Adv. Appl. Probab. **54,** 718–742 (2022)
*doi:*10.1017/apr.2021.52

MEASURING RECIPROCITY IN A DIRECTED PREFERENTIAL ATTACHMENT NETWORK

TIANDONG WANG,* *Texas A&M University*
SIDNEY I. RESNICK,** *Cornell University*

Abstract

Empirical studies (e.g. Jiang *et al.* (2015) and Mislove *et al.* (2007)) show that online social networks have not only in- and out-degree distributions with Pareto-like tails, but also a high proportion of reciprocal edges. A classical directed preferential attachment (PA) model generates in- and out-degree distributions with power-law tails, but the theoretical properties of the reciprocity feature in this model have not yet been studied. We derive asymptotic results on the number of reciprocal edges between two fixed nodes, as well as the proportion of reciprocal edges in the entire PA network. We see that with certain choices of parameters, the proportion of reciprocal edges in a directed PA network is close to 0, which differs from the empirical observation. This points out one potential problem of fitting a classical PA model to a given network dataset with high reciprocity, and indicates that alternative models need to be considered.

Keywords: Reciprocity; preferential attachment; in- and out-degrees

2020 Mathematics Subject Classification: Primary 60F15; 60G46
Secondary 05C82

1. Introduction

In social network analysis, reciprocal edges characterize communication between two users. For instance, on Facebook, one user leaves messages on another user's wall page, and a response from the target user then creates a reciprocal edge. *Reciprocity*, which is classically defined as the proportion of reciprocal edges (cf. [12, 21]), is one important network metric to measure interactions among individual users. The directed network constructed from Facebook wall posts [17] is one example of social networks with a large proportion of reciprocal edges. The study on eight different types of networks in [7] shows that online social networks (e.g. [3, 6, 9, 10, 17]) tend to have a higher proportion of reciprocal edges, compared to other types of networks such as biological networks, communication networks, software call graphs, and P2P networks.

Another widely observed feature of directed social networks is the scale-free property, where both in- and out-degree distributions have Pareto-like tails. The directed preferential attachment (PA) network model is appealing (cf. [2, 8]), since theoretically the directed PA

Received 12 March 2021; revision received 18 August 2021.

* Postal address: Department of Statistics, Texas A&M University, College Station, TX 77843, U.S.A. Email address: twang@stat.tamu.edu

** Postal address: School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853, U.S.A. Email address: sir1@cornell.edu

mechanism generates a network with the scale-free property because nodes with large degrees are likely to attract more edges than those with small degrees (cf. [14, 15, 18, 20]). However, the asymptotic behavior of the proportion of reciprocal edges in a directed PA model has not yet been explored in the literature. In this paper, we derive asymptotic results about (1) the number of reciprocal edges between two fixed nodes; (2) the proportion of reciprocal edges in a directed PA network, provided that the network has a large number of edges; and (3) the first time a reciprocal edge appears between two distinct fixed nodes.

Our theoretical results suggest that for certain choices of model parameters, especially when the proportion of edges added between two existing nodes is small, the proportion of reciprocal edges in the entire graph is close to 0, even though the total number of reciprocal edges between two fixed nodes may be of order $O(n^a)$, $a \in (0, 1)$. Such behavior flags potential problems for fitting a directed PA model in practice. When fitting a directed PA model to a real network with high reciprocity using existing methods developed in [18, 19], there is no guarantee that the calibrated model will also have a high proportion of reciprocal edges. Such a discrepancy indicates a poor fit of the PA model, since the fitted model fails to capture the important feature of high reciprocity. In these cases, variants of the directed PA model need to be considered. For instance, [4] provides several different ways to predict the reciprocal edges between two given nodes, and one may incorporate those features to construct a refined network model that is both scale-free and of high reciprocity.

The rest of the paper is organized as follows. Section 1.1 provides the notation and definitions necessary to specify a growing sequence of graphs that evolve according to PA. The definitions use model parameters that control the growth of power law sequences. In Section 2, we give the power law growth of the in- and out-degree sequences. This leads to the main results of the paper, in Section 3:

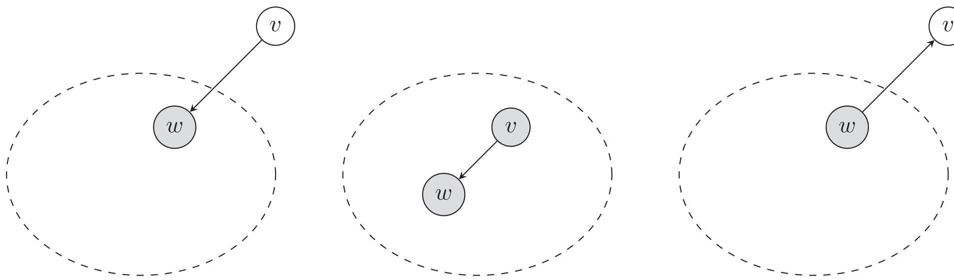
1. For fixed nodes i, j, the number of reciprocal edges between i and j evolves as a power law.
2. For specified subsets of the parameters, the proportion of reciprocal edges in the graph goes to 0, showing that many real data sets with significant reciprocity do not follow this model.
3. For fixed nodes i, j we provide information about the first time a reciprocal pair of edges forms between i and j.

Section 4 contains a short discussion of theoretical results and future research directions, and the appendix gives some lemmas and proofs.

1.1. Model setup

Here is the specification of the classical directed PA model. Initialize the model with the graph G(0), which consists of one node, labeled node 1, and a self-loop. After n steps in the construction, $G(n) = (V(n), E(n))$ is the graph with node set $V(n)$ with $V(0) = \{1\}$ and $|V(0)| = 1$ and set of directed edges $E(n)$ such that an ordered pair $(i, j) \in E(n)$, $i, j \in V(n)$, represents a directed edge $i \mapsto j$. When $n = 0$, we have $E(0) = \{(1, 1)\}$. For later use, self-loops are not counted for reciprocity.

Let $\left(D_v^{\text{in}}(n), D_v^{\text{out}}(n)\right)$ be the in- and out-degrees of node $v \in V(n)$ with the convention that if $v \notin V(n)$, $D_v^{\text{in}}(n) = 0 = D_v^{\text{out}}(n)$. From $G(n)$ to $G(n+1)$, one of three scenarios happens:



1. With probability α, we add a new edge (v, w), where $w \in V(n)$, and $v \notin V(n)$ is a new node. The existing node w is chosen with probability

$$\frac{D_w^{\text{in}}(n) + \delta_{\text{in}}}{\sum_{w \in V(n)} \left(D_w^{\text{in}}(n) + \delta_{\text{in}}\right)} = \frac{D_w^{\text{in}}(n) + \delta_{\text{in}}}{n + 1 + \delta_{\text{in}}|V(n)|}.$$

2. With probability β, a new edge (v, w) is added between two existing nodes $v, w \in V(n)$, where the starting node v and the ending node w are chosen independently with probability

$$\frac{D_w^{\text{in}}(n) + \delta_{\text{in}}}{\sum_{w \in V(n)} \left(D_w^{\text{in}}(n) + \delta_{\text{in}}\right)} \frac{D_v^{\text{out}}(n) + \delta_{\text{out}}}{\sum_{v \in V(n)} \left(D_v^{\text{out}}(n) + \delta_{\text{out}}\right)}$$
$$= \frac{D_w^{\text{in}}(n) + \delta_{\text{in}}}{n + 1 + \delta_{\text{in}}|V(n)|} \frac{D_v^{\text{out}}(n) + \delta_{\text{out}}}{n + 1 + \delta_{\text{out}}|V(n)|}.$$

   For brevity of notation, we set

$$A_{wv}(n) := \frac{D_w^{\text{in}}(n) + \delta_{\text{in}}}{n + 1 + \delta_{\text{in}}|V(n)|} \frac{D_v^{\text{out}}(n) + \delta_{\text{out}}}{n + 1 + \delta_{\text{out}}|V(n)|}; \tag{1.1}$$

   then the attachment probability in the β-scheme is $\beta A_{wv}(n)$.

3. With probability γ, we add a new edge (w, v), where $w \in V(n)$, and $v \notin V(n)$ is a new node. The existing node w is chosen with probability

$$\frac{D_w^{\text{out}}(n) + \delta_{\text{out}}}{\sum_{w \in V(n)} \left(D_w^{\text{out}}(n) + \delta_{\text{out}}\right)} = \frac{D_w^{\text{out}}(n) + \delta_{\text{out}}}{n + 1 + \delta_{\text{out}}|V(n)|}.$$

We assume $\alpha + \beta + \gamma = 1$, $\beta \in [0, 1)$, and $\delta_{\text{in}}, \delta_{\text{out}} > 0$. Owing to the α- and γ-schemes, $|V(n)| - 1$ follows a binomial distribution with size n and success probability $\alpha + \gamma = 1 - \beta$, so that $|V(n)| \xrightarrow{\text{a.s.}} \infty$ as $n \to \infty$. For $v \geq 1$, we define S_v to be the time when node v is created, i.e.

$$S_v := \inf\left\{n \geq 0 : |V(n)| = v\right\}. \tag{1.2}$$

Since we use the convention that $D_v^{\text{in}}(n) = 0$ and $D_v^{\text{out}}(n) = 0$ if $S_v > n$, we have by (1.1) that $A_{vw}(n) \in [0, 1]$ for all n.

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

Let $\mathcal{F}_n$ be the σ-field generated by observing the network up to the creation of the nth new edge. Suppose τ is a stopping time with respect to the filtration $\left(\mathcal{F}_n\right)_{n\geq 0}$; then

$$\mathcal{F}_\tau = \{F : F \cap \{\tau = n\} \in \mathcal{F}_n\}.$$

By (1.2), we see that S_v is a stopping time with respect to $\left(\mathcal{F}_n\right)_{n\geq 0}$. For $n \geq k \geq 0$, we have

$$\{S_v + k = n\} = \{S_v = n - k\} \in \mathcal{F}_{n-k} \subset \mathcal{F}_n,$$

so $S_v + k$, $k \geq 0$, is a stopping time with respect to $\left(\mathcal{F}_n\right)_{n\geq 0}$. Note that for $v > n - k$, $\{S_v = n - k\} = \emptyset \in \mathcal{F}_{n-k} \subset \mathcal{F}_n$. In what follows, we write $\mathbb{E}^{\mathcal{F}_n}(\,\cdot\,) := \mathbb{E}\big(\cdot|\mathcal{F}_n\big)$.

2. Growth of the degree sequences $D_v^{\text{in}}(n)$, $D_v^{\text{out}}(n)$

The parameters controlling the behavior of the model are $\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}}$, and we now define two new parameters which will serve as power-law exponents:

$$c_1 = \frac{\alpha + \beta}{1 + \delta_{\text{in}}(1 - \beta)}, \qquad c_2 = \frac{\beta + \gamma}{1 + \delta_{\text{out}}(1 - \beta)}. \tag{2.1}$$

From (2.1), $0 < c_1 < \alpha + \beta$, $0 < c_2 < \beta + \gamma$, and $0 < c_1 + c_2 < 1 + \beta$. In fact, we can reparametrize the PA model using $(\alpha, \beta, \gamma, c_1, c_2)$, which leads to an estimation method for model parameters alternative to unavailable maximum likelihood techniques; see [19] for details.

The following proposition summarizes the power-law growth of $\big(D_v^{\text{in}}(n), D_v^{\text{out}}(n)\big)$, which is controlled by the parameters c_1, c_2. From these growth rates, we will derive the limiting behavior of the number of reciprocal edges between two fixed nodes in Section 3.1.

Proposition 2.1. *For $v \geq 1$, there are random variables ξ_v^{in}, ξ_v^{out} satisfying $\mathbb{P}\big(\xi_v^{in} \in (0, \infty)\big) = 1 = \mathbb{P}\big(\xi_v^{out} \in (0, \infty)\big)$, such that*

$$\frac{D_v^{in}(n)}{n^{c_1}} \xrightarrow{a.s.} \xi_v^{in}, \qquad \frac{D_v^{out}(n)}{n^{c_2}} \xrightarrow{a.s.} \xi_v^{out}.$$

The proof of Proposition 2.1 is given in Appendix A.3, after the statement and proof of two lemmas.

3. Reciprocity in the preferential attachment network

In this section, we focus on the asymptotic behavior of the number of reciprocal edges between two fixed, distinct nodes i and j as well as that of the proportion of reciprocal edges in the entire graph. We also consider the first time a reciprocal edge forms between two distinct nodes.

To assess goodness of fit of the directed PA model to a particular dataset, it is useful to evaluate statistics to see whether the empirical values match those from the fitted model. The statistic we focus on here is *reciprocity*. If there is a significant discrepancy between the reciprocity measure for the fitted theoretical PA model and that of the actual network data, then we conclude that variants of the classical PA model should be considered.

Given a directed graph $G = (V, E)$, let $L_{(i,j)} = L_{(i,j)}(G)$ be the number of directed edges (i, j) in the graph G, for $i \neq j$. Then define the *reciprocity coefficient*, $R = R(G)$, as

$$R = \frac{2}{|E|} \sum_{i,j\in V: i<j} \min\left\{L_{(i,j)}, L_{(j,i)}\right\}. \tag{3.1}$$

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

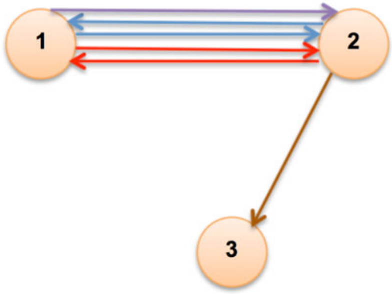


FIGURE 1. A graphical illustration of reciprocal edges with edge set $E = \{(1, 2), (1, 2), (1, 2), (2, 1), (2, 1), (2, 3)\}$ and $V = \{1, 2, 3\}$. If a pair of reciprocal edges are observed, we label them with the same color. Here we have $R_6 = 4/6 = 0.667$.

Note that a node pair can be counted more than once but self-loops are never counted. For example, consider the graph given in Figure 1, where there are $|E| = 6$ edges and node set $V = \{1, 2, 3\}$. We distinguish multiple edges between two nodes by different colors, and if a pair of reciprocal edges are observed, we label the pair with the same color. The graph in Figure 1 contains a pair of blue edges and a pair of red edges, thus giving $R_6 = 4/6 = 0.667$. In `R`, we can easily compute the reciprocity coefficient by applying the `dyad_census()` function in the `igraph` package to the graph object, and its `mut` value outputs the total number of unordered node pairs $\{i, j\}$ with reciprocal connections (i, j) and (j, i), allowing multiplicity.

Consider a sequence of graphs $\{G(n) = (V(n), E(n)), n \geq 1\}$ constructed following the PA rule with parameters $(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}})$ as outlined in Section 1.1. For two nodes $i \neq j \in V(n)$, write $L_{(i,j)}(n) = L_{(i,j)}(G(n))$, and write the reciprocity coefficient for the PA network $G(n)$ as $R_n^{\text{pa}} := R(G(n))$. We emphasize that the definition in (3.1) excludes self-loops when counting reciprocal edges. In Sections 3.1 and 3.2, we study the asymptotic behavior of $L_{(i,j)}(n)$ for fixed $i \neq j$, and R_n^{pa} in a PA network $G(n)$, respectively.

3.1. Reciprocal edges between two fixed nodes

In a directed PA network, the total number of reciprocal edges between two fixed nodes i, j is equal to

$$L_{i\leftrightarrow j}(n) := 2 \min \{L_{(i,j)}(n), L_{(j,i)}(n)\} .$$

We first study the limiting behavior of the number of edges between two fixed nodes $i \neq j$, $L_{(i,j)}(n)$, when n is large. The asymptotics of $L_{i\leftrightarrow j}(n)$ then follow from a continuous mapping argument, and this leads in Section 3.2 to a study of the asymptotic behavior of R_n^{PA}. The asymptotic behavior of $L_{(i,j)}(n)$ also assists in the study, in Section 3.3, of the behavior of the first time a reciprocal pair is formed between i and j.

3.1.1. *Convergence of* $L_{(i,j)}(n)$. Theorem 3.1 gives the main asymptotic results, but we start by presenting a lemma on $A_{ij}(n)$ which is useful for the proof of Theorem 3.1.

Lemma 3.1. *Recall the definition of* $A_{ij}(n)$ *in* (1.1) *and the notation in Proposition* 2.1. *For fixed* $1 \leq i < j$,

$$\frac{A_{ij}(n)}{n^{c_1+c_2-2}} \longrightarrow \frac{1}{(1 + \delta_{in}(1-\beta))(1 + \delta_{out}(1-\beta))} \xi_j^{in} \xi_i^{out}, \quad \textit{almost surely (a.s.) and in } L_1. \tag{3.2}$$

This further gives the following:

1. *When* $c_1 + c_2 > 1$*, there exist constants* $C_{ij} > 0$ *such that*

$$\frac{c_1 + c_2 - 1}{n^{c_1+c_2-1}} \sum_{k=S_j}^{S_j+n} A_{ij}(k) \longrightarrow C_{ij}\xi_j^{in}\xi_i^{out}, \qquad a.s.\ and\ in\ L_1. \tag{3.3}$$

2. *When* $c_1 + c_2 = 1$*, there exist constants* $C'_{ij} > 0$ *such that*

$$\frac{1}{\log n} \sum_{k=S_j}^{S_j+n} A_{ij}(k) \longrightarrow C'_{ij}\xi_j^{in}\xi_i^{out}, \qquad a.s.\ and\ in\ L_1. \tag{3.4}$$

3. *When* $c_1 + c_2 < 1$*, there exist constants* $C''_{ij} > 0$ *such that*

$$\frac{1}{n^{c_1+c_2-1}} \sum_{k=n}^{\infty} A_{ij}(k) \longrightarrow C''_{ij}\xi_j^{in}\xi_i^{out}, \qquad a.s.\ and\ in\ L_1, \tag{3.5}$$

   which further implies $\sum_{k\geq 1} \mathbb{E}(A_{ij}(k)) < \infty$ *and* $\sum_{k\geq 1} A_{ij}(k) < \infty$ *a.s.*

In addition, we have similar convergence results for $A_{ji}(n)$ *by replacing* $A_{ij}(n)$*,* ξ_j^{in}*, and* ξ_i^{out} *with* $A_{ji}(n)$*,* ξ_i^{in}*, and* ξ_j^{out}*, respectively.*

Proof. By Lemma 2.1 as well as the fact that $|V(n)|/n \xrightarrow{\text{a.s.}} 1 - \beta$, we have

$$\frac{A_{ij}(n)}{n^{c_1+c_2-2}} = \frac{\big(D_i^{\text{out}}(n) + \delta_{\text{out}}\big)\big(D_j^{\text{in}}(n) + \delta_{\text{in}}\big)}{n^{c_1+c_2}} \frac{n^2}{\big(n + 1 + \delta_{\text{in}}|V(n)|\big)\big(n + 1 + \delta_{\text{out}}|V(n)|\big)}$$
$$\xrightarrow{\text{a.s.}} \frac{1}{\big(1 + \delta_{\text{in}}(1-\beta)\big)\big(1 + \delta_{\text{out}}(1-\beta)\big)} \xi_j^{\text{in}}\xi_i^{\text{out}}.$$

Note that once we show

$$\sup_{n\geq 1} \mathbb{E}\left[\left(\frac{A_{ij}(n)}{n^{c_1+c_2-2}}\right)^2\right] < \infty, \tag{3.6}$$

we have by [5, Theorem 4.6.2] that $\{A_{ij}(n)/n^{c_1+c_2-2} : n \geq 1\}$ is uniformly integrable, which gives the L_1-convergence in (3.2).

To prove (3.6), we now use the Cauchy–Schwarz inequality to obtain

$$\mathbb{E}\left[\left(\frac{A_{ij}(n)}{n^{c_1+c_2-2}}\right)^2\right] \leq \mathbb{E}\left[\frac{\big(D_i^{\text{out}}(n) + \delta_{\text{out}}\big)^2\big(D_j^{\text{in}}(n) + \delta_{\text{in}}\big)^2}{n^{2(c_1+c_2)}}\right]$$
$$\leq \frac{1}{n^{2(c_1+c_2)}} \left(\mathbb{E}\left[\big(D_i^{\text{out}}(n) + \delta_{\text{out}}\big)^4\right] \mathbb{E}\left[\big(D_i^{\text{in}}(n) + \delta_{\text{in}}\big)^4\right]\right)^{1/2}.$$

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

Then by Lemma A.2 we have

$$\sup_{n\geq 1}\mathbb{E}\left[\left(\frac{A_{ij}(n)}{n^{c_1+c_2-2}}\right)^2\right]\leq \sup_{n\geq 1}\frac{\sqrt{\mathbb{E}\left[\left(D_i^{\text{out}}(n)+\delta_{\text{out}}\right)^4\right]}}{n^{2c_2}}\times\sup_{n\geq 1}\frac{\sqrt{\mathbb{E}\left[\left(D_j^{\text{in}}(n)+\delta_{\text{in}}\right)^4\right]}}{n^{2c_1}}<\infty.$$

With (3.2) established, the results in (3.3)–(3.5) follow directly from Karamata's theorem (cf. [13, Theorem 2.1]).

The results for $A_{ji}(n)$ follow from similar reasoning. □

We now give the asymptotic behavior of $L_{(i,j)}(n)$ in a directed PA model.

Theorem 3.1. *Consider two fixed nodes $1\leq i<j$. We have the following:*

(i) If $c_1+c_2>1$, then there exists some random variable ξ_{ij}, satisfying $\mathbb{P}\big(\xi_{ij}\in(0,\infty)\big)=1$, such that as $n\to\infty$,

$$\frac{L_{(i,j)}(n)}{n^{c_1+c_2-1}}\xrightarrow{a.s.}\xi_{ij}. \tag{3.7}$$

So for large n, the number of (i, j) edges is of order $n^{c_1+c_2-1}$.

(ii) If $c_1+c_2=1$, then there exists some random variable ζ_{ij}, satisfying $\mathbb{P}\big(\zeta_{ij}\in(0,\infty)\big)=1$, such that

$$\frac{L_{(i,j)}(n)}{\log n}\xrightarrow{a.s.}\zeta_{ij}. \tag{3.8}$$

So for large n, the number of (i, j) edges is of order $\log n$.

(iii) If $c_1+c_2<1$, then for any $i<j$, there is a last time for an (i, j) edge to form. Furthermore, as $n\to\infty$,

$$L_{(i,j)}(n)\uparrow L_{(i,j)}(\infty)<\infty,\qquad a.s., \tag{3.9}$$

and $L_{(i,j)}(n)-L_{(i,j)}(\infty)=0$ a.s. for n large.

In addition, we obtain similar convergence results for $L_{(j,i)}(n)$ by replacing $L_{(i,j)}(n)$, $L_{(i,j)}(\infty)$, ξ_{ij}, and ζ_{ij} with $L_{(j,i)}(n)$, $L_{(j,i)}(\infty)$, ξ_{ji}, and ζ_{ji}, respectively.

Proof. Set

$$\Delta_k(i,j)=\mathbf{1}_{\left\{E(k)=E(k-1)\cup\{(i,j)\}\right\}};$$

i.e. $\Delta_k(i,j)=1$ if a directed edge (i, j) is created from $G(k-1)$ to $G(k)$. For $1\leq i<j$, notice that

$$L_{(i,j)}\big(S_j+n\big)=\sum_{k=1}^{S_j+n}\Delta_k(i,j)=\sum_{k=0}^{n}\Delta_{S_j+k}(i,j). \tag{3.10}$$

For $n\geq 0$,

$$\mathbb{E}^{\mathcal{F}_{S_j+n}}\left(\Delta_{S_j+n+1}(i,j)\right)=\beta A_{ji}\big(S_j+n\big), \tag{3.11}$$

and

$$\mathbb{E}^{\mathcal{F}_{S_j-1}}\big(\Delta_{S_j}(i,j)\big)=\gamma\frac{D_i^{\text{out}}\big(S_j-1\big)+\delta_{\text{out}}}{S_j+\delta_{\text{out}}(j-1)}. \tag{3.12}$$

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

When $c_1 + c_2 \geq 1$, (3.3) and (3.4) suggest that $\sum_{k=0}^{\infty} A_{ji}\big(S_j + k\big) = \infty$ a.s.; then we apply [5, Theorem 4.5.5] to get

$$\frac{L_{(i,j)}\big(S_j + n\big)}{\gamma \frac{D_i^{\text{out}}\big(S_j - 1\big) + \delta_{\text{out}}}{S_j + \delta_{\text{out}}(j-1)} + \sum_{k=0}^{n-1} \beta A_{ji}\big(S_j + k\big)} \xrightarrow{\text{a.s.}} 1. \tag{3.13}$$

Also, by a similar argument as in (3.3), we have that when $c_1 + c_2 > 1$, there exists some constant $\widetilde{C}_{ij} > 0$ such that

$$\frac{1}{n^{c_1+c_2-1}} \sum_{k=0}^{n-1} \beta A_{ji}\big(S_j + k\big) \xrightarrow{\text{a.s.}} \beta \widetilde{C}_{ij} \xi_i^{\text{out}} \xi_j^{\text{in}}. \tag{3.14}$$

Then combining (3.13) with (3.14) gives

$$\frac{L_{(i,j)}\big(S_j + n\big)}{n^{c_1+c_2-1}} \xrightarrow{\text{a.s.}} \beta \widetilde{C}_j \xi_i^{\text{out}} \xi_j^{\text{in}}. \tag{3.15}$$

Analogous reasoning is also applicable to the case $c_1 + c_2 = 1$, where the scaling function $n^{c_1+c_2-1}$ is replaced with $\log n$, according to (3.4).

Next, consider the case $c_1 + c_2 < 1$. By the corollary in [11, Chapter IV.6, p. 151], we have a.s.

$$\begin{aligned}\left\{\sum_{k \geq S_j} \Delta_{k+1}(i, j) < \infty\right\} &= \left\{\sum_{k \geq S_j} \mathbb{E}^{\mathcal{F}_k}\big(\Delta_{k+1}(i, j)\big) < \infty\right\} \\ &= \left\{\sum_{k \geq S_j} \beta A_{ji}(k) < \infty\right\}.\end{aligned}$$

Using a similar argument as in Lemma 3.1(3), we have $\sum_{k \geq S_j} A_{ji}(k) < \infty$ a.s., thus giving a.s.

$$\sum_{k \geq S_j} \Delta_{k+1}(i, j) < \infty.$$

Hence, with probability 1, there is a finite number of (i, j) edges that can be formed, and there exists a last time for an (i, j) edge to form. □

Note that by the definition of $L_{i\leftrightarrow j}(n)$, applying continuous mapping gives the asymptotic results for $L_{i\leftrightarrow j}(n)$, which also depends on the value of $c_1 + c_2$:

(1) If $c_1 + c_2 > 1$, then there exists some random variable $\overline{\xi}_{ij}$, satisfying $\mathbb{P}\big(\overline{\xi}_{ij} \in (0, \infty)\big) = 1$, such that as $n \to \infty$,

$$\frac{L_{i\leftrightarrow j}(n)}{n^{c_1+c_2-1}} \xrightarrow{\text{a.s.}} \overline{\xi}_{ij}.$$

So for large n, the number of reciprocal edges between i and j is of order $n^{c_1+c_2-1}$.

(2) If $c_1 + c_2 = 1$, then there exists some random variable $\overline{\zeta}_{ij}$, satisfying $\mathbb{P}\big(\overline{\zeta}_{ij} \in (0, \infty)\big) = 1$, such that

$$\frac{L_{i\leftrightarrow j}(n)}{\log n} \xrightarrow{\text{a.s.}} \overline{\zeta}_{ij}.$$

So for large n, the number of reciprocal edges between i and j is of order $\log n$.

(3) If $c_1 + c_2 < 1$, then a.s.

$$L_{i\leftrightarrow j}(n) \uparrow L_{i\leftrightarrow j}(\infty) < \infty.$$

Now consider a special case with $\gamma = 0$ and $\delta_{\text{in}} = \delta_{\text{out}} = \delta > 0$; then

$$c_1 + c_2 < 1 \qquad \Leftrightarrow \qquad \alpha > \frac{1}{1+\delta}.$$

From Theorem 3.1, we see that when the probability of generating a new node in a PA network at each step is too high, the node set grows strongly and it is difficult to form reciprocal edges. Then the number of reciprocal edges between two nodes is finite a.s. If $\beta = 0$, then for $\delta_{\text{in}}, \delta_{\text{out}} > 0$, we always have $c_1 + c_2 < 1$, indicating that the number of edges between two fixed nodes is finite a.s. when no edge is added between two existing nodes.

3.2. Reciprocity in the entire graph

Here we consider the proportion of reciprocal edges in the entire PA network, R_n^{pa}, and the next theorem specifies the asymptotic behavior of R_n for $0 < c_1 + c_2 < 5/3$.

Theorem 3.2. *Suppose R_n^{pa} is as defined in (3.1). Then for $0 < c_1 + c_2 < 5/3$, we have*

$$R_n^{pa} \xrightarrow{p} 0, \qquad n \to \infty.$$

When $c_1 + c_2 < 5/3$, the reciprocity coefficient R_n^{pa} is likely to be small, provided that the number of edges in the PA network is large. In particular, when $c_1 + c_2 \le 1$, i.e. in the second and third cases in Theorem 3.1, we have $R_n^{\text{pa}} \xrightarrow{p} 0$.

Proof. It suffices to show that $\mathbb{E}\big(R_n^{\text{pa}}\big) \to 0$ for $0 < c_1 + c_2 < 5/3$, as $n \to \infty$. Recall that

$$\Delta_k(i,j) = \mathbf{1}_{\big\{E(k)=E(k-1)\cup\{(i,j)\}\big\}},$$

so that $\Delta_k(i,j)\mathbf{1}_{\{(j,i)\in E(k-1)\}}$ indicates the event that from $G(k-1)$ to $G(k)$, an edge (i, j) is created when (j, i) already exists in $G(k-1)$. By the definition of R_n^{pa}, we have

$$\begin{aligned}
R_n^{\text{pa}} &\le \frac{2}{n+1} \sum_{j=1}^{n} \sum_{i<j} \sum_{k=S_j}^{n} \Delta_k(i,j)\mathbf{1}_{\{(j,i)\in E(k-1)\}} \\
&\quad + \frac{2}{n+1} \sum_{j=1}^{n} \sum_{i<j} \sum_{k=S_j}^{n} \Delta_k(i,j)\mathbf{1}_{\{(i,j)\in E(k-1)\}} \\
&= \frac{2}{n+1} \sum_{j=1}^{n} \sum_{i<j} \sum_{k=S_j+1}^{n} \Delta_k(i,j)\mathbf{1}_{\{(j,i)\in E(k-1)\}} \\
&\quad + \frac{2}{n+1} \sum_{j=1}^{n} \sum_{i<j} \sum_{k=S_j+1}^{n} \Delta_k(i,j)\mathbf{1}_{\{(i,j)\in E(k-1)\}} \\
&=: Q_1(n) + Q_2(n).
\end{aligned}$$

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

For $Q_1(n)$, we have

$$\begin{aligned}
\mathbb{E}(Q_1(n)) &= \frac{2}{n+1}\sum_{j=1}^{n}\sum_{i<j}\mathbb{E}\left(\sum_{k=S_j+1}^{n}\Delta_k(i,j)\mathbf{1}_{\{(j,i)\in E(k-1)\}}\right)\\
&= \frac{2}{n+1}\sum_{j=1}^{n}\sum_{i<j}\mathbb{E}\left(\sum_{k=S_j+1}^{n}\mathbb{E}^{\mathcal{F}_{k-1}}\left(\Delta_k(i,j)\right)\mathbf{1}_{\{(j,i)\in E(k-1)\}}\right)\\
&= \frac{2}{n+1}\sum_{j=1}^{n}\sum_{i<j}\mathbb{E}\left(\sum_{k=S_j+1}^{n}\beta A_{ji}(k-1)\mathbf{1}_{\{(j,i)\in E(k-1)\}}\right). \qquad (3.16)
\end{aligned}$$

Since $S_j \geq j-1$ for $j\geq 2$, (3.16) implies

$$\mathbb{E}(Q_1(n)) \leq \frac{2}{n+1}\sum_{j=2}^{n}\sum_{i=1}^{j-1}\sum_{k=j}^{n}\mathbb{E}\left(A_{ji}(k-1)\mathbf{1}_{\left\{(j,i)\in E(k-1),k\geq S_j+1\right\}}\right). \qquad (3.17)$$

By the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
&\mathbb{E}\left(A_{ji}(k-1)\mathbf{1}_{\left\{(j,i)\in E(k-1),k\geq S_j+1\right\}}\right)\\
&\leq \left[\mathbb{E}\left(A_{ji}^2(k-1)\right)\right]^{1/2}\left[\mathbb{P}\big((j,i)\in E(k-1),\, k\geq S_j+1\big)\right]^{1/2}\\
&\leq \left[\mathbb{E}\left(A_{ji}^2(k-1)\right)\right]^{1/2}\left[\sum_{l=j}^{k-1}\mathbb{P}\big(E(l)=E(l-1)\cup\{(j,i)\},\, l\geq S_j+1\big)\right]^{1/2}\\
&\leq \left[\mathbb{E}\left(A_{ji}^2(k-1)\right)\right]^{1/2}\left[\sum_{l=j}^{k-1}\mathbb{E}\left(A_{ij}(l)\right)+\alpha\mathbb{E}\left(\frac{D_i^{\text{in}}\left(S_j-1\right)+\delta_{\text{in}}}{S_j+\delta_{\text{in}}(j-1)}\right)\right]^{1/2}. \qquad (3.18)
\end{aligned}$$

When $c_1+c_2>1$, Lemma A.2 and (A.11) together imply that there exist constants $M_1, M_2, M_3>0$ such that for $i<j\leq k\leq n$,

$$\mathbb{E}\left(A_{ji}^2(k-1)\right)\leq M_1\frac{k^{2\left(c_1+c_2-2\right)}}{i^{2c_2}j^{2c_1}},\quad \mathbb{E}\left(\frac{D_i^{\text{in}}\left(S_j-1\right)+\delta_{\text{in}}}{S_j+\delta_{\text{in}}(j-1)}\right)\leq M_3\frac{j^{c_1-1}}{i^{c_1}},$$

and that for $j\leq l\leq k-1$,

$$\mathbb{E}\left(A_{ij}(l)\right)\leq M_2\frac{l^{c_1+c_2-2}}{i^{c_1}j^{c_2}}.$$

Therefore, when $c_1+c_2>1$, we have

$$\begin{aligned}
&\mathbb{E}\left(A_{ji}(k-1)\mathbf{1}_{\left\{(j,i)\in E(k-1),k\geq S_j+1\right\}}\right)\\
&\leq \sqrt{M_1}\frac{k^{c_1+c_2-2}}{i^{c_2}j^{c_1}}\left(\frac{M_2}{c_1+c_2-1}\frac{k^{c_1+c_2-1}}{i^{c_1}j^{c_2}}+\alpha M_3 j^{c_1-1}/i^{c_1}\right)^{1/2},
\end{aligned}$$

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

and since $k^{c_1+c_2-1}/\left(i^{c_1}j^{c_2}\right) \geq j^{c_1-1}/i^{c_1}$ for $k \geq j$, there exists some constant $M > 0$ such that

$$\mathbb{E}\left(A_{ji}(k-1)\mathbf{1}_{\left\{(j,i)\in E(k-1), k\geq S_j+1\right\}}\right) \leq M\frac{k^{\frac{3}{2}(c_1+c_2)-\frac{5}{2}}}{i^{c_1/2+c_2}j^{c_1+c_2/2}}. \tag{3.19}$$

Then (3.17) leads to

$$\begin{aligned}\mathbb{E}(Q_1(n)) &\leq \frac{2M}{n+1}\sum_{j=2}^{n}\sum_{i=1}^{j-1}\sum_{k=j}^{n}\frac{k^{\frac{3}{2}(c_1+c_2)-\frac{5}{2}}}{i^{c_1/2+c_2}j^{c_1+c_2/2}} \\ &\leq \frac{2M}{3/2(c_1+c_2-1)}n^{\frac{3}{2}(c_1+c_2)-\frac{5}{2}}\sum_{j=2}^{n}\sum_{i=1}^{j-1} i^{-(c_1/2+c_2)}j^{-(c_1+c_2/2)}.\end{aligned}$$

If $c_1/2+c_2 > 1$, $c_1+c_2/2 > 1$, and $c_1+c_2 < 5/3$, then

$$\mathbb{E}(Q_1(n)) \leq \frac{2M}{3/2(c_1+c_2-1)}n^{\frac{3}{2}(c_1+c_2)-\frac{5}{2}}\sum_{i=1}^{\infty} i^{-(c_1/2+c_2)}\sum_{j=2}^{\infty} j^{-(c_1+c_2/2)} \to 0.$$

If $c_1/2+c_2 < 1$, $c_1+c_2/2 < 1$, then

$$\begin{aligned}\mathbb{E}(Q_1(n)) &\leq \frac{2M}{3/2(c_1+c_2-1)}n^{\frac{3}{2}(c_1+c_2)-\frac{5}{2}}\frac{n^{2-3/2(c_1+c_2)}}{(1-c_1/2-c_2)(1-c_1-c_2/2)} \\ &= \frac{2Mn^{-1/2}}{3/2(c_1+c_2-1)(1-c_1/2-c_2)(1-c_1-c_2/2)} \to 0.\end{aligned}$$

If $c_1/2+c_2 < 1$, $c_1+c_2/2 > 1$, and $c_1+c_2 < 5/3$, then

$$\begin{aligned}\mathbb{E}(Q_1(n)) &\leq \frac{2M}{3/2(c_1+c_2-1)}n^{\frac{3}{2}(c_1+c_2)-\frac{5}{2}}\frac{1}{1-(c_1/2+c_2)}\sum_{j=2}^{n} j^{1-(c_1/2+c_2)}j^{-1} \\ &\leq \frac{2Mn^{c_1+c_2/2-3/2}}{3/2(c_1+c_2-1)(1-(c_1/2+c_2))^2} \to 0,\end{aligned}$$

as $c_1+c_2/2 < 1+1/2 = 3/2$. Similarly, $\mathbb{E}(Q_1(n)) \to 0$, when $c_1/2+c_2 > 1$, $c_1+c_2/2 < 1$, and $c_1+c_2 < 5/3$. The proof machinery also applies to the case where either $c_1/2+c_2 = 1$ or $c_1+c_2/2 = 1$, and $c_1+c_2 < 5/3$, which gives the conclusion that for $1 < c_1+c_2 < 5/3$, $\mathbb{E}(Q_1(n)) \to 0$. Following the same reasoning, we have $\mathbb{E}(Q_2(n)) \to 0$, for $1 < c_1+c_2 < 5/3$, thus implying $R_n^{\text{pa}} \xrightarrow{p} 0$ for $1 < c_1+c_2 < 5/3$.

When $c_1+c_2 = 1$, we revise the bound in (3.19) to get the following: for some constant $\widetilde{M} > 0$,

$$\mathbb{E}\left(A_{ji}(k-1)\mathbf{1}_{\left\{(j,i)\in E(k-1),\, k\geq S_j+1\right\}}\right) \leq \widetilde{M}\frac{k^{-1}(\log k)^{1/2}}{i^{c_1/2+c_2}j^{c_1+c_2/2}}.$$

Then we have

$$\mathbb{E}(Q_1(n)) \leq \frac{2\widetilde{M}}{n}(\log n)^{3/2}\sum_{j=2}^{n}\sum_{i=1}^{j-1} i^{-(c_1/2+c_2)}j^{-(c_1+c_2/2)} \leq 4\widetilde{M}\frac{(\log n)^{3/2}}{n^{1/2}} \to 0.$$

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

Meanwhile, for some constant $\widetilde{M}'>0$, we have

$$\mathbb{E}(Q_2(n)) \le \frac{2\widetilde{M}'}{n}(\log n)^{3/2}\sum_{j=2}^{n}\sum_{i=1}^{j-1} i^{-(c_2/2+c_1)}j^{-(c_2+c_1/2)} \le 4\widetilde{M}'\frac{(\log n)^{3/2}}{n^{1/2}} \to 0.$$

Hence, we have $R_n^{\text{pa}} \xrightarrow{p} 0$ when $c_1+c_2=1$.

When $c_1+c_2<1$, the bound in (3.18) implies that there exists some constant $\bar{M}>0$ such that

$$\begin{aligned}\mathbb{E}\Big(A_{ji}(k-1)\mathbf{1}_{\{(j,i)\in E(k-1),k\ge S_j+1\}}\Big) &\le \bar{M}\frac{k^{c_1+c_2-2}}{i^{c_2}j^{c_1}}\left(\frac{j^{c_1+c_2-1}}{i^{c_1}j^{c_2}}\right)^{1/2}\\ &= \bar{M}\frac{k^{c_1+c_2-2}}{i^{c_1/2+c_2}j^{c_1/2+1/2}},\end{aligned} \tag{3.20}$$

which gives

$$\mathbb{E}(Q_1(n)) \le \frac{4\bar{M}}{(1-c_1-c_2)(1-c_1/2-c_2)}n^{-1/2} \to 0.$$

Similar reasoning also gives $\mathbb{E}(Q_2(n)) \to 0$ when $c_1+c_2<1$, thus giving $\mathbb{E}\big(R_n^{\text{pa}}\big) \to 0$ and completing the proof of the theorem. □

Remark 3.1. (i) From the definition of c_1 and c_2 in (2.1), if $\beta \le 2/3$, then $c_1+c_2<1+\beta\le 5/3$. Theorem 3.2 suggests that if the proportion of edges added between two existing nodes is less than $2/3$ and n is sufficiently large, the corresponding PA network will have R_n^{pa} close to 0.

(ii) Note also that

$$\mathbf{1}_{\left\{N_0^{i\leftrightarrow j}\le n\right\}} \le \sum_{k=1}^{n}\Delta_k(i,j)\mathbf{1}_{\{(j,i)\in E(k-1)\}} + \sum_{k=1}^{n}\Delta_k(j,i)\mathbf{1}_{\{(i,j)\in E(k-1)\}}.$$

Hence, when $c_1+2c_2<1$, applying the bound in (3.20) gives the following: for fixed i, there exists some constant $\tilde{M}>0$ such that

$$\lim_{n\to\infty}\mathbb{E}\left(\sum_{j\in V(n)}\mathbf{1}_{\left\{N_0^{i\leftrightarrow j}\le n\right\}}\right) \le \tilde{M}i^{-(c_1/2+c_2)}\sum_{j=1}^{\infty}j^{c_1/2+c_2-3/2}<\infty.$$

This indicates that when $c_1+2c_2<1$, a fixed node i can form a reciprocal pair of edges only with finitely many nodes.

3.2.1. *Simulation for* $c_1+c_2\ge 5/3$. Theorem 3.2 does not explain the asymptotic behavior of R_n^{pa} for $c_1+c_2\in[5/3,1+\beta)$, provided that $\beta\in(2/3,1)$. For comparison, we choose three sets of parameters,

$$\boldsymbol{\theta}_1=(\alpha,\beta,\gamma,\delta_{\text{in}},\delta_{\text{out}})=(0.1,0.8,0.1,2,1),$$
$$\boldsymbol{\theta}_2=(0.1,0.8,0.1,.4,.4),\qquad\text{and}\qquad \boldsymbol{\theta}_3=(0.05,0.9,0.05,1,1),$$

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

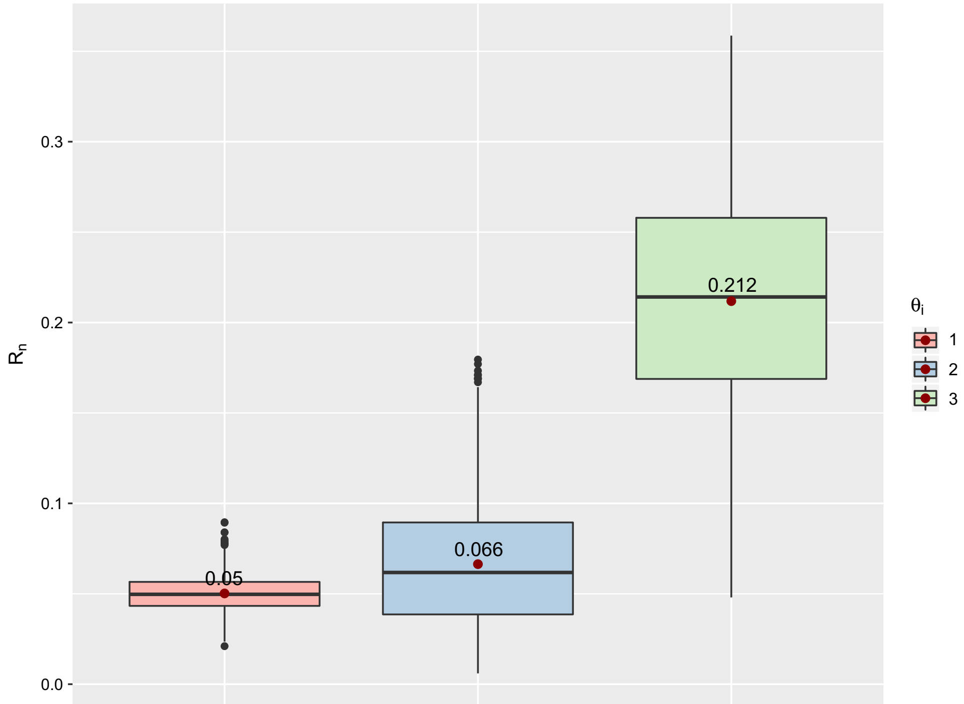


FIGURE 2. Box plots of R_n^{pa} for directed PA models simulated using $\boldsymbol{\theta}_i$, $i = 1, 2, 3$. The red dots represent the averaged empirical R_n^{pa} for each $\boldsymbol{\theta}_i$, $i = 1, 2, 3$.

such that the values of $c_1 + c_2$ are equal to $1.393 < 5/3$, $1.667 = 5/3$, and $1.727 > 5/3$, respectively. For each $\boldsymbol{\theta}_i$, $i = 1, 2, 3$, we simulate 1000 replications of the directed PA network with 10^5 edges, and compute the value of R_n^{pa} for each replication.

The numerical results are summarized as box plots in Figure 2. For each box plot, we use a dark red dot to mark the corresponding averaged empirical R_n^{pa}. Under $\boldsymbol{\theta}_1$, all 1000 empirical R_n^{pa} values are close to 0, with a maximum of 0.090 and a minimum of 0.021. The empirical R_n^{pa} values under $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$ are more variable, but both have higher mean than in the $\boldsymbol{\theta}_1$ case. This simulation experiment confirms that the asymptotic behavior of R_n^{pa} in a directed PA model depends on the value of $c_1 + c_2$. Meanwhile, when $c_1 + c_2 \geq 5/3$, the value of R_n^{pa} may not necessarily concentrate around a specific value, but may vary over a certain range.

3.3. The first time when a reciprocal pair forms

Theorem 3.1 gives results about the first time when a reciprocal pair forms between nodes $i \neq j$. The $c_1 + c_2 \geq 1$ scenario is discussed in Corollary 3.1, while the $c_1 + c_2 < 1$ case is analyzed in Proposition 3.1.

Corollary 3.1. *Let $N_0^{i\leftrightarrow j}$ be the first time when a reciprocal pair of edges $i \leftrightarrow j$ is formed between nodes $i \neq j$; i.e.*

$$N_0^{i\leftrightarrow j} := \inf\left\{n \geq 0 : (i, j) \in E(n),\ \textit{and}\ (j, i) \in E(n)\right\}, \tag{3.21}$$

with the convention that $\inf \emptyset = \infty$. *If $1 \leq i < j$ are fixed, and c_1, c_2 are as given in (2.1) and satisfy $c_1 + c_2 \geq 1$, then $N_0^{i\leftrightarrow j} < \infty$ a.s.*

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

Proof. We will show that for $c_1+c_2\geq 1$, $\lim_{n\to\infty}\mathbb{P}\Big(N_0^{i\leftrightarrow j}>S_j+n\Big)=\mathbb{P}\Big(N_0^{i\leftrightarrow j}=\infty\Big)=0$. Note that when $c_1+c_2\geq 1$, (3.15) indicates that $L_{(i,j)}\big(S_j+n\big)\xrightarrow{\text{a.s.}}\infty$, and similarly, $L_{(j,i)}\big(S_j+n\big)\xrightarrow{\text{a.s.}}\infty$. Therefore,

$$\begin{aligned}\mathbb{P}\Big(N_0^{i\leftrightarrow j}>S_j+n\Big)&\leq\mathbb{P}\big((i,j)\notin E\big(S_j+n\big)\big)+\mathbb{P}\big((j,i)\notin E\big(S_j+n\big)\big)\\&=\mathbb{P}\big(L_{(i,j)}\big(S_j+n\big)=0\big)+\mathbb{P}\big(L_{(j,i)}\big(S_j+n\big)=0\big)\to 0,\end{aligned}$$

as $n\to\infty$. □

When $c_1+c_2<1$, if we have $\mathbb{E}\big(L_{(i,j)}(\infty)\big)<1$, then

$$\mathbb{P}\big(L_{(i,j)}(\infty)<1\big)>0,$$

which implies $\mathbb{P}\big(L_{(i,j)}(\infty)=0\big)>0$, and $\mathbb{P}\Big(N_0^{i\leftrightarrow j}=\infty\Big)>0$. Note that

$$\mathbb{E}\big(L_{(i,j)}(\infty)\big)=\gamma\mathbb{E}\left(\frac{D_i^{\text{out}}\big(S_j-1\big)+\delta_{\text{out}}}{S_j+\delta_{\text{out}}(j-1)}\right)+\beta\sum_{k=0}^{\infty}\mathbb{E}\left(A_{ji}\big(S_j+k\big)\right),$$

and by (A.11), there exist some constants $K_1, K_2>0$ such that for $n\geq i\geq 1$,

$$\mathbb{E}\Big(\big(D_i^{\text{in}}(n)\big)^2\Big)\leq K_1(n/i)^{2c_1},\qquad \mathbb{E}\Big(\big(D_i^{\text{out}}(n)\big)^2\Big)\leq K_2(n/i)^{2c_2}.$$

Then applying the Cauchy–Schwarz inequality gives that for $n\geq j>i\geq 1$,

$$\begin{aligned}\mathbb{E}\big(A_{ji}(n)\big)&\leq\left[\mathbb{E}\left(\left(\frac{D_i^{\text{in}}(n)+\delta_{\text{in}}}{n+1+\delta_{\text{in}}|V(n)|}\right)^2\right)\right]^{1/2}\left[\mathbb{E}\left(\left(\frac{D_j^{\text{out}}(n)+\delta_{\text{out}}}{n+1+\delta_{\text{out}}|V(n)|}\right)^2\right)\right]^{1/2}\\&\leq\frac{1}{n^2}\left[\mathbb{E}\Big(\big(D_i^{\text{in}}(n)+\delta_{\text{in}}\big)^2\Big)\mathbb{E}\Big(\big(D_j^{\text{out}}(n)+\delta_{\text{out}}\big)^2\Big)\right]^{1/2}\\&\leq(K_1K_2)^{1/2}\frac{n^{c_1+c_2-2}}{i^{c_1}j^{c_2}}=:K'\frac{n^{c_1+c_2-2}}{i^{c_1}j^{c_2}}.\end{aligned}$$

Therefore, we have

$$\sum_{k=0}^{\infty}\mathbb{E}\left(A_{ji}\big(S_j+k\big)\right)\leq K'\sum_{k=j}^{\infty}\frac{k^{c_1+c_2-2}}{i^{c_1}j^{c_2}}\leq\frac{K'}{1-c_1-c_2}i^{-c_1}j^{c_1-1},$$

which gives

$$\mathbb{E}\big(L_{(i,j)}(\infty)\big)\leq\gamma+\frac{\beta K'}{1-c_1-c_2}i^{-c_1}j^{c_1-1}. \tag{3.22}$$

Since $c_1-1<0$, for j sufficiently large we have $\mathbb{E}\big(L_{(i,j)}(\infty)\big)<1$, thus giving $\mathbb{P}\Big(N_0^{i\leftrightarrow j}=\infty\Big)>0$. In other words, when $c_1+c_2<1$, it is possible to have zero pairs of reciprocal edges between two fixed nodes i, j, if j is created at a late stage of network evolution.

Equation (3.22) also suggests that for arbitrarily chosen i, j, $\mathbb{E}\big(L_{(i,j)}(\infty)\big) < 1$ if β is small enough. Hence, if few edges are created between existing nodes, it is possible for two fixed nodes never to form a reciprocal pair of edges. In particular, when $\beta = 0$, there is no reciprocal pair of edges, i.e. $\mathbb{P}\Big(N_0^{i\leftrightarrow j} = \infty\Big) = 1$, for all fixed i, j.

The following proposition assumes $c_1 + c_2 < 1$ and gives the asymptotic behavior of $\sup_{j\geq n\epsilon} \mathbb{P}\Big(N_0^{i\leftrightarrow j} \leq n\Big)$.

Proposition 3.1. *If $c_1 + c_2 < 1$, then for fixed $i \geq 1$, $\epsilon > 0$, as $n \to \infty$,*

$$\sup_{j\geq n\epsilon} \mathbb{P}\Big(N_0^{i\leftrightarrow j} \leq n\Big) \to 0.$$

Proof. Applying the union bound to $\mathbb{P}\Big(N_0^{i\leftrightarrow j} \leq S_j + n\Big)$ gives

$$\begin{aligned}
\mathbb{P}\Big(N_0^{i\leftrightarrow j} \leq S_j + n\Big) &\leq \sum_{k=1}^{n} \mathbb{P}\big(E\big(S_j + k\big) = E(S_j + k - 1) \cup \{(i, j)\}\big) \\
&\quad + \sum_{k=1}^{n} \mathbb{P}\big(E\big(S_j + k\big) = E(S_j + k - 1) \cup \{(j, i)\}\big) \\
&= \mathbb{E}\left(\sum_{k=0}^{n-1} \beta A_{ji}(S_j + k)\right) + \mathbb{E}\left(\sum_{k=0}^{n-1} \beta A_{ij}(S_j + k)\right) \\
&\leq \mathbb{E}\left(\sum_{k=j}^{\infty} \beta A_{ji}(k)\right) + \mathbb{E}\left(\sum_{k=j}^{\infty} \beta A_{ij}(k)\right).
\end{aligned}$$

By (3.5), we see that for $\epsilon > 0$,

$$\sup_{j\geq n\epsilon} \mathbb{E}\left(\sum_{k=j}^{\infty} A_{ji}(k)\right) \to 0, \qquad \sup_{j\geq n\epsilon} \mathbb{E}\left(\sum_{k=j}^{\infty} A_{ij}(k)\right) \to 0,$$

as $n \to \infty$, which gives

$$\sup_{j\geq n\epsilon} \mathbb{P}\Big(N_0^{i\leftrightarrow j} \leq n\Big) \leq \sup_{j\geq n\epsilon} \mathbb{P}\Big(N_0^{i\leftrightarrow j} \leq S_j + n\Big) \to 0. \qquad \square$$

4. Discussion

Suppose that we are given a scale-free network with a large proportion of reciprocal edges, e.g. Facebook wall posts [17], Twitter [6], Google+ [9], or Flickr [3, 10]. In fitting a directed PA model to such a dataset using inference methods developed in [18, 19], there is no guarantee that the calibrated model also has a large R_n^{pa}. In fact, estimated values of $\hat{c}_1$ and $\hat{c}_2$ do not necessarily satisfy $\hat{c}_1 + \hat{c}_2 \geq 5/3$. If we have $\hat{c}_1 + \hat{c}_2 < 5/3$ in the calibrated model, then by Theorem 3.2, the corresponding R_n^{pa} is close to 0, which differs from the feature of high reciprocity in the given dataset. This flags modeling error and suggests the consideration of alternative models or variants of the classical PA network. For instance, once a directed

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

edge (i, j) is created following the PA rule, we may add a reciprocal edge (j, i) with probability $\rho \in (0, 1)$. The study in [4] also provides other features that can be employed to predict reciprocal edges; we will defer the analysis of these variants of directed PA models to future research.

Appendix A. Lemmas and proofs needed in Section 2

In this section, we state and prove two lemmas needed for the proof of Proposition 2.1, which is also given.

A.1. Statement and proof of Lemma A.1.

Lemma A.1. *For some integer $p \geq 1$, $k \geq 0$, and $v \geq 1$, we have the following:*

(i) For $p = 1$,

$$\mathbb{E}^{\mathcal{F}_{S_v+k}}\left(D_v^{in}(S_v+k+1)+\delta_{in}\right) = \left(D_v^{in}(S_v+k)+\delta_{in}\right)\left(1+\frac{\alpha+\beta}{S_v+k+1+\delta_{in}|V(S_v+k)|}\right). \tag{A.1}$$

(ii) For an integer $p \geq 2$,

$$\begin{aligned}\mathbb{E}^{\mathcal{F}_{S_v+k}}&\left(\left(D_v^{in}(S_v+k+1)+\delta_{in}\right)^p\right)\\ &= \left(D_v^{in}(S_v+k)+\delta_{in}\right)^p\left(1+p\frac{\alpha+\beta}{S_v+k+1+\delta_{in}|V(S_v+k)|}\right)\\ &\quad+\frac{\alpha+\beta}{S_v+k+1+\delta_{in}|V(S_v+k)|}\sum_{r=2}^{p}\binom{p}{r}\left(D_v^{in}(S_v+k)+\delta_{in}\right)^{p-r+1}.\end{aligned} \tag{A.2}$$

Proof. Note that the right-hand sides of (A.1) and (A.2) are both $\mathcal{F}_{S_v+k}$-measurable. We prove the results for $p \geq 2$, and the case $p = 1$ follows by a similar argument. Let $F \in \mathcal{F}_{S_v+k}$; then

$$\begin{aligned}\int_F \mathbb{E}^{\mathcal{F}_{S_v+k}}&\left(\left(D_v^{\text{in}}(S_v+k+1)+\delta_{\text{in}}\right)^p\right)\mathrm{d}\mathbb{P}\\ &= \int_F \left(D_v^{\text{in}}(S_v+k+1)+\delta_{\text{in}}\right)^p \mathrm{d}\mathbb{P}\\ &= \sum_{l\geq k}\int_{F\cap\{S_v+k=l\}}\left(D_v^{\text{in}}(l+1)+\delta_{\text{in}}\right)^p \mathrm{d}\mathbb{P},\end{aligned}$$

and since $D_v^{\text{in}}(l+1) = D_v^{\text{in}}(l) + \mathbf{1}_{\{\text{Node } v \text{ is chosen at step } l+1\}} =: D_v^{\text{in}}(l) + \Delta_v(l+1)$, this equals

$$\begin{aligned}&= \sum_{l\geq k}\int_{F\cap\{S_v+k=l\}}\left(D_v^{\text{in}}(l)+\delta_{\text{in}}+\Delta_v(l+1)\right)^p \mathrm{d}\mathbb{P}\\ &= \sum_{l\geq k}\int_{F\cap\{S_v+k=l\}}\left(\left(D_v^{\text{in}}(l)+\delta_{\text{in}}\right)^p+\sum_{r=1}^{p}\binom{p}{2}\left(D_v^{\text{in}}(l)+\delta_{\text{in}}\right)^{p-r}\Delta_v(l+1)\right)\mathrm{d}\mathbb{P}.\end{aligned} \tag{A.3}$$

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

Since $F \cap \{S_v + k = l\} \in \mathcal{F}_l$, the quantity in (A.3) is equal to

$$\sum_{l\geq k} \int_{F\cap\{S_v+k=l\}} \left(D_v^{\text{in}}(l) + \delta_{\text{in}}\right)^p \mathrm{d}\mathbb{P}$$
$$+ \sum_{l\geq k} \int_{F\cap\{S_v+k=l\}} p\left(D_v^{\text{in}}(l) + \delta_{\text{in}}\right)^{p-1} \mathbb{E}^{\mathcal{F}_l}\left(\Delta_v(l+1)\right)\mathrm{d}\mathbb{P}$$
$$+ \sum_{r=2}^{p} \sum_{l\geq k} \int_{F\cap\{S_v+k=l\}} \binom{p}{r} \left(D_v^{\text{in}}(l) + \delta_{\text{in}}\right)^{p-r} \mathbb{E}^{\mathcal{F}_l}\left(\Delta_v(l+1)\right)\mathrm{d}\mathbb{P}.$$

Note also that $\mathbb{E}^{\mathcal{F}_l}\left(\Delta_v(l+1)\right) = (\alpha+\beta)\left(D_v^{\text{in}}(l) + \delta_{\text{in}}\right)/\left(l+1+\delta_{\text{in}}|V(l)|\right)$, so we have

$$\int_F \mathbb{E}^{\mathcal{F}_{S_v+k}}\left(\left(D_v^{\text{in}}(S_v+k+1) + \delta_{\text{in}}\right)^p\right) \mathrm{d}\mathbb{P}$$
$$= \int_F \left(D_v^{\text{in}}(S_v+k) + \delta_{\text{in}}\right)^p \left(1 + p\frac{\alpha+\beta}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right) \mathrm{d}\mathbb{P}$$
$$+ \int_F \frac{\alpha+\beta}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|} \sum_{r=2}^{p} \binom{p}{r} \left(D_v^{\text{in}}(S_v+k) + \delta_{\text{in}}\right)^{p-r+1} \mathrm{d}\mathbb{P}. \qquad \square$$

A.2. Statement and proof of Lemma A.2.

Next, we study properties of $\mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)\right)^p\right]$ and $\mathbb{E}\left[\left(D_v^{\text{out}}(S_v+k)\right)^p\right]$, $k \geq 1$, which are needed for deriving the theorems in Section 3 as well as for the proof of Proposition 2.1.

Lemma A.2. *For $v \in V(n)$ and $p \geq 1$, we have*

$$\sup_{k\geq 1} \frac{\mathbb{E}\left[\left(D_v^{in}(k)\right)^p\right]}{k^{c_1 p}} < \infty, \qquad \sup_{k\geq 1} \frac{\mathbb{E}\left[\left(D_v^{out}(k)\right)^p\right]}{k^{c_2 p}} < \infty.$$

Proof. For $p = 1$, we see from (A.1) that

$$\mathbb{E}\left(D_v^{\text{in}}(S_v+k+1) + \delta_{\text{in}}\right)$$
$$= \mathbb{E}\left(\left(D_v^{\text{in}}(S_v+k) + \delta_{\text{in}}\right)\left(1 + \frac{\alpha+\beta}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right)\right);$$

since $S_v \geq 0$ and $|V(S_v+k)| \geq |V(k)|$, this is

$$\leq \mathbb{E}\left(\left(D_v^{\text{in}}(S_v+k) + \delta_{\text{in}}\right)\left(1 + \frac{\alpha+\beta}{k+1+\delta_{\text{in}}|V(k)|}\right)\right),$$
$$\leq \mathbb{E}\left(D_v^{\text{in}}(S_v+k) + \delta_{\text{in}}\right)\left(1 + \frac{c_1}{k}\right)$$
$$+ \mathbb{E}\left(\left(D_v^{\text{in}}(S_v+k) + \delta_{\text{in}}\right)\frac{(\alpha+\beta)\delta_{\text{in}}\left||V(k)| - (1-\beta)k\right|}{(k+\delta_{\text{in}}|V(k)|)\left(1+\delta_{\text{in}}(1-\beta)\right)k}\right),$$

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

and as $\alpha+\beta\le 1$, this is

$$\begin{aligned}
&\le \mathbb{E}\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)\left(1+\frac{c_1}{k}\right)\\
&\quad+\mathbb{E}\left(\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)\frac{\delta_{\text{in}}\Big||V(k)|-(1-\beta)k\Big|}{(k+\delta_{\text{in}}|V(k)|)\big(1+\delta_{\text{in}}(1-\beta)\big)k}\right)\\
&=: H_v^{(1)}(k)+H_v^{(2)}(k).
\end{aligned}\tag{A.4}$$

Since $\big|V(k)\big|-1$ follows a binomial distribution with size $k\ge 1$ and success probability $1-\beta$, by applying the Chernoff bound we obtain

$$\mathbb{P}\left(\Big|\big|V(k)\big|-(1-\beta)k\Big|\ge 1+\sqrt{12(1-\beta)k\log k}\right)\le\frac{2}{k^4},\tag{A.5}$$

and rewrite the term $H_v^{(2)}(k)$ in (A.4) as

$$\begin{aligned}
&H_v^{(2)}(k)\\
&=\mathbb{E}\left(\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)\frac{\delta_{\text{in}}\Big||V(k)|-(1-\beta)k\Big|}{(k+\delta_{\text{in}}|V(k)|)\big(1+\delta_{\text{in}}(1-\beta)\big)k}\right.\\
&\qquad\left.\times\mathbf{1}_{\left\{\Big||V(k)|-(1-\beta)k\Big|\le 1+\sqrt{12(1-\beta)k\log k}\right\}}\right)\\
&+\mathbb{E}\left(\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)\frac{\delta_{\text{in}}\Big||V(k)|-(1-\beta)k\Big|}{\big(k+\delta_{\text{in}}|V(k)|\big)\big(1+\delta_{\text{in}}(1-\beta)\big)k}\right.\\
&\qquad\left.\times\mathbf{1}_{\left\{\Big||V(k)|-(1-\beta)k\Big|> 1+\sqrt{12(1-\beta)k\log k}\right\}}\right).
\end{aligned}$$

Since $D_v^{\text{in}}(S_v+k)\le k+1$ and $\Big||V(k)|-(1-\beta)k\Big|\le k+1$, the foregoing term is bounded by

$$\begin{aligned}
&\mathbb{E}\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)\frac{\delta_{\text{in}}\big(1+\sqrt{12(1-\beta)k\log k}\big)}{k^2}+\frac{\delta_{\text{in}}\big(k+1+\delta_{\text{in}}\big)(k+1)}{\big(1+\delta_{\text{in}}(1-\beta)\big)k^2}\frac{2}{k^4}\\
&\le\mathbb{E}\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)\frac{\delta_{\text{in}}\big(1+\sqrt{12k\log k}\big)}{k^2}+\frac{2\delta_{\text{in}}\big(k+1+\delta_{\text{in}}\big)(k+1)}{k^6}.
\end{aligned}\tag{A.6}$$

Combining the bound in (A.6) with (A.4) gives

$$\begin{aligned}
\mathbb{E}\left(D_v^{\text{in}}(S_v+k+1)+\delta_{\text{in}}\right)&\le\mathbb{E}\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)\left(1+\frac{c_1}{k}\right)\\
&\quad+\mathbb{E}\left(\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)\frac{\delta_{\text{in}}\Big||V(k)|-(1-\beta)k\Big|}{(k+\delta_{\text{in}}|V(k)|)\big(1+\delta_{\text{in}}(1-\beta)\big)k}\right)\\
&\le\mathbb{E}\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)\left(1+\frac{c_1}{k}+\frac{\delta_{\text{in}}\left(1+\sqrt{12k\log k}\right)}{k^2}\right)\\
&\quad+\frac{2\delta_{\text{in}}\big(k+1+\delta_{\text{in}}\big)(k+1)}{k^6}.
\end{aligned}$$

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

Recursively applying the inequality above k times, we have

$$\begin{aligned}
&\mathbb{E}\left(D_v^{\text{in}}(S_v+k+1)+\delta_{\text{in}}\right)\\
&\le \mathbb{E}\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)\left(1+\frac{c_1}{k}+\frac{\delta_{\text{in}}\left(1+\sqrt{12k\log k}\right)}{k^2}\right)\\
&\quad+\frac{2\delta_{\text{in}}\big(k+1+\delta_{\text{in}}\big)(k+1)}{k^6}\\
&\le \ldots \le \mathbb{E}\big(D_v^{\text{in}}(S_v)+\delta_{\text{in}}\big)\prod_{l=1}^{k}\left(1+\frac{c_1}{l}+\frac{\delta_{\text{in}}\left(1+\sqrt{12l\log l}\right)}{l^2}\right)\\
&\quad+2\delta_{\text{in}}\sum_{l=1}^{k}\frac{\big(l+1+\delta_{\text{in}}\big)(l+1)}{l^6}\prod_{s=l+1}^{k}\left(1+\frac{c_1}{s}+\frac{\delta_{\text{in}}\left(1+\sqrt{12s\log s}\right)}{s^2}\right).
\end{aligned}\tag{A.7}$$

Here, $\mathbb{E}\big(D_v^{\text{in}}(S_v)+\delta_{\text{in}}\big)=\alpha\delta_{\text{in}}+\gamma(1+\delta_{\text{in}})$, depending on whether the α- or the γ-scenario occurs. Note that there exists a constant $M>0$ such that

$$\begin{aligned}
\prod_{l=1}^{k}\left(1+\frac{c_1}{l}+\frac{\delta_{\text{in}}\left(1+\sqrt{12l\log l}\right)}{l^2}\right)&\le \exp\left\{\sum_{l=1}^{k}\left(\frac{c_1}{l}+\frac{\delta_{\text{in}}\left(1+\sqrt{12l\log l}\right)}{l^2}\right)\right\}\\
&\le Mk^{c_1},
\end{aligned}\tag{A.8}$$

and it follows from (A.7) that

$$\sup_{k\ge1}\frac{\mathbb{E}\left[D_v^{\text{in}}(S_v+k)\right]}{k^{c_1}}\le\sup_{k\ge1}\frac{\mathbb{E}\left[D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right]}{k^{c_1}}<\infty.$$

For $p\ge2$, suppose

$$\sup_{k\ge1}\frac{\mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^r\right]}{k^{c_1 r}}\le A_r<\infty$$

holds for some constants, A_r, $r=1,\ldots,p-1$. Let $A_0=\max\{A_r: r=1,\ldots,p-1\}$; then by (A.2), we have

$$\begin{aligned}
&\mathbb{E}\left(\left(D_v^{\text{in}}(S_v+k+1)+\delta_{\text{in}}\right)^p\right)\\
&\le \mathbb{E}\left(\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p\left(1+p\frac{\alpha+\beta}{S_v+k+\delta_{\text{in}}|V(S_v+k)|}\right)\right)\\
&\quad+\sum_{r=1}^{p-1}(\alpha+\beta)\binom{p}{2}A_r k^{c_1 r-1}\\
&\le \mathbb{E}\left(\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p\left(1+p\frac{\alpha+\beta}{S_v+k+\delta_{\text{in}}|V(S_v+k)|}\right)\right)\\
&\quad+\frac{1}{2}(\alpha+\beta)p(p-1)^2A_0k^{c_1(p-1)-1}\\
&=: C_v^{(1)}(k)+C_v^{(2)}(k).
\end{aligned}\tag{A.9}$$

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

We rewrite the $C_v^{(1)}(k)$ term in (A.9) to get

$$\begin{aligned}
C_v^{(1)}(k) &= \left(1+\frac{c_1 p}{k}\right)\mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p\right] \\
&\quad + \mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p\left(\frac{p(\alpha+\beta)}{S_v+k+\delta_{\text{in}}|V(S_v+k)|}-\frac{c_1 p}{k}\right)\right] \\
&\le \left(1+\frac{c_1 p}{k}\right)\mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p\right] \\
&\quad + \mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p \frac{p\delta_{\text{in}}\Big||V(k)|-(1-\beta)k\Big|}{(k+\delta_{\text{in}}|V(k)|)(1+\delta_{\text{in}}(1-\beta))k}\right] \\
&\le \left(1+\frac{c_1 p}{k}\right)\mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p\right] \\
&\quad + \mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p \frac{p\delta_{\text{in}}\Big||V(k)|-(1-\beta)k\Big|}{k^2}\right].
\end{aligned}$$

Similarly to the Chernoff bound in (A.5), we have, for $p\ge 2$,

$$\mathbb{P}\left(\Big||V(k)|-(1-\beta)k\Big|\ge 1+\sqrt{6p(1-\beta)k\log k}\right)\le \frac{2}{k^{2p}}. \tag{A.10}$$

Therefore, analogously to the calculation in (A.6), we have

$$\begin{aligned}
&\mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p \frac{p\delta_{\text{in}}\Big||V(k)|-(1-\beta)k\Big|}{k^2}\right] \\
&= \mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p \frac{p\delta_{\text{in}}\Big||V(k)|-(1-\beta)k\Big|}{k^2}\mathbf{1}_{\left\{\left||V(k)|-(1-\beta)k\right|\le 1+\sqrt{6p(1-\beta)k\log k}\right\}}\right] \\
&\quad + \mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p \frac{p\delta_{\text{in}}\Big||V(k)|-(1-\beta)k\Big|}{k^2}\mathbf{1}_{\left\{\left||V(k)|-(1-\beta)k\right|> 1+\sqrt{6p(1-\beta)k\log k}\right\}}\right],
\end{aligned}$$

and since $\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p\le\left(k+1+\delta_{\text{in}}\right)^p$, this is

$$\le \mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p\right]\frac{p\delta_{\text{in}}\left(1+\sqrt{6p\,k\log k}\right)}{k^2}+\frac{p\delta_{\text{in}}\left(k+1+\delta_{\text{in}}\right)^p(k+1)}{k^2}\frac{2}{k^{2p}}.$$

Hence,

$$\begin{aligned}
C_v^{(1)}(k) &\le \mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p\right]\left(1+\frac{c_1 p}{k}+\frac{p\delta_{\text{in}}\left(1+\sqrt{6p\,k\log k}\right)}{k^2}\right) \\
&\quad + \frac{2p\delta_{\text{in}}\left(k+1+\delta_{\text{in}}\right)^p(k+1)}{k^{2p+2}}.
\end{aligned}$$

Note also that

$$\left(k+1+\delta_{\text{in}}\right)^p(k+1)k^{-2p-2}\le 2(2+\delta_{\text{in}})^p k^{-p-1}\le 2\left(2+\delta_{\text{in}}\right)^p k^{c_1(p-1)-1}$$

for all $k \geq 1$, $p \geq 2$. We conclude from (A.9) that

$$\begin{aligned}\mathbb{E}\left(\left(D_v^{\text{in}}(S_v+k+1)+\delta_{\text{in}}\right)^p\right) &\leq \mathbb{E}\left(\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p\right)\left(1+\frac{c_1p}{k}+\frac{p\delta_{\text{in}}\left(1+\sqrt{6pk\log k}\right)}{k^2}\right)\\ &\quad+\left(A_0(\alpha+\beta)p(p-1)^2/2+4p\delta_{\text{in}}\left(2+\delta_{\text{in}}\right)^p\right)k^{c_1(p-1)-1}.\end{aligned}$$

Following the recursive step as for the $p=1$ case gives

$$\sup_{k\geq 1}\frac{\mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)\right)^p\right]}{k^{c_1p}} \leq \sup_{k\geq 1}\frac{\mathbb{E}\left[\left(D_v^{\text{in}}(S_v+k)+\delta_{\text{in}}\right)^p\right]}{k^{c_1p}} < \infty.$$

Note that for $p \geq 1$,

$$\begin{aligned}\mathbb{E}\left(\left(\frac{D_v^{\text{in}}(n)}{n^{c_1}}\right)^p\right) &= \mathbb{E}\left(\left(\frac{D_v^{\text{in}}(n)}{n^{c_1}}\right)^p \mathbf{1}_{\{S_v\leq n-1\}}\right)+\mathbb{E}\left(\left(\frac{D_v^{\text{in}}(n)}{n^{c_1}}\right)^p \mathbf{1}_{\{S_v\geq n\}}\right)\\ &\leq \mathbb{E}\left(\left(\frac{D_v^{\text{in}}(S_v+n)}{n^{c_1}}\right)^p\right)+\left(\frac{1}{n^{c_1}}\right)^p,\end{aligned}$$

since $D_v^{\text{in}}(n)$ is monotone in n. Then we have, for $v \geq 1$,

$$\sup_{n\geq 1}\mathbb{E}\left(\left(\frac{D_v^{\text{in}}(n)}{n^{c_1}}\right)^p\right) < \infty.$$

Applying a similar argument to the out-degrees completes the proof. □

Remark A.1. In the proof of Lemma A.2, if we revise the inequality $|V(S_v+k)| \geq |V(k)|$ to $|V(S_v+k)| \geq |V(v+k-1)|$, then we have for $p \geq 1$

$$\sup_{n\geq 1}\sup_{v\geq 1}\mathbb{E}\left(\left(\frac{D_v^{\text{in}}(n)}{(n/v)^{c_1}}\right)^p\right) < \infty, \qquad \sup_{n\geq 1}\sup_{v\geq 1}\mathbb{E}\left(\left(\frac{D_v^{\text{out}}(n)}{(n/v)^{c_2}}\right)^p\right) < \infty. \tag{A.11}$$

These results are used in the proof of Theorem 3.2, in Section 3.2.

A.3. Proof of Proposition 2.1

We prove only the results for $D_v^{\text{in}}(n)$; those for $D_v^{\text{out}}(n)$ follow from a similar argument. First, by Lemma A.1, we see that for $n \geq 1$,

$$\frac{D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}}{\prod_{k=0}^{n-1}\left(1+\frac{\alpha+\beta}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right)} =: \frac{D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}}{X_n^{(v)}} \tag{A.12}$$

is a nonnegative $\left(\mathcal{F}_{S_v+n}\right)_{n\geq 0}$-martingale, which by the martingale convergence theorem converges to some limit L_v a.s. as $n \to \infty$. It remains to analyze the denominator and to verify that $\mathbb{P}(L_v \in (0, \infty)) = 1$. We do this by applying some proof machinery similar to that of [16, Lemma 8.17].

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

By Markov's inequality, we see that for $\epsilon > 0$, and $\max\{-1, -\delta_{\text{in}}\} < m < 0$,

$$\begin{aligned}\mathbb{P}(L_v \le \epsilon) &= \limsup_{n\to\infty} \mathbb{P}\left(\frac{D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}}{X_n^{(v)}} \le \epsilon\right)\\ &\le \epsilon^{|m|} \limsup_{n\to\infty} \mathbb{E}\left[\left(\frac{D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}}{X_n^{(v)}}\right)^m\right]\\ &\le \epsilon^{|m|} \limsup_{n\to\infty} \mathbb{E}\left[\frac{\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}\right)^m}{\prod_{k=0}^{n-1}\left(1+\frac{(\alpha+\beta)m}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right)}\right]. \end{aligned} \tag{A.13}$$

By (A.13), it suffices to show

$$\limsup_{n\to\infty} \mathbb{E}\left[\frac{\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}\right)^m}{\prod_{k=0}^{n-1}\left(1+\frac{(\alpha+\beta)m}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right)}\right] < \infty.$$

Similarly to [16, Equation (8.7.23)], there exists some constant C_m such that

$$\begin{aligned}&\limsup_{n\to\infty} \mathbb{E}\left[\frac{\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}\right)^m}{\prod_{k=0}^{n-1}\left(1+\frac{(\alpha+\beta)m}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right)}\right]\\ &\qquad \le C_m \limsup_{n\to\infty}\ \mathbb{E}\left[\frac{\Gamma\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}+m\right)/\,\Gamma\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}\right)}{\prod_{k=0}^{n-1}\left(1+\frac{(\alpha+\beta)m}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right)}\right].\end{aligned}$$

Hence, once we show

$$\limsup_{n\to\infty}\ \mathbb{E}\left[\frac{\Gamma\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}+m\right)/\,\Gamma\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}\right)}{\prod_{k=0}^{n-1}\left(1+\frac{(\alpha+\beta)m}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right)}\right] < \infty, \tag{A.14}$$

the inequality in (A.13) implies $\mathbb{P}(L_v \in (0, \infty)) = 1$.

We prove (A.14) by showing that

$$M_n^{(m)} := \frac{\Gamma(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}+m)/\,\Gamma(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}})}{\prod_{k=0}^{n-1}\left(1+\frac{(\alpha+\beta)m}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right)}$$

is an $\left(\mathcal{F}_{S_v+n}\right)_{n\ge 0}$-martingale. Note that

$$\begin{aligned}&\mathbb{E}^{\mathcal{F}_{S_v+n}}\left(\frac{\Gamma\left(D_v^{\text{in}}(S_v+n+1)+\delta_{\text{in}}+m\right)}{\Gamma\left(D_v^{\text{in}}(S_v+n+1)+\delta_{\text{in}}\right)}\right)\\ &\quad= \frac{\Gamma\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}+m\right)}{\Gamma\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}\right)}\left(1-\frac{(\alpha+\beta)\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}\right)}{S_v+n+1+\delta_{\text{in}}|V(S_v+n)|}\right)\\ &\qquad+\frac{(\alpha+\beta)\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}\right)}{S_v+n+1+\delta_{\text{in}}|V(S_v+n)|}\,\frac{\Gamma\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}+m\right)}{\Gamma\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}\right)}\,\frac{D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}+m}{D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}}\\ &\quad= \frac{\Gamma\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}+m\right)}{\Gamma\left(D_v^{\text{in}}(S_v+n)+\delta_{\text{in}}\right)}\left(1+\frac{(\alpha+\beta)m}{S_v+n+1+\delta_{\text{in}}|V(S_v+n)|}\right),\end{aligned}$$

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

which confirms that $M_n^{(m)}$ is an $\left(\mathcal{F}_{S_v+n}\right)_{n\ge 0}$-martingale. Also,

$$\begin{aligned}
\mathbb{E}\big(M_n^{(m)}\big) &= \mathbb{E}\big(M_0^{(m)}\big) \\
&= \mathbb{E}\left(\frac{\Gamma\big(D_v^{\text{in}}(S_v)+\delta_{\text{in}}+m\big)/\,\Gamma\big(D_v^{\text{in}}(S_v)+\delta_{\text{in}}\big)}{1+\frac{(\alpha+\beta)m}{S_v+1+\delta_{\text{in}}v}}\right).
\end{aligned}$$

Since $m<0$ and $S_v\ge v-1\ge 0$, we have

$$\left(1+\frac{(\alpha+\beta)m}{S_v+1+\delta_{\text{in}}v}\right)^{-1}\le\left(1+\frac{(\alpha+\beta)m}{(1+\delta_{\text{in}})v}\right)^{-1}.$$

This further implies

$$\begin{aligned}
\mathbb{E}\big(M_n^{(m)}\big) &\le \left(1+\frac{(\alpha+\beta)m}{(1+\delta_{\text{in}})v}\right)^{-1}\mathbb{E}\left(\Gamma\big(D_v^{\text{in}}(S_v)+\delta_{\text{in}}+m\big)/\,\Gamma\big(D_v^{\text{in}}(S_v)+\delta_{\text{in}}\big)\right) \\
&= \left(1+\frac{(\alpha+\beta)m}{(1+\delta_{\text{in}})v}\right)^{-1}\left(\alpha\frac{\Gamma(\delta_{\text{in}}+m)}{\Gamma(\delta_{\text{in}})}+\gamma\frac{\Gamma(\delta_{\text{in}}+1+m)}{\Gamma(1+\delta_{\text{in}})}\right)<\infty,
\end{aligned}$$

thus completing the proof of (A.14).

Next, we consider the convergence of $X_n^{(v)}$ by noting that

$$\begin{aligned}
&\log X_n^{(v)} \\
&= \sum_{k=0}^{n-1}\left[\log\left(1+\frac{\alpha+\beta}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right)-\frac{\alpha+\beta}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right] \\
&\quad+\sum_{k=0}^{n-1}\left(\frac{\alpha+\beta}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}-\frac{c_1}{S_v+k+1}\right) \\
&\quad+\left(\sum_{k=0}^{n-1}\frac{c_1}{S_v+k+1}-c_1\log\frac{S_v+n}{S_v+1}\right)+c_1\log\frac{S_v+n}{S_v+1} \\
&=: \mathrm{I}_v(n)+\mathrm{II}_v(n)+\mathrm{III}_v(n)+c_1\log\frac{S_v+n}{S_v+1}.
\end{aligned}$$

Since $\log(1+x)\le x$, for all $x\ge 0$, we have $\mathrm{I}_v(n+1)-\mathrm{I}_v(n)\le 0$ for all n; i.e. $\mathrm{I}_v(n)$ is decreasing in n. Note also that $|\log(1+x)-x|\le x^2/2$ for all $x\ge 0$, so we have

$$\begin{aligned}
&\mathbb{E}\left|\sum_{k=0}^{\infty}\left(\log\left(1+\frac{\alpha+\beta}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right)-\frac{\alpha+\beta}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right)\right| \\
&\le \sum_{k=0}^{\infty}\mathbb{E}\left|\log\left(1+\frac{\alpha+\beta}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right)-\frac{\alpha+\beta}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right| \\
&\le \frac{(\alpha+\beta)^2}{2}\sum_{k=0}^{\infty}\mathbb{E}\left(\frac{1}{S_v+k+1+\delta_{\text{in}}|V(S_v+k)|}\right)^2\le\sum_{k=1}^{\infty}\frac{1}{k^2}<\infty,
\end{aligned}$$

which implies $\mathrm{I}_v(\infty)<\infty$ a.s., and $\mathrm{I}_v(n)\xrightarrow{\text{a.s.}}\mathrm{I}_v(\infty)$ as $n\to\infty$.

https://doi.org/10.1017/apr.2021.52 Published online by Cambridge University Press

By [1, Theorem 3.9.4], we see that there exists a finite random variable Z such that

$$\sum_{k=1}^{\infty}\left(\frac{\alpha+\beta}{k+\delta_{\text{in}}|V(k-1)|}-\frac{c_1}{k}\right)\xrightarrow{\text{a.s.}} Z;$$

then

$$\text{II}_v(n)\xrightarrow{\text{a.s.}} Z-\sum_{k=1}^{S_v}\left(\frac{\alpha+\beta}{k+\delta_{\text{in}}|V(k-1)|}-\frac{c_1}{k}\right)=:\text{II}_v(\infty).$$

Since $\sum_{k=1}^{n} 1/k-\log n\to\tilde{c}$, where $\tilde{c}$ is Euler's constant, for $v=1$ we have

$$\text{III}_1(n)\xrightarrow{\text{a.s.}} c_1\tilde{c}\ =:\text{III}_1(\infty),$$

and for $v\geq 2$,

$$\text{III}_v(n)\xrightarrow{\text{a.s.}} c_1\left(\tilde{c}+\log(S_v+1)-\sum_{k=1}^{S_v}\frac{1}{k}\right)=:\text{III}_v(\infty).$$

Hence, as $n\to\infty$,

$$\frac{X_n^{(v)}}{((S_v+n)/(S_v+1))^{c_1}}\xrightarrow{\text{a.s.}}\exp\{\text{I}_v(\infty)+\text{II}_v(\infty)+\text{III}_v(\infty)\}. \tag{A.15}$$

Combining (A.15) with the convergence of the martingale in (A.12), we have

$$\frac{D_v^{\text{in}}(S_v+n)}{(S_v+n)^{c_1}}\xrightarrow{\text{a.s.}}\frac{L_v}{(S_v+1)^{c_1}}\exp\left\{-\big(\text{I}_v(\infty)+\text{II}_v(\infty)+\text{III}_v(\infty)\big)\right\}\ =:\xi_v^{\text{in}};$$

then

$$\lim_{n\to\infty}\frac{D_v^{\text{in}}(n)}{n^{c_1}}=\lim_{n\to\infty}\frac{D_v^{\text{in}}(S_v+n)}{(S_v+n)^{c_1}}\overset{\text{a.s.}}{=}\xi_v^{\text{in}}.$$

Since $\mathbb{P}(L_v\in(0,\infty))=1$ and $S_v+1\geq v\geq 1$, we also have $\mathbb{P}(\xi_v^{\text{in}}\in(0,\infty))=1$.

Acknowledgements

We thank the two anonymous referees for their valuable comments on the manuscript.

Funding information

There are no funding bodies to thank in relation to this creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process for this article.

References

[1] Athreya, K. and Ney, P. (2004). *Branching Processes*. Springer, New York.

[2] Bollobás, B., Borgs, C., Chayes, J. and Riordan, O. (2003). Directed scale-free graphs. In *Proceedings of the Fourteenth Annual ACM–SIAM Symposium on Discrete Algorithms (Baltimore, 2003)*, Association for Computing Machinery, New York, pp. 132–139.

[3] CHA, M., MISLOVE, A. AND GUMMADI, K. (2009). A measurement-driven analysis of information propagation in the Flickr social network. In *Proceedings of the 18th International Conference on World Wide Web (WWW '09)*, Association for Computing Machinery, New York, pp. 721–730.
[4] CHENG, J., ROMERO, D., MEEDER, B. AND KLEINBERG, J. (2011). Predicting reciprocity in social networks. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 49–56.
[5] DURRETT, R. (2019). *Probability: Theory and Examples*, 5th edn. Cambridge University Press.
[6] JAVA, A., SONG, X., FININ, T. AND TSENG, B. (2007). Why we Twitter: understanding microblogging usage and communities. In *WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, Association for Computing Machinery, New York, pp. 56–65.
[7] JIANG, B., ZHANG, Z. AND TOWSLEY, D. (2015). Reciprocity in social networks with capacity constraints. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*, Association for Computing Machinery, New York, pp. 457–466.
[8] KRAPIVSKY, P. AND REDNER, S. (2001). Organization of growing random networks. *Phys. Rev. E* **63**, 066123.
[9] MAGNO, G. *et al.* (2012). New kid on the block: exploring the Google + social graph. In *Proceedings of the 2012 Internet Measurement Conference (IMC '12)*, Association for Computing Machinery, New York, pp. 159–170.
[10] MISLOVE, A. *et al.* (2007). Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement (IMC '07)*, Association for Computing Machinery, New York, pp. 29–42.
[11] NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Holden-Day, San Francisco.
[12] NEWMAN, M., FORREST, S. AND BALTHROP, J. (2002). Email networks and the spread of computer viruses. *Phys. Rev. E* **66**, 035101.
[13] RESNICK, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.
[14] RESNICK, S. AND SAMORODNITSKY, G. (2015). Tauberian theory for multivariate regularly varying distributions with application to preferential attachment networks. *Extremes* **18**, 349–367.
[15] SAMORODNITSKY, G. *et al.* (2016). Nonstandard regular variation of in-degree and out-degree in the preferential attachment model. *J. Appl. Prob.* **53**, 146–161.
[16] VAN DER HOFSTAD, R. (2017). *Random Graphs and Complex Networks*, Vol. **1**. Cambridge University Press.
[17] VISWANATH, B., MISLOVE, A., CHA, M. AND GUMMADI, K. (2009). On the evolution of user interaction in Facebook. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Social Networks (WOSN '09)*, Association for Computing Machinery, New York, pp. 37–42.
[18] WAN, P., WANG, T., DAVIS, R., and RESNICK, S. (2017). Fitting the linear preferential attachment model. *Electron. J. Statist.* **11**, 3738–3780.
[19] WAN, P., WANG, T., DAVIS, R. AND RESNICK, S. (2020). Are extreme value estimation methods useful for network data? *Extremes* **23**, 171–195.
[20] WANG, T. AND RESNICK, S. (2020). Degree growth rates and index estimation in a directed preferential attachment model. *Stoch. Process. Appl.* **130**, 878–906.
[21] WASSERMAN, S. AND FAUST, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.