 4.1

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

$$
\sum_{i=0}^\infty\frac{N^{\text{in}}_{>i}(n)}{n} = \sum_{i=0}^\infty\frac{iN^{\text{in}}_i(n)}{n} = 1.
$$

We can re-write (4.4a) as

$$\begin{aligned}\alpha+\tilde{\beta} &= \left(\frac{1}{\delta_{\text{in}}}-\sum_{i=0}^{\infty}\frac{N^{\text{in}}_{>i}(n)/n}{i+\delta_{\text{in}}}\right)\Bigg/\left(\frac{1}{\delta_{\text{in}}}-\frac{1-\tilde{\beta}}{1+\delta_{\text{in}}(1-\tilde{\beta})}\right)\\ &= \left(\sum_{i=0}^{\infty}\frac{N^{\text{in}}_{>i}(n)/n}{\delta_{\text{in}}}-\sum_{i=0}^{\infty}\frac{N^{\text{in}}_{>i}(n)/n}{i+\delta_{\text{in}}}\right)\Bigg/\left(\frac{1}{\delta_{\text{in}}(1+\delta_{\text{in}}(1-\tilde{\beta}))}\right)\\ &= \sum_{i=1}^{\infty}\frac{N^{\text{in}}_{>i}(n)}{n}\frac{i}{i+\delta_{\text{in}}}\left(1+\delta_{\text{in}}(1-\tilde{\beta})\right) =:\ f_n(\delta_{\text{in}}),\end{aligned}\tag{A.12}$$

and (4.4b) as

$$\alpha+\tilde{\beta} = \left(\frac{N^{\text{in}}_0(n)}{n}+\tilde{\beta}\right)\Bigg/\left(1-\frac{N^{\text{in}}_0(n)}{n}\frac{\delta_{\text{in}}}{1+(1-\tilde{\beta})\delta_{\text{in}}}\right) =:\ g_n(\delta_{\text{in}}).$$

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:

$$\begin{aligned}f(\delta) &:= \sum_{i=1}^{\infty}p^{\text{in}}_{>i}\frac{i}{i+\delta}(1+\delta(1-\beta)),\\ g(\delta) &:= \left(p^{\text{in}}_0+\beta\right)\Bigg/\left(1-p^{\text{in}}_0\frac{\delta}{1+(1-\beta)\delta}\right), \qquad \delta\in[\epsilon,K].\end{aligned}$$

Now we apply the re-parametrization

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

to f and g, such that

$$\begin{aligned}\tilde{f}(\eta) &:= f(\delta(\eta)) = \sum_{i=1}^{\infty}\frac{p^{\text{in}}_{>i}}{1+(i^{-1}-(1-\beta))\eta},\\ \tilde{g}(\eta) &:= g(\delta(\eta)) = \frac{p^{\text{in}}_0+\beta}{1-\eta p^{\text{in}}_0}.\end{aligned}$$

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 $i\ge 1$. So $\tilde{f}(\eta)>0$ on $\mathcal{I}$;
- $\tilde{f}(\eta)\le\frac{1}{1-(1-\beta)\eta}\sum_{i=0}^{\infty}p^{\text{in}}_{>i}\le 1+(1-\beta)K<\infty$.

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

$$1/p^{\text{in}}_0 > 1/(1-\beta) > \eta.\tag{A.14}$$

The first inequality holds since:

$$\begin{aligned}1/p^{\text{in}}_0 > 1/(1-\beta) &\Leftrightarrow p^{\text{in}}_0 < 1-\beta\\ &\Leftrightarrow \frac{\alpha}{1+\frac{(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}}} < 1-\beta\end{aligned}$$

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

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

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

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

then it follows that

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

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

First observe that

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

We now claim that

$$\frac{\partial}{\partial\eta}\left(\sum_{i=1}^{\infty}\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right) = \sum_{i=1}^{\infty}\frac{\partial}{\partial\eta}\left(\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right) = -\sum_{i=1}^{\infty}\frac{p^{\text{in}}_{>i}(i^{-1}-(1-\beta))}{(1+b_i(\eta))^2}, \tag{A.16}$$
$$\frac{\partial^2}{\partial\eta^2}\left(\sum_{i=1}^{\infty}\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right) = \sum_{i=1}^{\infty}\frac{\partial^2}{\partial\eta^2}\left(\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right) = 2\sum_{i=1}^{\infty}\frac{p^{\text{in}}_{>i}(i^{-1}-(1-\beta))^2}{(1+b_i(\eta))^3}. \tag{A.17}$$

It suffices to check:

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

Note that for $i\ge 1$,

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

Recall (A.6), we then have

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

Hence,

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

$$= (2-\beta)(1+(1-\beta)K)^2 < \infty,$$

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

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

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

From Lemma A.1, $\psi(\delta_{\text{in}}) = 0$ where $\psi(\cdot)$ is as defined in (3.11). Hence we have $f(\delta_{\text{in}}) = \alpha+\beta$ in a similar derivation to that of (A.12). Also from (4.2), 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 (A.13), 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 A.2, we have

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

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

□