The Buckley-Osthus model and the block preferential attachment model: statistical analysis and application

Xin Guo [1]   **Fengmin Tang** [2]   **Wenpin Tang** [3]

Abstract

This paper is concerned with statistical estimation of two preferential attachment models: the Buckley-Osthus model and a new block preferential attachment model. We prove that the maximum likelihood estimates for both models are consistent. We perform simulation studies to corroborate our theoretical findings. We also apply both models to study the evolution of a real-world network. A list of open problems are presented.

1. Introduction

Networks are ubiquitous. A network consists of elements or actors represented by nodes or vertices, with interactions modeled by links or edges. Network studies are usually data-driven via graph theoretical analysis, aiming to determine the influence of its constituents. In the era of data deluge, there is a surge of interest in exploring the features of large networks such as the Internet (Gyöngyi et al., 2004; Bychkovsky et al., 2006), social media (Centola, 2010; Wei et al., 2016), biological systems (Jeong et al., 2001; Barabási et al., 2011), and more recently blockchains (Nakamoto, 2008; Croman et al., 2016).

Large networks are complex, and their intricate structures can be modeled by random graphs. The best-known random graph model is the *Erdös-Rényi graph* (Erdős & Rényi, 1959; 1961), where any two nodes are linked independently with a fixed probability. See (Bollobás, 2001; Durrett, 2010) for further results of the Erdös-Rényi graph such as the limiting degree distribution and emergence of the giant component.

Despite its simple form and wide applications, the Erdös-

---
[*]Equal contribution [1]Department of Industrial Engineering and Operations Research, University of California, Berkeley, USA [2]ICME, Stanford University, USA [3]Department of Industrial Engineering and Operations Research, Columbia University, USA. Correspondence to: Wenpin Tang <wt2319@columbia.edu>.

Rényi model is often criticized for the following two reasons.

- The Erdös-Rényi model does not take into account the heterogeneity, whereas a real-world network is heterogeneous, composed of a few communities or groups and exhibiting different characteristics across these communities.

- The Erdös-Rényi graph has an unrealistic Poisson degree distribution, whereas in many real-world networks, the degree of a typical node is observed to be power-law or heavy-tailed distributed.

To tackle the first point, (Holland et al., 1983) introduced the *stochastic blockmodel* (SBM) which is a generalization of the Erdös-Rényi graph with community structure. In the past decade, there has been considerable progress on the SBM including community detection (Girvan & Newman, 2002; Bickel & Chen, 2009; Mossel et al., 2015; Abbe et al., 2016; Mossel et al., 2018) and statistical estimation (Daudin et al., 2008; Bickel et al., 2011; 2013; Wang & Bickel, 2017). To address the second point, (Barabási & Albert, 1999) proposed the *preferential attachment model* (PAM), which is an instance of the *scale-free network*. It attaches a new node to existing ones according to the popularity, i.e., the degrees of the existing nodes, and the resulting degree distribution follows a power law. See also (Karrer & Newman, 2011; Zhao et al., 2012; Chen et al., 2018) for the degree-corrected SBM.

Since the work of Barabási and Albert, preferential attachment and related models have attracted much attention from combinatorics and probability communities, see e.g. (Bollobás et al., 2001; Bollobás et al., 2003; Bollobás & Riordan, 2003; 2004; Peköz et al., 2013; Bhamidi et al., 2015; Bubeck et al., 2015). But it was not until recently that statistical estimation of the PAM was investigated. (Gao & van der Vaart, 2017; Gao et al., 2017; Wang & Resnick, 2019) considered estimates of an undirected PAM, while the counterpart of a directed PAM was studied by (Wan et al., 2017).

This paper is concerned with two classes of PAM:

- the *Buckley-Osthus model* (Buckley & Osthus, 2004),

which is a one-parameter PAM allowing for self-loops;

- a new model that adds a community structure to the PAM, which we call *block preferential attachment model* (BPAM).

These two models capture different features of the underlying network. The parameter in the Buckley-Osthus model identifies the exponent of the power-law degree distribution, and the community structure in the BPAM characterizes the interactions across various groups. The relation between the BPAM and the PAM is the same as that between the SBM and the Erdös-Rényi graph.

Meanwhile, inn contrast to the SBM, the BPAM hinges on the order of nodes added, thus describing the evolution of the network. Thus, the BPAM naturally combines the idea of the SBM and the PAM, and may overcome the two drawbacks of the Erdös-Rényi model. Though some related characteristics of the BPAM are implicitly discussed in the work of (Jordan, 2013; Hajek & Sankagiri, 2018), it is the first time in this work the BPAM is explicitly formulated. We also provide statistical analysis under this framework.

In this paper, we show the asymptotic consistency of the maximum likelihood estimate (MLE) for the aforementioned two models. Precisely, we prove that (1) the MLE of the Buckley-Osthus model is asymptotically normal, (2) the MLE of the BPAM with observed community memberships is consistent. Contrary to the SBM, the likelihood of either the Buckley-Osthus model or the BPAM does not belong to the exponential family. Thus, the proof of these results does not follow from standard theory, and requires extra mathematical tools and careful analysis. We also compare the BPAM with the SBM, and apply both models to study the evolution of real-world networks, and in particular the Bitcoin network. We present a list of open problems such as community detection and efficient estimation of the BPAM. We hope that this work serves as a first step towards understanding the BPAM, a natural model for modeling network evolution with community structures.

The rest of this paper is organized as follows.

- In Section 2, we present the models of interest, and state the main results.

- In Section 3, we conduct simulation studies to corroborate our theoretical findings. We also apply the BPAM to a real-world data – Bitcoin network.

- In Section 4, we conclude with further extensions of the PAM, and a comparison of the BPAM and the SBM. There a list of open problems are presented.

The proofs of the lemma are given in the Supplementary Material.

2. Models and main results

2.1. Preferential attachment models

The PAM was proposed by Barabási and Albert (Barabási & Albert, 1999), capturing the idea that popular nodes get more attracted than less popular ones. It is closely related to the Yule model (Yule, 1925), the Matthew effect (Merton, 1968), the Price model (Price, 1976) and the Chinese restaurant process (Aldous, 1985). Roughly speaking, it generates the graph in a sequential way by attaching a new node to existing ones with probability proportional to the degree of those nodes. Thus, the resulting degree distribution has a power law. As pointed out in (Bollobás et al., 2001), the original definition of the PAM is mathematically ambiguous. A more precise description is given as follows.

- Start with two nodes linked by $m \geq 1$ edges.

- When adding a new node, add edges one a time, with the second and subsequent edges performing preferential attachment using the updated degrees.

This construction allows to reduce the PAM for any $m \geq 1$ to $m = 1$ by collapsing nodes $(k-1)m+1, \ldots, km$ to node k. See also (Bianconi & Barabási, 2001; Móri, 2002; Bollobás et al., 2003; Bollobás & Riordan, 2003) for various extensions of the PAM. In the sequel, we focus on the undirected linear PAM.

An important variant of the PAM was introduced in (Bollobás & Riordan, 2004), which is referred to as the *Linearized Chord Diagram* (LCD) or the *Bollobás-Riordan model*. It starts with a single node, labelled 1, with a self-loop. For $i \geq 1$, node $i + 1$ is attached to the graph by the following rule:

$$\mathbb{P}(i + 1 \sim v) = \begin{cases} \frac{d_i(v)}{2i+1} & \text{for} \quad v \leq i, \\ \frac{1}{2i+1} & \text{for} \quad v = i + 1, \end{cases} \quad (2.1)$$

where $d_i(v)$ is the degree of the node v at time i. It can be shown that the graph under the dynamics (2.1) has a degree distribution $P(k)$ proportional to k^{-3}. A more general model was proposed (Buckley & Osthus, 2004), allowing the degree power as a parameter. Precisely, the attachment rule is given by

$$\mathbb{P}(i + 1 \sim v) = \begin{cases} \frac{d_i(v)+a-1}{(a+1)i+a} & \text{for} \quad v \leq i, \\ \frac{a}{(a+1)i+a} & \text{for} \quad v = i + 1. \end{cases} \quad (2.2)$$

By taking $a = 1$, we get the LCD model. The degree distribution $P(k)$ of the Buckley-Osthus model is proportional to k^{-2-a}. A nice property of the Buckley-Osthus model is the exchangeability, i.e. the graph distribution does not depend

on the order of nodes added. Given a graph of n nodes and edges, the Buckley-Osthus likelihood is

$$L_n^{BO}(a) := \frac{\prod_v a^{(d(v)-1)}}{a(2a+1)(3a+2)\cdots(na+n-1)}, \quad (2.3)$$

where $a^{(n)} := \prod_{k=0}^{n-1}(a+k)$ is the *Pochhammer rising factorial*, with the convention $a^{(0)} := 1$, and $d(v)$ is the total degree of the node v. One can fit the Buckley-Osthus model by MLE, and a natural question is whether the MLE is consistent.

2.2. Block preferential attachment model

In view of the stochastic blockmodel, it is natural to add a community structure to the PAM. It allows for extra parameters to model the incentive of attachment between different communities. Precisely, the attachment rule is as follows.

- There are K communities $C_1, \ldots, C_K$. Each added node belongs to community C_j with probability π_j, $1 \leq j \leq K$ with $\sum_{j=1}^{K} \pi_j = 1$. The memberships of nodes are independent.

- $(\gamma_{kl}; 1 \leq k, l \leq K)$ is a symmetric matrix representing the interaction between communities. The probability that node $i+1$ is attached to v is proportional to $\gamma_{kl} d_i(v)$ with $i+1 \in C_k$, $v \in C_l$.

In this case,

$$\mathbb{P}(i+1 \sim v) = \begin{cases} \frac{\gamma_{kl} d_i(v)}{\sum_{v'} \gamma_{kl'} d_i(v') + \gamma_{kk}} & \text{for} \quad v \leq i, \\ \frac{\gamma_{kk}}{\sum_{v'} \gamma_{kl'} d_i(v') + \gamma_{kk}} & \text{for} \quad v = i+1. \end{cases} \quad (2.4)$$

By taking γ_{kl}'s all equal, we get the LCD model. It is easy to see that the attachment probability (2.4) is homogenous in $(\gamma_{kl}; 1 \leq k, l \leq K)$. One can take for instance $\gamma_{11} = 1$ for normalization.

The main difference between the PAM and the BPAM is that the latter is not exchangeable. That is, the graph distribution depends on the order of nodes added. To see this, we need some notations. For $n \geq 1$ and $1 \leq i, j \leq K$, let

- T_j^n be the number of nodes belonging to community C_j up to time n;

- N_j^n be the sum of degrees of nodes belonging to C_j up to time n;

- M_{ij}^n be the number of edges linking a node in C_i with one in C_j.

For $1 \leq k \leq n$, let l_k be the membership of node k, i.e. $l_k = j$ if node k belongs to community j. The BPAM

likelihood is given by

$$L_n^{BPA}(\boldsymbol{\pi}, \boldsymbol{\gamma}) := \prod_{j=1}^{K} \pi_j^{T_j^n} \frac{\prod_{i,j=1}^{K} \gamma_{ij}^{M_{ij}^n} \prod_v (d(v)-1)!}{\prod_{k=1}^{n} \left(\sum_{j=1}^{K} \gamma_{l_k j} N_j^{k-1} + \gamma_{l_k l_k} \right)}. \quad (2.5)$$

It can be seen from the likelihood (2.5) that the denominator is node order dependent, while the numerator is not. Thus, the graph likelihood depends on the whole history of the network expansion, not merely the final configuration. A first question is whether the MLE of L_n^{BPA} is consistent.

By letting t_k be the membership of the node attached by node k, the BPAM likelihood (2.5) can be rewritten as

$$L_n^{BPA}(\boldsymbol{l}; \boldsymbol{\pi}, \boldsymbol{\gamma}) = \prod_{k=1}^{n} \pi_{l_k} \frac{\prod_{k=1}^{n} \gamma_{l_k t_k} \prod_v (d(v)-1)!}{\prod_{k=1}^{n} \left(\sum_{j=1}^{K} \gamma_{l_k j} N_j^{k-1} + \gamma_{l_k l_k} \right)}. \quad (2.6)$$

Often the memberships of nodes are not observed, and this leads to considering the marginal likelihood:

$$G_n^{BPA}(\boldsymbol{\pi}, \boldsymbol{\gamma}) := \sum_{\boldsymbol{l} \in \{1,\ldots,K\}^n} L_n^{BPA}(\boldsymbol{l}; \boldsymbol{\pi}, \boldsymbol{\gamma}). \quad (2.7)$$

Note that the marginal likelihood (2.7) is invariant under relabeling the names of communities, i.e. $G_n^{BPA}(\Pi\boldsymbol{\pi}, \Pi\boldsymbol{\gamma}\Pi^T) = G_n^{BPA}(\boldsymbol{\pi}, \boldsymbol{\gamma})$ for Π a permutation matrix. It is also interesting to ask whether the MLE of G_n^{BPA} is consistent.

2.3. Theoretical results

Now we present the main results – consistency of the MLE for both the Buckley-Osthus model and the BPAM. Recall the Buckley-Osthus likelihood from (2.3). We restrict to the domain $\mathcal{D} := [\varepsilon, M] \subset (0, \infty)$, and consider the following scaled log-likelihood:

$$\ell_n^{BO}(a) := \frac{1}{n} \left(\sum_{v : d(v) \geq 2} \sum_{k=0}^{d(v)-2} \log(a+k) \right.$$
$$\left. - \sum_{k=1}^{n} \log \left(a + \frac{k-1}{k} \right) \right). \quad (2.8)$$

The MLE of the Buckley-Osthus model is defined by $\widehat{a}_n^{BO} := \operatorname{argmax}_{a \in \mathcal{D}} \ell_n^{BO}(a)$. Denote a_0 to be the true value of a. Our first result shows the consistency of $\widehat{a}_n^{BO}$.

Theorem 2.1. *Assume that* $a_0 \in \mathcal{D}$. *Then* $\widehat{a}_n^{BO} \to a_0$ *almost surely as* $n \to \infty$

Moreover, we can prove the asymptotic normality of $\widehat{a}_n^{BO}$.

Theorem 2.2. *As* $n \to \infty$,

$$\sqrt{n}(\widehat{a}_n^{BO} - a_0) \xrightarrow{(d)} \mathcal{N}\left(0, \frac{\sigma^2}{\beta^2}\right), \quad (2.9)$$

where

$$\sigma^2 := \sum_{k \geq 0} \frac{p_{>k+1}}{(a_0 + k)^2} - \frac{2}{a_0 + 1} \sum_{k \geq 0} \frac{p_{>k+1}}{a_0 + k} + \frac{1}{(a_0 + 1)^2}, \tag{2.10}$$

and

$$\beta := \sum_{k \geq 0} \frac{p_{>k+1}}{(a_0 + k)^2} - \frac{1}{(a_0 + 1)^2}, \tag{2.11}$$

with

$$p_k := \frac{(a_0 + 1)a_0^{(k-1)}}{(2a_0 + 1)^{(k)}}, \tag{2.12}$$

and $p_{>k} := \sum_{j=k+1}^{\infty} p_j$.

Now let us move onto the BPAM whose likelihood is given by (2.5). We consider the following scaled log-likelihood:

$$\ell_n^{BPA}(\boldsymbol{\pi}, \boldsymbol{\gamma}) := \frac{1}{n} \left(\sum_{j=1}^{K} T_j^n \log \pi_j + \sum_{i,j=1}^{K} M_{ij}^n \log \gamma_{ij} \right.$$
$$\left. - \sum_{k=1}^{n} \log \left(\sum_{j=1}^{K} \gamma_{l_k j} \frac{N_j^{k-1}}{k} \right) \right). \tag{2.13}$$

We write $\ell_n^{BPA}(\boldsymbol{\pi}, \boldsymbol{\gamma}) = \ell_n^{BPA}(\boldsymbol{\pi}) + \ell_n^{BPA}(\boldsymbol{\gamma})$ where

$$\ell_n^{BPA}(\boldsymbol{\pi}) := \frac{1}{n} \sum_{j=1}^{K} T_j^n \log \pi_j \tag{2.14}$$

and

$$\ell_n^{BPA}(\boldsymbol{\gamma}) := \frac{1}{n} \left(\sum_{i,j=1}^{K} M_{ij}^n \log \gamma_{ij} \right.$$
$$\left. - \sum_{k=1}^{n} \log \left(\sum_{j=1}^{K} \gamma_{l_k j} \frac{N_j^{k-1}}{k} \right) \right). \tag{2.15}$$

The MLE $(\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\gamma}})$ is defined by $\widehat{\boldsymbol{\pi}} := \operatorname{argmax}_{\boldsymbol{\pi} \in \mathcal{D}} \ell_n^{BPA}(\boldsymbol{\pi})$ over the set of probability distributions $\mathcal{D}$, and $\widehat{\boldsymbol{\gamma}} := \operatorname{argmax}_{\boldsymbol{\gamma} \in \mathcal{S}} \ell_n^{BPA}(\boldsymbol{\gamma})$ over the set of symmetric stochastic matrices $\mathcal{S}$. Denote $(\boldsymbol{\pi}^0, \boldsymbol{\gamma}^0)$ to the true parameter values. The main result is the consistency of $(\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\gamma}})$.

Theorem 2.3. *We have* $\widehat{\boldsymbol{\pi}} \to \boldsymbol{\pi}^0$, $\widehat{\boldsymbol{\gamma}} \to \boldsymbol{\gamma}^0$ *almost surely as* $n \to \infty$.

2.4. Roadmap to the proofs

Let us start with the consistency of $\widehat{a}_n^{BO}$. For $k \geq 1$, let Z_k^n be the number of nodes of degree k, and $Z_{>k}^n = \sum_{j=k+1}^{\infty} Z_j^n$ be the number of nodes of degree greater than k. The log-likelihood (2.8) can be expressed as

$$\ell_n^{BO}(a) = \sum_{k \geq 0} \frac{Z_{>k+1}^n}{n} \log(a+k) - \frac{1}{n} \sum_{k=1}^{n} \log \left(a + \frac{k-1}{k} \right). \tag{2.16}$$

The key idea to prove Theorem 2.1 is that Z_k^n/n converges almost surely to p_k, which forms a probability distribution.

Proposition 2.4. *For $k \geq 1$, $Z_k^n/n \to p_k$ almost surely, where p_k is defined by* (2.12).

Proof. For $k \geq 1$, let $N_k^n := \mathbb{E}Z_k^n$. We prove that $N_k^n/n \to p_k$ as $n \to \infty$. The idea is to establish the recursion for N_k^n. For $k = 1$,

$$N_1^{n+1} - N_1^n = \frac{(a_0 + 1)n}{(a_0 + 1)n + a_0} - \frac{a_0 N_1^n}{(a_0 + 1)n + a_0}, \tag{2.17}$$

where the first term on the r.h.s. comes from attaching the new node to any existing one, and the second term is due to loss of a degree one node being attached by the new node. For $k = 2$,

$$N_2^{n+1} - N_2^n = \frac{a_0}{(a_0 + 1)n + a_0} + \frac{a_0 N_1^n}{(a_0 + 1)n + a_0} - \frac{(a_0 + 1)N_2^n}{(a_0 + 1)n + a_0}, \tag{2.18}$$

where the first term on the r.h.s. is due to creation of a loop by the new node, the second term due to creation of a degree two node by attaching the new node to a degree one node, and the third term due to loss of a degree two node being attached by the new node. Similarly, for $k \geq 3$,

$$N_k^{n+1} - N_k^n = \frac{(k + a_0 - 2)N_{k-1}^n}{(a_0 + 1)n + a_0} - \frac{(k + a_0 - 1)N_k^n}{(a_0 + 1)n + a_0}. \tag{2.19}$$

From (2.17), we get

$$N_1^{n+1} = \left(1 - \frac{a_0}{(a_0 + 1)n + a_0} \right) N_1^n + \frac{(a_0 + 1)n}{(a_0 + 1)n + a_0}.$$

By Lemma 4.1.2 in (Durrett, 2010), $N_n^n/n \to 1/(1 + \frac{a_0}{a_0+1}) = p_1$ as $n \to \infty$. Similarly, we have $N_n^k/n \to p_k$ for all k. During a time epoch, the number of nodes with degree k differs at most two. A standard argument by combining the Azuma-Hoeffding inequality and the Borel-Cantelli lemma yields the desired result. $\square$

This leads to considering the limit of (2.16):

$$\ell_\infty^{BO}(a) := \sum_{k \geq 0} p_{>k+1} \log(a + k) - \log(a + 1), \tag{2.20}$$

where $p_{>k+1} := \sum_{j=k+2}^{\infty} p_j$. Next we show that ℓ_∞^{BO} attains its unique maximum at a_0, from which we derive the consistency of $\widehat{a}_n^{BO}$.

Lemma 2.5. *The function $\ell_\infty^{BO}(\cdot)$ has a unique maximum at a_0. Moreover, for any $\varepsilon > 0$,*

$$\sup_{a > \varepsilon} |\ell_n^{BO'}(a) - \ell_\infty^{BO'}(a)| \to 0 \quad a.s. \tag{2.21}$$

Proof of Theorem 2.1. Fix $\eta > 0$. According to Lemma 2.5, there exists $m_\eta > 0$ such that

$$\ell'_\infty(a) > m_\eta \text{ for } a \in [\varepsilon, a_0 - \eta]$$

and

$$\ell'_\infty(a) < -m_\eta \text{ for } a \in [a_0 + \eta, M],$$

and a.s. for n large enough, $\sup_{\mathcal{D}} |\ell'_n(a) - \ell'_\infty(a)| < m_\eta/2$. As a consequence,

$$\ell'_n(a) > m_\eta/2 \text{ for } a \in [\varepsilon, a_0 - \eta]$$

and

$$\ell'_n(a) < -m_\eta/2 \text{ for } a \in [a_0 + \eta, M].$$

This implies that $\widehat{a}_n^{BO} \in [a_0 - \eta, a_0 + \eta]$ a.s. Since η can be taken arbitrarily small, we prove the consistency of $\widehat{a}_n^{BO}$. $\square$

We proceed to proving the asymptotic normality of $\widehat{a}_n^{BO}$. To this end, let $v^{(k)}$ be the node being attached at time k. The scaled log-likelihood can be written as

$$\ell_n(a) = \frac{1}{n} \sum_{k=1}^n F_k(a),$$

where $F_k(a) := \log(a + d_k(v^{(k)}) - 1) - \log(a + 1 - k^{-1})$ with the convention that $d_k(v^{(k)}) = 1$ if a self-loop is formed at epoch k. Therefore,

$$\ell'_n(a) = \frac{1}{n} \sum_{k=1}^n f_k(a),$$

where

$$f_k(a) := \frac{1}{a + d_k(v^{(k)}) - 1} - \frac{1}{a + 1 - k^{-1}}. \quad (2.22)$$

Note that $\ell'_n(\widehat{a}_n^{BO}) = 0$. By Taylor expanding f_k's, we get

$$0 = \sum_{k=1}^n f_k(\widehat{a}_n^{BO}) = \sum_{k=1}^n f_k(a_0) + (\widehat{a}_n^{BO} - a_0) \sum_{k=1}^n f'_k(a^\star),$$

where $a^\star \in (a_0, \widehat{a}_n^{BO})$. Consequently,

$$\sqrt{n}(\widehat{a}_n^{BO} - a_0) = \left(-\frac{1}{n} \sum_{k=1}^n f'_k(a^\star) \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n f_k(a_0) \right). \quad (2.23)$$

The proof of Theorem 2.2 boils down to the following two lemmas.

Lemma 2.6. *As* $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n f_k(a_0) \xrightarrow{(d)} \mathcal{N}(0, \sigma^2), \quad (2.24)$$

where σ^2 *is defined by* (2.10).

Lemma 2.7. *As* $n \to \infty$,

$$\frac{1}{n} \sum_{k=1}^n f'_k(a^\star) \longrightarrow -\beta \quad \text{in probability}, \quad (2.25)$$

where β *is defined by* (2.11).

Now we consider the BPAM. Note that $\widehat{\pi}$ is the MLE of the Multinomial$(1, \pi)$ distribution. It follows from standard exponential family theory (Brown, 1986) that $\widehat{\pi}$ is consistent and asymptotically normal. The main difficulty of Theorem 2.3 is to prove that $\widehat{\gamma} \to \gamma^0$. The idea is similar to that of Theorem 2.1. We prove that $M_{ij}^n/n \to \theta_{ij}^0$ and $N_j^{n-1}/n \to p_j^0$ almost surely, with $\sum_{i,j=1}^K \theta_{ij}^0 = 1$ and $\sum_{j=1}^K p_j^0 = 2$.

Proposition 2.8. *For* $1 \le i, j \le K$, $N_j^{n-1}/n \to p_j^0$ *and* $M_{ij}^n/n \to \theta_{ij}^0$ *almost surely, where* $(p_j^0; 1 \le j \le K)$ *satisfies*

$$p_j^0 \left(1 - \sum_{l=1}^K \frac{\pi_l^0 \gamma_{lj}^0}{\sum_{k=1}^K \gamma_{lk}^0 p_k^0} \right) = \pi_j^0, \quad (2.26)$$

and

$$\theta_{ij}^0 := \begin{cases} \dfrac{\pi_i^0 \gamma_{ij}^0 p_j^0}{\sum_{k=1}^K \gamma_{ik}^0 p_k^0} + \dfrac{\pi_j^0 \gamma_{ji}^0 p_i^0}{\sum_{k=1}^K \gamma_{jk}^0 p_k^0} & \text{for} \quad i \ne j, \\[4mm] \dfrac{\pi^0 \gamma_{ii}^0 p_i^0}{\sum_{k=1}^K \gamma_{ik}^0 p_k^0} & \text{for} \quad i = j. \end{cases} \quad (2.27)$$

Proof. The idea is to establish a recursion for N_j^n, $1 \le j \le K$. Denote $(\mathcal{F}_n; n \ge 1)$ to be the natural filtration of the attachment process. We have

$$\mathbb{E}(N_j^{n+1}|\mathcal{F}_n) - N_j^n = \sum_{l \ne j} \frac{\pi_l^0 \gamma_{lj}^0 N_j^n}{\sum_k \gamma_{lk}^0 N_k^n + \gamma_{ll}^0}$$

$$+ \pi_j^0 \left[\sum_{l \ne j} \frac{\gamma_{lj}^0 N_l^n}{\sum_k \gamma_{jk}^0 N_k^n + \gamma_{jj}^0} + \frac{2\gamma_{jj}^0 (1 + N_j^n)}{\sum_k \gamma_{jk}^0 N_k^n + \gamma_{jj}^0} \right], \quad (2.28)$$

where the first term on the r.h.s. includes the contributions from a new node not in C_j attached to an existing one in C_j, the second term from a new node in C_j attached to an existing one not in C_j, and the third term from a new node in C_j attached to an existing one in C_j. Rearranging the terms in (2.28) yields

$$\mathbb{E}(N_j^{n+1}|N_j^n) = \left[1 + \sum_{l \ne j} \frac{\pi_l^0 \gamma_{lj}^0}{\sum_k \gamma_{lk}^0 N_k^n + \gamma_{ll}^0} \right.$$

$$\left. + \frac{2\pi_j^0 \gamma_{jj}^0}{\sum_k \gamma_{jk}^0 N_k^n + \gamma_{jj}^0} \right] N_j^n + \pi_j^0 \frac{\sum_{l \ne j} \gamma_{lj}^0 N_l^n + 2\gamma_{jj}^0}{\sum_k \gamma_{jk}^0 N_k^n + \gamma_{jj}^0}.$$

Now we explain how (2.26) comes from. One expects $N_k^n \sim p_k^0 n$. Then the r.h.s. of the above expression is approximately

$$\left[1 + \frac{1}{n} \sum_{l \neq j} \frac{\pi_l^0 \gamma_{lj}^0}{\sum_k \gamma_{lk}^0 p_k^0} + \frac{1}{n} \frac{2\pi_j^0 \gamma_{jj}^0}{\sum_k \gamma_{jk}^0 p_k^0} \right] N_j^n$$

$$+ \pi_j^0 \frac{\sum_{l \neq j} \gamma_{lj}^0 p_l^0}{\sum_k \gamma_{jk}^0 p_k^0} = \left[1 + \frac{1}{n} \sum_l \frac{\pi_l^0 \gamma_{lj}^0}{\sum_k \gamma_{lk}^0 p_k^0} \right] N_j^n + \pi_j^0.$$

Now by Lemma 4.1.2 in (Durrett, 2010), $(p_j^0; 1 \leq j \leq K)$ satisfies (2.26). These facts can be justified by a routine argument as in Section 3, (Jordan, 2013). The full details are left for the readers.

Now we move to the convergence of M_{ij}^n/n. Similarly, we get the recursion

$$\mathbb{E}(M_{ij}^{n+1}|\mathcal{F}_n) - M_{ij}^n$$

$$= \begin{cases} \dfrac{\pi_i^0 \gamma_{ij}^0 N_j^n}{\sum_k \gamma_{ik}^0 N_k^n + \gamma_{ii}^0} + \dfrac{\pi_j^0 \gamma_{ji}^0 N_i^n}{\sum_k \gamma_{jk}^0 N_k^n + \gamma_{jj}^0} & \text{for} \quad i \neq j, \\[1em] \dfrac{\pi_i^0 \gamma_{ii}^0 (1+N_i^n)}{\sum_k \gamma_{ik}^0 N_k^n + \gamma_{ii}^0} & \text{for} \quad i = j. \end{cases}$$

$$(2.29)$$

It is easy to see from the asymptotics of N_j^n that M_{ij}^n/n converges to θ_{ij}^0 satisfying (2.27). □

The above result yields the limit of (2.15):

$$\ell_\infty^{BPA}(\boldsymbol{\gamma}) := \sum_{i,j=1}^K \theta_{ij}^0 \log \gamma_{ij} - \sum_{i=1}^K \pi_i^0 \log \left(\sum_{j=1}^K \gamma_{ij} p_j^0 \right).$$

$$(2.30)$$

Note that ℓ_∞^{BPA} is homogeneous of order 0, i.e. $\ell_\infty^{BPA}(\theta\boldsymbol{\gamma}) = \ell_\infty^{BPA}(\boldsymbol{\gamma})$ for each $\theta > 0$. It suffices to prove that ℓ_∞^{BPA} attains its unique maximum at the equivalent class $[\boldsymbol{\gamma}^0]$, from which the consistency of $\widehat{\boldsymbol{\gamma}}$ follows.

Lemma 2.9. *The function $\ell_\infty^{BPA}(\cdot)$ has a unique maximum at $\boldsymbol{\gamma}_0$ up to a constant multiple.*

All the remaining proofs are deferred to the Supplementary Material.

3. Empirical results

3.1. Buckley-Osthus model

Theorems 2.1 and 2.2 show that the MLE of the Buckley-Osthus model is consistent and asymptotically normal. Here we present some simulation experiments to illustrate statistical inference for finite samples. We pick three different values for the parameter, $a_0 \in \{0.5, 1.0, 2.0\}$. For each parameter value, we generate 200 realizations from

the Buckley-Osthus dynamics (2.2) with sample sizes $n = 100, 200, 500, 1000$.

We fit the Buckley-Osthus model by MLE, and calculate the MLE $\widehat{a}^{BO}$ via the gradient ascent algorithm. So for each parameter value and sample size, there are 200 estimates. Figure 1 gives the boxplots of these MLEs for each parameter $a_0 \in \{0, 5, 1.0, 2.0\}$ with sample sizes $n = 100.200, 500, 1000$. Table 1 summarizes a few statistics of the MLEs in different experiment settings.
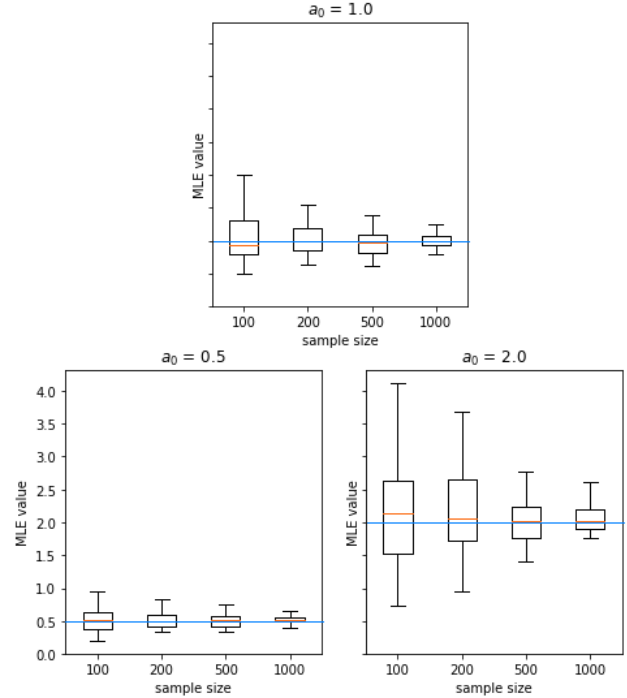

Figure 1. Boxplots of the MLEs for the Buckley-Osthus model with different parameter values and sample sizes.

It can be seen from Figure 1 and Table 1 that for $n \gtrsim 500$, the mean/median of the MLEs are close to the true parameter values. For each parameter value, the standard deviation of the MLEs decreases as the sample size increases. Given the sample size, the standard deviation of the MLEs increases as the parameter value gets larger. These observations agree with the theoretical findings in Section 2.3.

3.2. Block preferential attachment model

Theorem 2.3 proves that the MLE of the BPAM is consistent. Here we corroborate this result with some simulations. We take $K = 2$ with $(\pi_1, \pi_2) = (0.3, 0.7)$, and $\gamma_{12} = \gamma_{21} = 0.5$ and $\gamma_{22} = 1.5$. For this parameter setting, we generate 100 realizations from the BPAM dynamics (2.4) with sample sizes $n = 100, 200, 500, 1000$.

$a_0 = 0.5$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Mean	0.543	0.511	0.510	0.506
Median	0.526	0.516	0.515	0.511
Std	0.200	0.114	0.100	0.062
$a_0 = 1.0$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Mean	1.125	1.098	0.977	1.011
Median	0.936	0.997	0.977	0.990
Std	0.645	0.365	0.182	0.123
$a_0 = 2.0$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Mean	2.540	2.354	2.081	2.085
Median	2.130	2.060	2.027	2.025
Std	1.708	1.008	0.487	0.251

Table 1. The mean, median and standard deviation of the MLEs for the Buckley-Osthus model with different parameter values and sample sizes.

We calculate the MLEs $(\widehat{\pi}_1, \widehat{\pi}_2)$ by counting frequency, and the MLEs $(\widehat{\gamma}_{12}, \widehat{\gamma}_{22})$ via the gradient ascent algorithm. For each sample size, we get 100 sets of estimates. Figure 2 provides boxplots of the MLEs with sample sizes $n = 100, 200, 500, 1000$. Table 2 displays some statistics of the MLEs with different sample sizes.

$\pi_1 = 0.3$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Mean	0.291	0.297	0.302	0.303
Median	0.300	0.300	0.301	0.301
Std	0.071	0.036	0.033	0.024
$\pi_2 = 0.7$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Mean	0.709	0.703	0.698	0.697
Median	0.700	0.700	0.698	0.699
Std	0.071	0.036	0.033	0.024
$\gamma_{12} = 0.5$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Mean	0.736	0.486	0.476	0.520
Median	0.478	0.435	0.460	0.496
Std	0.837	0.203	0.122	0.098
$\gamma_{22} = 1.5$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Mean	1.680	1.555	1.480	1.529
Median	1.529	1.609	1.409	1.432
Std	1.842	1.457	0.588	0.514

Table 2. The mean, median and standard deviation of the MLEs for the HPAM with different sample sizes.

It can be observed from Figure 2 and Table 2 that the MLEs $(\widehat{\pi}_1, \widehat{\pi}_2)$ get very close to the true values for $n \gtrsim 100$, while the MLEs $(\widehat{\gamma}_{12}, \widehat{\gamma}_{22})$ seem to converge when $n \approx 1000$. This is partly due the stability of the gradient ascent algorithm to find the MLE of $\boldsymbol{\gamma}$. Also for each parameter, the standard deviation decreases as the sample size increases.
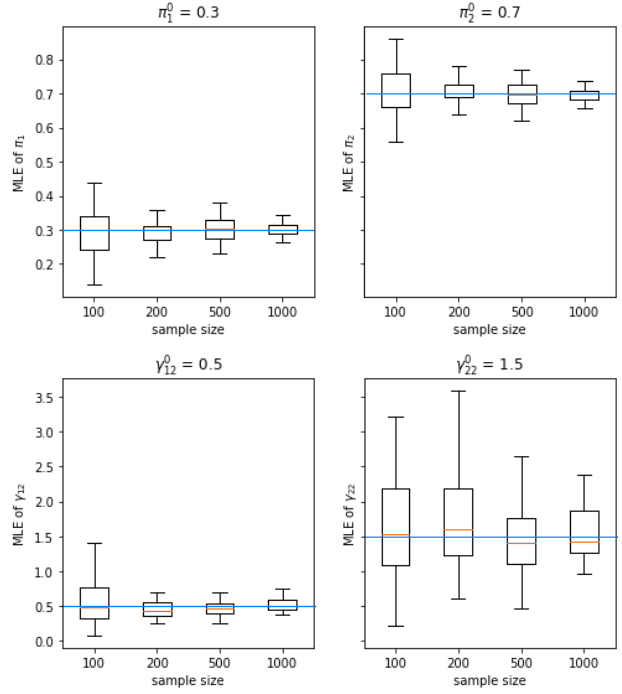

Figure 2. Boxplots of the MLEs of $(\pi_1, \pi_2, \gamma_{12}, \gamma_{22})$ for the HPAM with different sample sizes.

3.3. Evolution of the Bitcoin network

Bitcoin (Nakamoto, 2008) is a distributed digital currency system which works without central governing. Payments are processed by a peer-to-peer network of users connected through the internet. Here we apply the BPAM to the Bitcoin network. The data we use is available at http://www.vo.elte.hu/bitcoin/zipdescription.htm (Kondor et al., 2014).

In such a transaction, 'A' is the receiver and 'B' is the sender. The data set consists of 30048983 transactions up to December, 2013. All these transactions are recorded with timestamps. In the Bitcoin network, there are two types of nodes: regular nodes and super nodes. Regular nodes represents normal users, while super nodes are professional miners or digital currency companies. Super nodes are usually much more equipped than regular ones, and are more reliable in the transactions. Hence, we model the Bitcoin network by the BPAM with $K = 2$. The index '1' is used for super nodes, and '2' for regular nodes. The idea is to fit the Buckley-Osthus model and the BPAM by MLE, and calculate the MLEs for the parameters of interest.

We preprocess the data by removing a few abnormal transactions, for instance, those related to the *SatoshiDice gambling* whose addresses start with '1Dice'. Due to computation limits, we are unable to process all $3 \cdot 10^8$ transactions. In-

stead, we consider the first $n = 5000, 10000, 20000, 50000$ transactions to get the corresponding parameter estimates. We select a threshold of top 5% active users to distinguish super nodes from regular nodes. Table 3 displays the MLE for the power-law exponent with different network sizes. It can be seen that the exponent stabilizes around 4.4 for $n \gtrapprox 20000$.

	$n = 5000$	$n = 10000$	$n = 20000$	$n = 50000$
$\widehat{a}$	3.757	3.857	4.401	4.398

Table 3. The MLE for a of the Bitcoin network.

Table 4 displays the MLEs for the BPAM parameters with different network sizes. Observe that (1) the proportion of super nodes $\widehat{\pi}_1$ decreases as the network expands, (2) $\widehat{\gamma}_{12} > \widehat{\gamma}_{22}$ for $n = 5000$, and $\widehat{\gamma}_{12} < \widehat{\gamma}_{22}$ for $n \gtrapprox 10000$. These phenomena can be interpreted as follows. At the early stage of the Bitcoin, many transactions were made through super nodes, and professional miners were the main player. As the blockchain techniques are developed, more and more individuals participate in Bitcoin transactions. This is the reason why we see an order change between $\widehat{\gamma}_{12}$ and $\widehat{\gamma}_{22}$. The BPAM characterizes the evolution of the Bitcoin network.

	$n = 5000$	$n = 10000$	$n = 20000$	$n = 50000$
$\widehat{\pi}_1$	0.410	0.312	0.225	0.199
$\widehat{\pi}_2$	0.590	0.688	0.775	0.801
$\widehat{\gamma}_{12}$	3.974	9.226	16.351	34.596
$\widehat{\gamma}_{22}$	2.360	11.213	67.216	123.497

Table 4. The MLEs for $(\pi_1, \pi_2, \gamma_{12}, \gamma_{22})$ of the Bitcoin network.

4. Conclusion and discussion

This paper deals with statistical estimation of two preferential attachment related models: the Buckley-Osthus model, and the BPAM. We prove the consistency of the MLE for both models, and corroborate the theory with simulation studies. Though these models might be too simplistic for real-world networks (as shown in the Bitcoin example), they can capture some gross features and structural changes in the network. Despite their limitations, these models may be used as a building piece for more flexible systems.

In the remaining of this section, we discuss a few open problems related to the BPAM. As shown in Theorem 2.3, if the memberships of all nodes are correctly classified, then the MLEs are consistent. In fact, even if a small portion of memberships (e.g. of order $o(n)$) are misclassified, the consistency still holds. Consider the case where all but the last node are correctly classified. It is easy to see that the

statistics (M_{ij}^n, N_j^n) is affected by ± 1, and the limiting log-likelihood (2.30) stays the same. More generally, the degree of any node is of order n^β for some $\beta < 1$. Thus, misclassifying any node will affect (M_{ij}^n, N_j^n) by $\pm n^\beta$, which is absorbed by the $1/n$-scaling. This leads to the problem of community detection of the BPAM. (Hajek & Sankagiri, 2018) gave a message passing algorithm to recover the memberships of $\Theta(1)$ nodes. An important question is to which extent the memberships of nodes can be recovered for the BPAM.

Open Problem 4.1. *Is there any algorithm to recover the memberships of all nodes for the BPAM ? If not, what is the maximum number of memberships which can be recovered ? Is it of order $\Theta(n)$?*

For the SBM, the MLEs of the marginal likelihood were shown to be asymptotically normal (Bickel et al., 2013). The key idea is that the likelihood with misclassifications is much smaller than that with all correct memberships. However, this is not true for the BPAM since misclassifying a small portion of memberships will not affect much the likelihood. It is not clear whether the MLEs of the marginal likelihood (2.7) give the correct memberships and good parameter estimates.

Open Problem 4.2. *Are the MLEs for the marginal likelihood (2.7) consistent ?*

From the computational viewpoint, the marginal likelihood (2.7) is intractable so calculating the MLE is not efficient. In the case of the SBM, (Daudin et al., 2008) applied the variational inference to remove the normalizing term so that the optimization can be easily solved by the EM algorithm. The corresponding estimates were also shown to be asymptotically normal (Bickel et al., 2013). Here we can ask similar questions regarding the marginal likelihood for the BPAM.

Open Problem 4.3.

1. *Propose an approximation of the MLEs for the marginal likelihood (2.7) which is computationally feasible.*

2. *Is the approximation in (1) consistent ?*

3. *How to choose the number of communities K ?*

References

Abbe, E., Bandeira, A. S., and Hall, G. Exact recovery in the stochastic block model. *IEEE Trans. Inform. Theory*, 62(1):471–487, 2016.

Aldous, D. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pp. 1–198. Springer, Berlin, 1985.

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

Barabási, A.-L., Gulbahce, N., and Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1):56, 2011.

Bhamidi, S., Steele, J. M., and Zaman, T. Twitter event networks and the superstar model. *Ann. Appl. Probab.*, 25(5):2462–2502, 2015.

Bianconi, G. and Barabási, A. L. Competition and multi-scaling in evolving networks. *Europhysics Letters*, 54(4): 436, 2001.

Bickel, P. J. and Chen, A. A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA*, 106(50):21068–21073, 2009.

Bickel, P. J., Chen, A., and Levina, E. The method of moments and degree distributions for network models. *Ann. Statist.*, 39(5):2280–2301, 2011.

Bickel, P. J., Choi, D., Chang, X., and Zhang, H. Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.*, 41(4):1922–1943, 2013.

Bollobás, B. *Random Graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, second edition, 2001.

Bollobás, B. and Riordan, O. Mathematical results on scale-free random graphs. In *Handbook of graphs and networks*, pp. 1–34. Wiley-VCH, Weinheim, 2003.

Bollobás, B. and Riordan, O. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004.

Bollobás, B., Riordan, O., Spencer, J., and Tusnády, G. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, 18(3):279–290, 2001.

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

Brown, L. D. *Fundamentals of statistical exponential families with applications in statistical decision theory*, volume 9 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. 1986.

Bubeck, S., Mossel, E., and Rácz, M. On the influence of the seed graph in the preferential attachment model. *IEEE Trans. Network Sci. Eng.*, 2(1):30–39, 2015.

Buckley, P. G. and Osthus, D. Popularity based random graph models leading to a scale-free degree sequence. *Discrete Math.*, 282(1-3):53–68, 2004.

Bychkovsky, V., Hull, B., Miu, A., Balakrishnan, H., and Madden, S. A measurement study of vehicular internet access using in situ Wi-Fi networks. In *Proceedings of the Twelvth Annual International Conference on Mobile Computing and Networking*, pp. 50–61, 2006.

Centola, D. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197, 2010.

Chen, Y., Li, X., and Xu, J. Convexified modularity maximization for degree-corrected stochastic block models. *Ann. Statist.*, 46(4):1573–1602, 2018.

Croman, K., Decker, C., Eyal, I., Gencer, A., Juels, A., Kosba, A., Miller, A., Saxena, P., Shi, E., and Sirer, E. On scaling decentralized blockchains. In *International Conference on Financial Cryptography and Data Security*, pp. 106–125, 2016.

Daudin, J. J., Picard, F., and Robin, S. A mixture model for random graphs. *Stat. Comput.*, 18(2):173–183, 2008.

Durrett, R. *Random graph dynamics*, volume 20 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2010.

Erdős, P. and Rényi, A. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

Erdős, P. and Rényi, A. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.*, 12: 261–267, 1961.

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

Gao, F., van der Vaart, A., Castro, R., and van der Hofstad, R. Consistent estimation in general sublinear preferential attachment trees. *Electron. J. Stat.*, 11(2):3979–3999, 2017.

Girvan, M. and Newman, M. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, 99 (12):7821–7826, 2002.

Gyöngyi, Z., Garcia-Molina, H., and Pedersen, J. Combating web spam with trustrank. In *Proceedings of the Thirtieth International Conference on Very Large Data Bases*, pp. 576–587, 2004.

Hajek, B. and Sankagiri, S. Community recovery in a preferential attachment graph. *arXiv:1801.06818*, 2018.

Hall, P. and Heyde, C. *Martingale Limit Theory and its Application*. Academic Press, 1980.

Holland, P. W., Laskey, K. B., and Leinhardt, S. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

Jeong, H., Mason, S., Barabási, A.-L., and Oltvai, Z. Lethality and centrality in protein networks. *Nature*, 411(6833): 41, 2001.

Jordan, J. Geometric preferential attachment in non-uniform metric spaces. *Electron. J. Probab.*, 18:no. 8, 15, 2013.

Karrer, B. and Newman, M. Stochastic blockmodels and community structure in networks. *Physical review E*, 83 (1):016107, 2011.

Kondor, D., Pósfai, M., Csabai, I., and Vattay, G. Do the rich get richer? An empirical analysis of the Bitcoin transaction network. *PLOS ONE*, 9(2):e86197, 2014.

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

Móri, T. F. On random trees. *Studia Sci. Math. Hungar.*, 39 (1-2):143–155, 2002.

Mossel, E., Neeman, J., and Sly, A. Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields*, 162(3-4):431–461, 2015.

Mossel, E., Neeman, J., and Sly, A. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018.

Nakamoto, S. Bitcoin: A peer-to-peer electronic cash system. 2008. Available at https://bitcoin.org/bitcoin.pdf.

Peköz, E. A., Röllin, A., and Ross, N. Degree asymptotics with rates for preferential attachment random graphs. *Ann. Appl. Probab.*, 23(3):1188–1218, 2013.

Price, D. A general theory of bibliometric and other cumulative advantage processes. *J. Am. Soc. Inf. Sci.*, 27(5): 292–306, 1976.

Wan, P., Wang, T., Davis, R. A., and Resnick, S. I. Fitting the linear preferential attachment model. *Electron. J. Stat.*, 11(2), 2017.

Wang, T. and Resnick, S. I. Consistency of Hill estimators in a linear preferential attachment model. *Extremes*, 22 (1):1–28, 2019.

Wang, Y. X. R. and Bickel, P. J. Likelihood-based model selection for stochastic block models. *Ann. Statist.*, 45 (2):500–528, 2017.

Wei, W., Joseph, K., Liu, H., and Carley, K. Exploring characteristics of suspended users and network stability on Twitter. *Social Network Analysis and Mining*, 6(1):51, 2016.

Yule, G. U. A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis, F.R.S. *J. R. Stat. Soc. Ser. B.*, 88(3):433–436, 1925.

Zhao, Y., Levina, E., and Zhu, J. Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.*, 40(4):2266–2292, 2012.