Estimation of a Scale-Free Network Formation Model*

Anton Kolotilin†

This version: December, 2013

Abstract

Growing evidence suggests that many social and economic networks are scale free in that their degree distribution $P(d)$ is approximately proportional to $d^{-\gamma}$. The most widespread explanation for this phenomenon is a random network formation process with preferential attachment. For a general version of such a process, we develop a class of GMM estimators. We show formally that these GMM estimators give consistent estimates of model parameters. Simulations suggest that the GMM estimates are asymptotically normally distributed. The commonly used NLLS estimator gives highly biased and inconsistent estimates; Hill (1975) estimator performs even worse.

JEL Codes: C15, C45, C51, D85

Keywords: consistency, degree distribution, network formation, scale-free network

*Most of this paper has been written during my PhD at MIT, whose hospitality is gratefully acknowledged. I thank Denis Chetverikov and Victor Chernozhukov for numerous helpful comments that significantly improved the paper. I also thank Isaiah Andrews, Arun Chandrasekhar, Jerry A. Hausman, Guido Imbens, Anna Mikusheva, Whitney Newey, Valentyn Panchenko, Chad Syverson, and the participants at MIT Econometrics Lunch for helpful comments. All errors are mine.

†UNSW, School of Economics. Email: a.kolotilin@unsw.edu.au.

1 Introduction

Many real networks have a *degree distribution with a power-law tail.*[1] That is, the fraction $P(d)$ of nodes that have d neighbors is approximately proportional to $d^{-\gamma}$ for large d, where γ is a positive constant called the *power-law parameter*. Such networks are called *scale-free*. The power-law parameter plays a crucial role for the topology of a network and for the network's statistical properties, such as learning, the spread of viruses, the size of the largest component, the connectivity, the searchability, and the robustness to errors and attacks (Albert and Barabasi (2002)). In this paper, we estimate the power-law parameter and other parameters for a general model of random scale-free network formation.

In their seminal article, Barabasi and Albert (1999) build the first theoretical model of scale-free network formation (hereafter the BA model):

> "...starting with a small number (m_0) of vertices, at every time step we add a new vertex with $m(\leq m_0)$ edges that link the new vertex to m different vertices already present in the system. To incorporate preferential attachment, we assume that the probability Π that a new vertex will be connected to vertex i depends on the connectivity k_i of that vertex, so that $\Pi(k_i) = k_i / \sum_j k_j$. After t time steps, the model leads to a random network with $t + m_0$ vertices and mt edges."

The idea of the model is that the rich get richer: more "popular" nodes get more links than less popular nodes as a network evolves. Such a process is called *preferential attachment*. The BA model initiated further wide-range investigation and modelling of scale-free networks.[2]

Cooper and Frieze (2003) and Cooper (2006) introduce and analyze a general model of scale-free network formation (hereafter the CF model). This model nests many models of scale-free network formation (Section 2.2), including the BA model and popular hybrid models, such as Jackson and Rogers (2007). In the CF model, initially, there is a small fixed network. At each subsequent period, a new node with a random number of links is added. Some of added links connect a new node with the existing network, and others connect old nodes between themselves. The sampling method for choosing link endpoints is

[1] Such networks include social networks (coauthorship, citation, movie actor, and sexual relation networks), biological networks (ecological and food webs, cellular, protein and neural networks) and other networks (WWW, Internet, Fedwire interbank market, and power grid networks). See Albert and Barabasi (2002) and Dorogovtsev and Mendes (2002).

[2] For a non-rigorous overview of scale-free networks literature, see Albert and Barabasi (2002), Dorogovtsev and Mendes (2002) and chapter 5 of Jackson (2008). For a rigorous treatment of scale-free random graph processes, see Bollobas and Riordan (2003).

decided uniformly at random with some probability and by preferential attachment with the complementary probability.

Cooper (2006) shows that the asymptotic degree distribution depends only on the following subset of parameters: the share of links inserted by preferential attachment (η), the number of link endpoints added uniformly at random (ν); and the initial degree distribution of added nodes ($\boldsymbol{p}$). Moreover, the asymptotic degree distribution is a linear combination (with coefficients $\boldsymbol{p}$) of Yule-Simon distributions (see Simon (1955)). Most interestingly, the asymptotic degree distribution has a power-law tail, where the power-law parameter γ is determined only by η: $\gamma = 1 + 1/\eta$. The goal of this paper is to estimate these parameters; these estimates give us an idea what the underlying network formation process was.

Despite a variety of theoretical models of network formation, there is a lack of rigorous econometrics papers that estimate parameters of scale-free network formation models. Many papers just plot the transformed degree distribution for real networks on the log-log scale and calculate the slope of the tail of the distribution to estimate the power-law parameter. More sophisticated papers, such as Pennock et al. (2002) and Jackson and Rogers (2007), use non-linear least squares procedures (hereafter NLLS) to fit the empirical degree distribution to an approximation of the asymptotic degree distribution. Goldstein et al. (2004) illustrate that although such procedures give good graphical fits of empirical and estimated distributions on the log-log scale, they give biased and inaccurate estimates of model parameters. Goldstein et al. (2004) also argue that the maximum likelihood estimation is much more robust. Unfortunately, as Jackson and Rogers (2007) note, deriving analytically and then computing numerically the true likelihood of all possible degree distributions appears to be intractable for scale-free network formation models.

Atalay et al. (2011) and Atalay (2013) estimate parameters of a network formation model using a pseudo maximum likelihood estimator (hereafter PML). Specifically, they calculate the likelihood assuming that each node degree is independent and identically distributed according to the asymptotic degree distribution, which is found using Dorogovtsev et al. (2000) method. However, in their network formation models, node degrees are interdependent and have different distributions even in asymptotics ("old" nodes have a much higher degree than "young" nodes). Therefore, it is not clear what asymptotic properties the PML estimator have.

To estimate parameters of the CF model, we develop a class of generalized method of moments (hereafter GMM) estimators, which includes the PML estimators as a special case. This GMM estimation is computationally simple, because it requires calculating only a sample average of a moment function, as opposed to the true likelihood of all possible degree distributions. Our main contribution is to show formally that the GMM estimators give consistent estimates of the CF model parameters. Standard consistency results rely on the uniform law of large numbers for independent or stationary data. Because node degrees are interdependent in a non-standard way in the CF model, we prove the uniform law of large numbers using first principles.

Simulations suggest that the GMM estimates are asymptotically normally distributed. Moreover, the GMM estimators have a small bias even for small networks. The NLLS method, on the other hand, gives inconsistent estimates with large biases. Moreover, the NLLS method can be used only for particular subcases of the CF model, whereas the GMM estimators can be used to estimate most general versions of the CF model. Hill (1975) estimates are also inconsistent with even larger biases and standard errors than the NLLS estimates.

In Chapters 4, 5, and 6, Jackson (2008) classifies network formation models into *random networks*, *growing random networks*, and *strategic networks*, respectively, where to some extent each subsequent class can be viewed as more realistic but harder to analyze. Chandrasekhar and Jackson (2013) estimate models in a general class of random networks and formally show that the maximum likelihood estimator is consistent and asymptotically normally distributed. This paper estimates models in a general class of growing random networks and formally shows that the GMM estimators are consistent. Christakis et al. (2010), Konig (2012), Sheng (2012), Comola and Fafchamps (2013), Goldsmith-Pinkham and Imbens (2013), and Mele (2013) use Bayesian methods to estimate models in a class of strategic networks.

The rest of the paper is organized as follows. Section 2 sets up and discusses the model. Section 3 analyzes the model. Section 4 describes the three estimation methods (the GMM, NLLS, and Hill estimators) and presents asymptotic properties of the GMM estimators. Section 5 compares statistical properties of estimators using simulations of the CF model. Section 6 concludes. Appendix A provides partial results on asymptotic normality of the GMM estimators. All proofs are relegated to Appendix B.

2 Model and Discussion

Section 2.1 sets up the CF model of random scale-free network formation, analyzed and estimated in this paper. Section 2.2 discusses the relation between the CF model and other models of scale-free network formation. In particular, it shows that the CF model nests many other scale-free network formation models.

2.1 CF Model

Consider the following random multigraph process, $(G(t))_{t\geq1} = (V(t), E(t))_{t\geq1}$, which is a slight modification of the model introduced by Cooper and Frieze (2003) and further analyzed by Cooper (2006).[3] $G(1)$ is a multigraph that contains $|V(1)| \geq 1$ vertices and $|E(1)| \geq 1$ edges (the number of elements of an arbitrary finite set X is denoted by $|X|$ hereafter). For $t \geq 2$ the random graph $G(t)$ is obtained from $G(t-1)$ by adding one vertex indexed by t with a random number of edges $m(t)$, whose initial endpoints are t and terminal endpoints lie in $V(t-1)$, and a random number of edges $M(t)$, whose endpoints lie in $V(t-1)$.[4] Both $m(t)$ and $M(t)$ are independently distributed (among themselves and across time) according to finite support distributions $\boldsymbol{p} = (p_0, ..., p_m, ...p_P)$ and $\boldsymbol{q} = (q_0, ..., q_M, ..., q_Q)$ respectively, where $p_m = \Pr(m(t) = m)$, $q_M = \Pr(M(t) = M)$, and $P, Q < \infty$. Denote $\overline{m} = \mathbb{E}(m(t))$ and $\overline{M} = \mathbb{E}(M(t))$. We assume that there is a positive probability that at least one edge will be added, i.e. $\overline{m} + \overline{M} > 0$. Denote the degree of a vertex v of the graph $G(t)$ by $d(v, t)$.

First, consider edges $e_i^m(t)$, $i = 1, ..., m(t)$, originated from t. Their terminal endpoints are chosen independently with probability A_1 by preferential attachment from $V(t-1)$ (i.e. the probability that v is the terminal endpoint of $e_i^m(t)$ is proportional to the degree of this vertex $d(v, t-1)$),[5] and with probability $A_2 = 1 - A_1$ uniformly at random from $V(t-1)$.

[3] *Graph* and *network* are used interchangeably throughout the paper. A *graph* G is an ordered pair (V, E), where V is a set of elements called *vertices* or *nodes*, and E is a set of unordered pairs of V called *edges* or *links*. Informally, a graph is a set of vertices together with a set of edges that join these vertices. Vertices x and y of an edge $\{x, y\}$ are called *endpoints* of this edge. A vertex y is called a *neighbor* of a vertex x if $\{x, y\} \in E$. The *degree* of a vertex x is the number of neighbors of x. The *degree distribution* of a graph, $P(.)$, is a description of the relative frequencies of vertices that have different degrees; specifically, $P(d)$ is the fraction of vertices that have degree d.

A *random multigraph process* $(G(t))_{t\geq1}$ is a sequence of random multigraphs $G(t) = (V(t), E(t))$, where $V(t)$ is a set of vertices and $E(t)$ is a set of edges at time step t. A *multigraph* as opposed to a graph can contain *loops* (i.e., edges joining a vertex to itself) and *multiple edges* (i.e., several edges joining the same two vertices). Inspired by Bollobas et al. (2001), we expect that the fraction of multiple edges and loops goes to 0 as $t \to \infty$ for the considered multigraph process.

[4] We distinguish between initial and terminal endpoints only to describe the random multigraph process. In the rest of the paper, we treat all edges as undirected, but it is straightforward to extend the analysis to directed graph processes.

[5] The reader who is interested in microfoundations of preferential attachment is referred to Section 2.2. Moreover, as discussed in Section 2.2, the network is scale-free if and only if the attachment probability is asymptotically linear. These facts serve as a justification of the assumption that the probabilities are proportional to the degrees.

To summarize,

$$p_A(v,t) \equiv \Pr(v \text{ is a terminal endpoint of } e_i^m(t)) = A_1 \frac{d(v,t-1)}{2|E(t-1)|} + A_2 \frac{1}{|V(t-1)|}.$$

Second, consider edges $e_i^M(t)$, $i = 1, ..., M(t)$. The initial endpoint (respectively the terminal endpoint) of each edge $e_i^M(t)$ is chosen independently with probability B_1 (respectively C_1) by preferential attachment from $V(t-1)$ and with probability $B_2 = 1 - B_1$ (respectively $C_2 = 1 - C_1$) uniformly at random from $V(t-1)$. Thus, we have

$$\begin{aligned} p_B(v,t) &\equiv \Pr(v \text{ is an initial endpoint of } e_i^M(t)) = B_1 \frac{d(v,t-1)}{2|E(t-1)|} + B_2 \frac{1}{|V(t-1)|}, \\ p_C(v,t) &\equiv \Pr(v \text{ is a terminal endpoint of } e_i^M(t)) = C_1 \frac{d(v,t-1)}{2|E(t-1)|} + C_2 \frac{1}{|V(t-1)|}. \end{aligned}$$

The degree distribution of a random graph is itself a random object. To clarify this issue let us define $D_t(d)$ as the number of vertices of the graph $G(t)$ that have degree d. Then the fraction of vertices of the random graph $G(t)$ that have degree d is $P_t(d) \equiv D_t(d)/|V(t)|$, which is a random variable. Section 3 shows that for all d the fraction $P_t(d)$ converges in probability to $P(d)$ as t goes to infinity. The limiting fractions $P(d)$ are called the *asymptotic degree distribution* of the multigraph process $(G(t))_{t\geq 1}$. Cooper (2006) showed that the asymptotic degree distribution of the multigraph process $(G(t))_{t\geq 1}$ depends only on $\boldsymbol{p}$ and on

$$\begin{aligned} \eta &\equiv \frac{\overline{m}A_1 + \overline{M}(B_1 + C_1)}{2(\overline{m} + \overline{M})}, \\ \nu &\equiv \overline{m}A_2 + \overline{M}(B_2 + C_2). \end{aligned}$$

> "The parameters η and ν summarize the preferential attachment and uniform at random aspects of the process respectively. The parameter η is the limiting ratio of the expected number of edge endpoints inserted by preferential attachment to the expected total degree of the graph. Similarly, ν is the average number of edges inserted uniformly at random per vertex."

As in Cooper and Frieze (2003), we assume that parameters are such that $0 < \eta < 1$ holds.[6]

[6] This model differs from the original CF model in that we allow vertices to have zero degree to conform with real network data (in the original CF model $p_0 = 0$ and $q_0 = 0$, i.e. each vertex has at least one neighbor). Nevertherless, if $0 < \eta < 1$, all results and corresponding proofs from Cooper (2006) remain valid.

2.2 Discussion

The CF model nests many other models of scale-free network formation in the sense that it can produce the same degree distribution or at least the same asymptotic degree distribution. However, the CF model is general only in the class of scale-free graph processes and it does not nest, for example, the exponential family of graphs. Moreover, the CF model is aimed to explain the degree distribution of a graph rather than other graph characteristics such as assortativity, betweenness centrality, clustering, and spectrum. For example, the CF and copying models have the same degree distribution, but they have very different clustering coefficients (Jackson and Rogers (2007)).

It is easy to see that the BA model described in the Introduction is a particular subcase of the CF model with $p_m = 1,\ q_0 = 1$, and $A_1 = 1$; so $\eta = 1/2$, and $\nu = 0$. Below we show that many scale-free network formation models that generalize the BA model are nested by the CF model.

There is a variety of hybrid models in which edge endpoints are chosen by a mix of random and preferential attachment. Most of these models are just particular subcases of the CF model. For example, Pennock et al. (2002) is a subcase with $p_0 = 1,\ q_m = 1,$ $B_1 = C_1 = \alpha$; the hybrid model considered in chapter 5.3 of Jackson (2008) is a subcase with $p_m = 1,\ q_0 = 1,\ A_1 = 1 - \alpha$.

Dorogovtsev et al. (2000) and Buckley and Osthus (2004) models are also nested by the CF model. In their models, each period t a new vertex indexed by t appears with m new edges originated from it. The probability that vertex v is chosen as a terminal endpoint of a new edge is proportional to the sum of initial attractiveness A and degree $d(v, t-1)$ of vertex v; i.e. the probability that vertex v is a terminal endpoint of a new edge is proportional to $A + d(v, t-1)$. These models are nested by the CF model with $p_m = 1,\ q_0 = 1$, and $A_1 = 1/\left(1 + A/2m\right)$. Moreover, Krapivsky et al. (2000) consider a more general setting where the probability that vertex v is a terminal endpoint of a new edge is a function of the degree of this vertex $\Pi(d(v, t-1))$. They show that the network is scale-free if and only if the attachment probability is asymptotically linear, i.e. $\Pi(d(v, t-1)) \sim a_\infty d(v, t-1)$ as $d(v, t-1) \to \infty$.

There are also many copying models, such as Kumar et al. (2000) and Jackson and Rogers (2007). They have the same degree distribution as usual hybrid models and thus these models are also subcases of the CF model. Copying models serve as microfoundations for hybrid models. A simple example of a copying model is as follows. Each period a new vertex is added which selects another vertex w from graph $G(t-1)$ uniformly at random. With probability α the edge (v,w) is added, and with probability $1-\alpha$ the edge (v,w') is added, where w' is a neighbor of vertex w chosen uniformly at random. This last choice is equivalent to preferential attachment, since the probability of given w' to be chosen is proportional to its degree $d(w',t-1)$. Thus this model is nested by the CF model with $p_m=1$, $q_0=1$, and $A_1=1-\alpha$.

3 Degree Distribution

Before turning to the main results presented in Section 4, we analyze degree distribution of the CF model. Section 3.1 approximates the asymptotic degree distribution of the CF model using a mean-field method. First, this method provides the intuition for the reader as to why the asymptotic degree distribution has a power-law tail. Second, this method suggests what parameters of the random multigraph process are crucial for the asymptotic degree distribution. Third, the NLLS method discussed in Section 4.3 is based on the mean-field approximation of the asymptotic degree distribution. Section 3.2 provides formal results on the degree distribution of the CF model. These results are used to establish the main results on asymptotic properties of the GMM estimators in Section 4.2.

3.1 Mean-Field Approximation

The mean-field method introduced by Barabasi and Albert (1999) approximates the network formation process by a continuous time process such that

$$\begin{aligned}\frac{d\mathbb{E}\left(d(v,t)\right)}{dt} &= \frac{\left(\overline{m}A_1+\overline{M}(B_1+C_1)\right)\mathbb{E}\left(d(v,t)\right)}{2\mathbb{E}|E(t-1)|}+\frac{\overline{m}A_2+\overline{M}(B_2+C_2)}{\mathbb{E}|V(t-1)|}\\ &= \frac{\left(\overline{m}A_1+\overline{M}(B_1+C_1)\right)\mathbb{E}\left(d(v,t)\right)}{2(\overline{m}+\overline{M})(t-2)+2|E(1)|}+\frac{\overline{m}A_2+\overline{M}(B_2+C_2)}{t-2+|V(1)|},\end{aligned}$$

where $\overline{m}A_1+\overline{M}(B_1+C_1)$ (respectively $\overline{m}A_2+\overline{M}(B_2+C_2)$) is the expected number of edge endpoints added at time t by preferential attachment (respectively uniformly at random).

Assuming $t\gg\max\{|V(1)|,|E(1)|\}$ the differential equation becomes:

$$\frac{d\mathbb{E}\left(d(v,t)\right)}{dt}=\frac{\eta\mathbb{E}\left(d(v,t)\right)}{t}+\frac{\nu}{t}.$$

The solution to this differential equation is:

$$\phi_t^m(v)\equiv\mathbb{E}\left(d(v,t)\right)=\left(m(v)+\frac{\nu}{\eta}\right)\left(\frac{t}{v}\right)^{\eta}-\frac{\nu}{\eta},$$

where $m(v)$ is the degree of a newly added vertex at time v. The function $\phi_t^m(v)$ is decreasing in v, which means that given an initial degree, "older" vertices have a larger expected degree than "younger" vertices. Thus the distribution of expected degrees of vertices with initial degree m can be approximated by (for $d \geq m$):

$$F_t^m(d) = \frac{p_m|\{i : \phi_t^m(i) \leq d\}|}{p_m t} = 1 - \frac{\phi_t^{m(-1)}(d)}{t} = 1 - \left(m + \frac{\nu}{\eta}\right)^{\frac{1}{\eta}} \left(d + \frac{\nu}{\eta}\right)^{-\frac{1}{\eta}}. \quad (1)$$

Thus the distribution of expected degrees of graph $G(t)$ can be approximated by:

$$F^{MF}(d) = \sum_{m=0}^{\min\{P,d\}} p_m F_t^m(d), \quad (2)$$

where $F_t^m(d)$ are given by (1).

3.2 Formal Results

This section provides formal results on the degree distribution of the CF model introduced in Section 2. Bollobas et al. (2001) obtain formal results for the BA model and Cooper (2006) extend them to the CF model. Following Cooper (2006) we define

$$\begin{aligned} n_m(d;\eta,\nu) &= \frac{B\left(d+\frac{\nu}{\eta}, 1+\frac{1}{\eta}\right)}{B\left(m+\frac{v}{\eta}, \frac{1}{\eta}\right)} = \frac{\Gamma\left(m+\frac{\nu+1}{\eta}\right)}{\eta\Gamma\left(m+\frac{\nu}{\eta}\right)} \frac{\Gamma\left(d+\frac{\nu}{\eta}\right)}{\Gamma\left(d+\frac{\nu+1}{\eta}+1\right)} \\ &= \frac{\left(d+\frac{\nu}{\eta}-1\right)\ldots\left(m+\frac{\nu}{\eta}\right)}{\eta\left(d+\frac{\nu+1}{\eta}\right)\ldots\left(m+\frac{\nu+1}{\eta}\right)} \end{aligned} \quad (3)$$

and

$$d^*(t;\eta) = \min\{t^{\eta/3}, t^{1/6}/\ln^2 t\},$$

where $\Gamma(z)$ is the Gamma function and $B(x,y)$ is the Beta integral:

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt, \quad (4)$$

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}. \quad (5)$$

Note that $n_m(d;\eta,\nu)$ is the Yule-Simon distribution (see Simon (1955)).[7]

Cooper (2006) establishes Lemma 1 that approximates the number of vertices of a given degree with a given initial degree. For this purpose we define $D_t(d,m)$ as the number of

[7] To get the second and third equalities in the definition of $n_m(d;\eta,\nu)$ given by (3) we use the fact $\Gamma(x+1) = x\Gamma(x)$.

vertices of the graph $G(t)$ with initial degree $d(v,v) = m$ and "current" degree $d(v,t) = d$. The approximation results are established for vertices of the graph $G(t)$ with degrees bounded by $d^*(t;\eta)$. This lemma shows that if we look only on vertices of the graph $G(t)$ with initial degree $d(v,v) = m$, then the fraction of these vertices that have degree d is close to $n_m(d;\eta,\nu)$ for large t.

Lemma 1 *For $m \leq d \leq d^*(t;\eta)$ we have the following:*

1. *expected degree sequence*[8]

$$\mathbb{E}D_t(d,m) = p_m n_m(d;\eta,\nu)t\left(1 + O\left(\frac{1}{\ln t}\right)\right),$$

2. *concentration*

$$\Pr\left(|D_t(d,m) - \mathbb{E}D_t(d,m)| \geq \frac{\mathbb{E}D_t(d,m)}{\sqrt{\ln t}}\right) = O\left(\frac{1}{\ln t}\right).$$

Since the initial degree of each vertex is not observed in real networks, we cannot use Lemma 1 directly. Proposition 1 extends Lemma 1 to an unconditional number of vertices of a given degree. The number of vertices of the graph $G(t)$ that have degree d is given by $D_t(d) = \sum_{m=0}^{P} D_t(d,m)$. Let $P(d;\eta,\nu,\boldsymbol{p})$ be a linear combination (with coefficients $\boldsymbol{p}$) of Yule-Simon distributions:

$$P(d;\eta,\nu,\boldsymbol{p}) = \sum_{m=0}^{\min\{P,d\}} p_m n_m(d;\eta,\nu), \tag{6}$$

Proposition 1 shows that the fraction of vertices of the graph $G(t)$ that have degree d is close to $P(d;\eta,\nu,\boldsymbol{p})$ for large t; so $P(d;\eta,\nu,\boldsymbol{p})$ is the asymptotic degree distribution.

Proposition 1 *For $0 \leq d \leq d^*(t;\eta)$ we have the following:*

1. *expected degree sequence*

$$\mathbb{E}D_t(d) = P(d;\eta,\nu,\boldsymbol{p})t\left(1 + O\left(\frac{1}{\ln t}\right)\right),$$

2. *concentration*

$$\Pr\left(|D_t(d) - \mathbb{E}D_t(d)| \geq \frac{\mathbb{E}D_t(d)}{\sqrt{\ln t}}\right) = O\left(\frac{1}{\ln t}\right).$$

[8] Following Cooper (2006), the equations with terms like $O(1/\ln t)$ should be treated as inequalities giving upper and lower bounds (no explicit functional form is implied). Constants in error terms like $O(1/\ln t)$ may depend on parameters of the model but not on d.

Corollary 1 restates expected degree sequence and concentration results of Proposition 1 in a more convenient form, which is used to establish asymptotic properties of the GMM estimators in Section 4.2.

Corollary 1 *For $0 \leq d \leq d^*(t;\eta)$ and for any $M > 1$ we have the following:*

$$\Pr\left(\left|\frac{D_t(d)}{|V(t)|} - P(d;\eta,\nu,\boldsymbol{p})\right| \geq M\frac{P(d;\eta,\nu,\boldsymbol{p})}{\sqrt{\ln t}}\right) = O\left(\frac{1}{\ln t}\right).$$

Claim 1 provides the intuition for the properties of the CF model.

Claim 1 1. *The fraction of vertices of degree d of the graph $G(t)$ converges in probability to $P(d;\eta,\nu,\boldsymbol{p})$ as $t \to \infty$.*

2. *The asymptotic degree distribution has a power-law tail with the power-law parameter $1 + 1/\eta$:*

$$P(d;\eta,\nu,\boldsymbol{p}) = \sum_{m=0}^{\min\{P,d\}} p_m \frac{\Gamma(m+\frac{\nu+1}{\eta})}{\eta\Gamma(m+\frac{\nu}{\eta})} d^{-1-1/\eta}\left(1 + O\left(\frac{1}{d}\right)\right).$$

3. *If the total probability of preferential attachment tends to zero, the asymptotic degree distribution approaches exponential distribution:*

$$\lim_{\eta\to 0} P(d;\eta,\nu,\boldsymbol{p}) = \frac{1}{\nu+1}\left(\frac{\nu}{\nu+1}\right)^{d-m}.$$

4 Methodology

This section introduces the GMM, NLLS, and Hill estimators of the CF model. Section 4.1 discusses what parameters we estimate and what asymptotic limits we use to get formal results on asymptotic properties of estimators. Section 4.2 introduces the GMM estimators, which include the PML estimators as a special case. We prove consistency of the GMM estimators and defer discussion of asymptotic normality to Appendix A. Section 4.3 describes and discusses the NLLS estimator commonly used in the literature. Finally, Section 4.4 describes Hill (1975) estimator.

4.1 Preliminaries

As shown in Corollary 1, the asymptotic degree distribution of the CF model depends only on the subset of parameters of the model; specifically on η, ν, and $\boldsymbol{p} = (p_{0,\ldots,}p_P)$. Section

4.2 shows that these parameters are identified, but this paper does not investigate whether other parameters of the CF model are identified. From the set up of the model it is clear that $\eta \in (0,1)$, $\nu \in [0,\infty)$ and $p \in \Delta^P$, where $\Delta^P = \{\boldsymbol{p} \in \mathbb{R}_+^{P+1} : \sum_{i=0}^P p_i = 1\}$ is P dimensional simplex.[9] We assume that the dimensionality P of $\boldsymbol{p}$ is known; i.e., it is known how many parameters we need to estimate. If P is unknown, then it can be chosen using information criteria such as AIC or BIC (see, e.g., Burnham and Anderson (2002)), but we do not explore asymptotic properties of such procedures.

Parameter η is of the highest interest in this model. First, it determines the power-law parameter $1+1/\eta$, which is important for the statistical properties of the network discussed in the introduction. Second, it is equal to the limiting share of edge endpoints inserted by preferential attachment.

For the rest of this section it is convenient to make the following change of variables: $\xi = 1/\eta \in (1,\infty)$ and $\kappa = \nu/\eta \in [0,\infty)$. Let $\boldsymbol{\theta} = (\xi, \kappa, \boldsymbol{p})$, i.e. $\boldsymbol{\theta}$ is $P+3$ dimensional parameter with the domain $\boldsymbol{\Theta} = (1,\infty) \times [0,\infty) \times \Delta^P$. With some abuse of notation, we write $n_m(d;\xi,\kappa) \equiv n_m\left(d;\eta(\xi,\kappa),\nu(\xi,\kappa)\right) = n_m\left(d;1/\xi,\kappa/\xi\right)$ and $P(d;\boldsymbol{\theta}) = P(d;\eta(\xi,\kappa),\nu(\xi,\kappa),\boldsymbol{p}) = P\left(d;1/\xi,\kappa/\xi,\boldsymbol{p}\right)$. To represent the true value, a generic value, and an estimate, we write $\boldsymbol{\theta}_0$, $\boldsymbol{\theta}$, and $\widehat{\boldsymbol{\theta}}$ respectively.

In Section 4.2 we derive asymptotic properties of estimates $\widehat{\boldsymbol{\theta}}$ as t goes to infinity. This asymptotic is similar to the standard large sample asymptotic, in which the number of observations goes to infinity. In the random multigraph process that we consider, one vertex and at most $P+Q$ edges are added at each time step t. Thus all asymptotic results will continue to hold if we consider an alternative asymptotic in which the number of vertices $|V(t)|$ or the number of edges $|E(t)|$ of the graph $G(t)$ goes to infinity, since $|V(t)| \to \infty$, $|E(t)| \to \infty$, and $t \to \infty$ are equivalent.

4.2 PML and GMM Estimators

We now introduce PML and GMM estimators and present the main results of the paper: consistency of these estimators. The standard regularity conditions for consistency are continuity and uniform convergence. We establish continuity by checking standard technical conditions for the distribution function $P(d;\boldsymbol{\theta})$ given by (6). But we cannot establish uniform convergence by using standard uniform laws of large numbers for independent or stationary data, because the CF model yields nonstandard vertex degree interdependencies. The main tech-

[9] Formally, because of the assumption $\overline{m} + \overline{M} > 0$, whenever $\overline{M} = 0$ we should eliminate point $\boldsymbol{p} = (1,0,...,0)$, which corresponds to $p_0 = 1$, from simplex Δ^P.

nical contribution of the paper is the uniform law of large numbers established for the CF model, using the concentration result of Corollary 1. This law allows us to extend standard consistency results to our network formation model.[10]

Lemma 2 *If $a(d;\boldsymbol{\theta})$ is a matrix of functions continuous in $\boldsymbol{\theta}$ on a compact set $\overline{\boldsymbol{\Theta}} \subset \boldsymbol{\Theta}$, and there is C with $\|a(d;\boldsymbol{\theta})\| < C \cdot (d+1)$ for all $d \in \mathbb{N}$ and all $\boldsymbol{\theta} \in \overline{\boldsymbol{\Theta}}$, where $\|a(d;\boldsymbol{\theta})\| = \left(\sum_{j,k} a_{jk}^2\right)^{1/2}$ is the Euclidean norm, then (i) $G_0(\boldsymbol{\theta}) = \sum_{d=0}^{\infty} a(d;\boldsymbol{\theta}) P(d;\boldsymbol{\theta}_0)$ is continuous in $\boldsymbol{\theta}$ and (ii) $\sup_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}} \left\| \widehat{G}_t(\boldsymbol{\theta}) - G_0(\boldsymbol{\theta}) \right\| \xrightarrow{P} 0$, where $\widehat{G}_t(\boldsymbol{\theta}) = \sum_{d=0}^{\infty} a(d;\boldsymbol{\theta}) \frac{D_t(d)}{|V(t)|}$.*

We now turn to consistency of extremum estimators based on the pseudo log-likelihood:[11]

$$\widehat{L}_t(\boldsymbol{\theta}) = \sum_{d=0}^{\infty} \frac{D_t(d)}{|V(t)|} \ln P(d;\boldsymbol{\theta}) = \sum_{d=0}^{\infty} \frac{D_t(d)}{|V(t)|} \ln \left(\sum_{m=0}^{\min\{P,d\}} p_m \xi \frac{\Gamma(m+\kappa+\xi)}{\Gamma(m+\kappa)} \frac{\Gamma(d+\kappa)}{\Gamma(d+\kappa+\xi+1)} \right). \tag{7}$$

We define the PML estimator as:[12]

$$\widehat{\boldsymbol{\theta}}^1 = \arg\max_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}} \widehat{L}_t(\boldsymbol{\theta}), \tag{8}$$

and the plug-in PML estimator as:

$$\widehat{\boldsymbol{\theta}}^2 = \arg\max_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}} \widehat{L}_t(\boldsymbol{\theta}) \tag{9}$$

$$s.t.\ \kappa = (\xi - 1) \sum_{d=0}^{\infty} d \frac{D_t(d)}{|V(t)|} - \xi \sum_{m=0}^{P} m p_m. \tag{10}$$

The plug-in PML estimator is obtained by replacing $\overline{m} + \overline{M}$ in (7) with an estimate:

$$\widehat{(\overline{m} + \overline{M})} = \frac{1}{2} \sum_{v=1}^{t} \frac{d(v,t)}{|V(t)|} = \frac{1}{2} \sum_{d=0}^{\infty} d \frac{D_t(d)}{|V(t)|} \tag{11}$$

[10] This section closely follows Newey and McFadden (1994) notation. Symbols $\rightsquigarrow$ and $\xrightarrow{P}$ stand for convergence in distribution and probability respectively. $O_P(1)$ and $o_P(1)$ are stochastic order symbols, formally defined in van der Vaart (2000).

[11] The true log-likelihood is different from the pseudo log-likelihood for two reasons: (i) degrees of vertices are interdependent; and (ii) a finite sample (not asymptotic) degree distribution should be used.

[12] Note that the PML estimator which solves (8) is numerically identical to the minimum chi-squared estimator (see, e.g., Harris and Kanji (1983)):

$$\chi_\lambda^2(\boldsymbol{d};\boldsymbol{\theta}) = 2 \sum_{d=0}^{\infty} D_t(d) \ln \left(\frac{D_t(d)}{|V(t)| P(d;\boldsymbol{\theta})} \right) \rightarrow \min_{\boldsymbol{\theta}}.$$

and maximizing (7) over the remaining parameters ξ and $\boldsymbol{p}$; so it is faster to compute the plug-in PML estimator, because it requires maximization over one less parameter (κ). A rationale for this estimator is that $(\widehat{\overline{m}+\overline{M}})$ is a consistent estimate of $\left(\overline{m}+\overline{M}\right) = \left(\kappa + \xi \sum_{m=0}^{P} m p_m\right)/2(\xi - 1)$ with the standard asymptotic:

$$\sqrt{|V(t)|}\left(\widehat{\overline{m}+\overline{M}} - \left(\overline{m}+\overline{M}\right)\right) \rightsquigarrow \mathcal{N}(0, Var(m(t)) + Var(M(t))),$$

because $m(t)$ and $M(t)$ are independent of each other and across time t.

Both the PML and the plug-in PML estimators are consistent.

Proposition 2 *Let $\overline{\boldsymbol{\Theta}} \subset \boldsymbol{\Theta}$ be compact and $\boldsymbol{\theta}_0 \in \overline{\boldsymbol{\Theta}}$. If $\widehat{\boldsymbol{\theta}}$ satisfies $\widehat{L}_t(\widehat{\boldsymbol{\theta}}) \geq \max_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}} \widehat{L}_t(\boldsymbol{\theta}) + o_P(1)$ where $\widehat{L}_t(\boldsymbol{\theta})$ is given by (7), then $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$. In particular, $\widehat{\boldsymbol{\theta}}^1 \xrightarrow{P} \boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}^2 \xrightarrow{P} \boldsymbol{\theta}_0$, where $\widehat{\boldsymbol{\theta}}^1$ and $\widehat{\boldsymbol{\theta}}^2$ are given by (8) and (9).*

We now consider the GMM estimators, which nest the PML estimators if the PML estimators are viewed as solutions to their first-order conditions. In particular, $\widehat{\boldsymbol{\theta}}^1$ is a solution to

$$\sum_{d=0}^{\infty} \nabla_{\boldsymbol{\theta}} \ln P(d;\boldsymbol{\theta}) \frac{D_t(d)}{|V(t)|} = 0,$$

so $\widehat{\boldsymbol{\theta}}^1$ can be viewed as a GMM estimator with a moment function vector given by $\nabla_{\boldsymbol{\theta}} \ln P(d;\boldsymbol{\theta})$.

More generally, let the *moment function* vector $g(d;\boldsymbol{\theta})$ be a vector of functions such that

$$\sum_{d=0}^{\infty} g(d;\boldsymbol{\theta}_0) P(d;\boldsymbol{\theta}_0) = 0 \tag{12}$$

and the dimensionality of $g(d;\boldsymbol{\theta})$ is greater or equal to the dimensionality of $\boldsymbol{\theta}$. Note that (12) requires the expectation of $g(d;\boldsymbol{\theta}_0)$ to be equal to zero with respect to the asymptotic rather than the finite sample distribution. Since (6) gives the explicit expression for the asymptotic distribution $P(d;\boldsymbol{\theta}_0)$, it is straightforward to verify whether a given $g(d;\boldsymbol{\theta})$ satisfies (12). For example, from the discussion of the PML estimators, we know that the *score function* vector $s(d;\boldsymbol{\theta}) \equiv \nabla_{\boldsymbol{\theta}} \ln P(d;\boldsymbol{\theta})$ and the *degree function* $g_d(d;\boldsymbol{\theta}) \equiv d - 2\left(\overline{m}+\overline{M}\right)$ satisfy (12).

The objective function for our GMM estimators is:

$$\widehat{Q}_t(\boldsymbol{\theta}) = -\left[\sum_{d=0}^{\infty} g(d;\boldsymbol{\theta}) \frac{D_t(d)}{|V(t)|}\right]' \widehat{W} \left[\sum_{d=0}^{\infty} g(d;\boldsymbol{\theta}) \frac{D_t(d)}{|V(t)|}\right], \tag{13}$$

where $g(d;\boldsymbol{\theta})$ satisfies (12) and $\widehat{W}$ is a positive semi-definite matrix.

Proposition 3 specifies sufficient conditions on moment function $g(d;\boldsymbol{\theta})$ and matrix $\widehat{W}$ for the GMM estimate $\widehat{\boldsymbol{\theta}}$ to be consistent.

Proposition 3 *Suppose that* $\widehat{W} \xrightarrow{P} W$*, and (i)* W *is positive semi-definite and* $W \sum_{d=0}^{\infty} g(d;\boldsymbol{\theta}) P(d;\boldsymbol{\theta}_0) = 0$ *only if* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$*; (ii)* $\boldsymbol{\theta}_0 \in \overline{\boldsymbol{\Theta}} \subset \boldsymbol{\Theta}$ *where* $\overline{\boldsymbol{\Theta}}$ *is compact; (iii)* $g(d;\boldsymbol{\theta})$ *is continuous on* $\overline{\boldsymbol{\Theta}}$*; and (iv) there is* C *with* $\|g(d;\boldsymbol{\theta})\| < C \cdot (d+1)$ *for all* $d \geq 0$*, and all* $\boldsymbol{\theta} \in \overline{\boldsymbol{\Theta}}$*. If* $\widehat{\boldsymbol{\theta}} \in \arg\max_{\boldsymbol{\theta} \in \overline{\boldsymbol{\Theta}}} \widehat{Q}_t(\boldsymbol{\theta})$*, then* $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}$*.*

Proposition 4 in Appendix A shows that if $P = 0$ then the moment function vector consisting of $g_d(d;\boldsymbol{\theta})$ and $s(d;\boldsymbol{\theta})$ satisfies (12) and all conditions of Proposition 3, except possibly the identification condition (i).[13]

Asymptotic normality of the GMM estimators is suggested by Lemma 3 part 2 in Appendix A, which shows that the dependence of degrees $d(v,t)$ and $d(v',t)$ of most vertices v and v' vanishes in asymptotic; so a central limit theorem may apply. Appendix A presents partial results on the asymptotic distribution of the GMM estimators.

4.3 NLLS Estimator

We now turn to the NLLS method commonly used to estimate scale-free network formation models (Pennock et al. (2002), Jackson and Rogers (2007), and Jackson (2008)). Pennock et al. (2002) assume $m(t) = 0$ and $M(t) = M$ for some M, whereas Jackson (2008) and Jackson and Rogers (2007) assume $m(t) = m$ and $M(t) = 0$ for some m.[14] More generally, the NLLS method can be used only when $m(t) = m$ for some m but $M(t)$ can have an arbitrary distribution $\boldsymbol{q}$. Since most real networks have vertices with zero degree, we use the NLLS method for the case of $m(t) = 0$.

The NLLS method is as follows. First, estimate $\overline{M}$ as $\sum_{v=1}^{t} d(v,t)/2|V(t)|$, which gives a consistent asymptotically normally distributed estimate as shown in (11). Thus, the only parameter to be estimated is ξ. Note that (2) can be rewritten as (taking into account that $\kappa = 2\overline{M}(\xi - 1)$ when $m(t) = 0$):

$$\ln(1 - F^{MF}(d)) = \xi \ln\left(2\overline{M}(\xi - 1)\right) - \xi \ln\left(d + 2\overline{M}(\xi - 1)\right) \tag{14}$$

Second, estimate ξ by finding a fixed point $\widehat{\xi}$ such that regression (14) gives an estimated coefficient of $\ln\left(d + 2\overline{M}(\widehat{\xi} - 1)\right)$ equal to exactly $-\widehat{\xi}$.

There are at least three possible alternatives for what we can use as observations for the NLLS estimator: (i) all observed distinct degrees (Barabasi and Albert (1999)), (ii) all degrees

[13] It is a common problem with GMM estimators that it is difficult to specify primitive conditions on $g(d;\boldsymbol{\theta})$ and W such that the identification condition holds.

[14] To be precise, Jackson and Rogers (2007) consider a directed multigraph process, but this distinction is not crucial, since the CF model and all results can be generalized to the directed multigraph processes.

in the range $[d_{\min}, d_{\max}]$ where $d_{\min} = \min_v d(v,t)$ and $d_{\max} = \max_v d(v,t)$ (Jackson (2008)), and (iii) all observed degrees with repetition (Newman (2005)); the number of observations is respectively (i) $\#\{d : D_t(d) \neq 0\}$, (ii) $d_{\max} - d_{\min} + 1$, and (iii) $|V(t)|$. There are also a few alternatives for what we can use as the dependent variable. Let $\widehat{F}_t(d)$ be the empirical degree distribution of the graph $G(t)$. By definition, $\widehat{F}_t(d_{\max}) = 1$; so $\ln(1 - \widehat{F}_t(d))$ is not defined. As a remedy, we can drop observations with $d = d_{\max}$ and use $\ln(1 - \widehat{F}_t(d))$ as the dependent variable (Jackson (2008)) or we can keep observations with $d = d_{\max}$ and use $\ln(1 - \widehat{F}_t(d) + \gamma/|V(t)|)$ as the dependent variable where $\gamma > 0$. Most of the literature uses $\gamma = 1$, but Gabaix and Ibragimov (2011) propose to use $\gamma = 1/2$. Moreover, we can use the rank as the dependent variable (Gabaix and Ibragimov (2011)); as opposed to the empirical distribution, the rank of vertices with the same degree d changes from $\widehat{F}_t(d-1)$ to $\widehat{F}_t(d)$ in discrete steps of $1/|V(t)|$.

The NLLS procedure is deficient for several reasons. First, this procedure does not guarantee that a fixed point $\widehat{\xi}$ exists or that it is unique. In fact, simulations show that often there is either no or two fixed points.[15] Second, this procedure uses an *approximation* to the asymptotic *expected* degree distribution. Third, even if $F^{MF}(d)$ was the true asymptotic degree distribution, the NLLS estimator, which finds the best fit to $F^{MF}(d)$ on the log-log scale, may still be very biased and inaccurate.

4.4 Hill Estimator

Finally, we describe Hill (1975) estimator that is often used to estimate the power-law parameter. As shown in Claim 1 part 2, the asymptotic degree distribution can be approximated by the limit degree distribution

$$P^{\lim}(d;\xi) = Cd^{-1-\xi} \tag{15}$$

for sufficiently large d. Suppose that we know a number $D \in \mathbb{N}$ after which $P(d;\boldsymbol{\theta})$ is well approximated by $P^{\lim}(d;\xi)$; so we can treat $P^{\lim}(d;\xi)$ as the true asymptotic degree distribution for $d \geq D$. Then conditional on having a degree not lower than D, a vertex chosen uniformly at random has degree d with probability

$$P^{H}(d) = \frac{d^{-1-\xi}}{\zeta(1+\xi, D)}, \tag{16}$$

[15] If there is no fixed point, then as the NLLS estimate, we choose ξ_0 such that $|\xi_0/\xi_1 - 1|$ attains its minimum, where ξ_1 is the estimated coefficient of the regression on $\ln\left(d + 2\overline{M}(\xi_0 - 1)\right)$. If there are multiple fixed points ξ, we choose the root $\widehat{\xi}$ which gives the smallest Kolmogorov-Smirnov statistic, which measures the distance between the actual degree distribution and the fitted approximation of the asymptotic degree distribution.

where $\zeta(1+\xi, D) = \sum_{i=0}^{\infty}(i+D)^{-(1+\xi)}$ is the Hurwitz zeta function. We define $t_{Hill} = |\{v : d(v,t) \geq D\}|$ as the number of vertices that have degree at least D. Thus the pseudo log-likelihood conditional on having degree not lower than D is given by

$$\widehat{L}_t(\xi|D) = -\frac{1}{|V(t)|}\sum_{v:d(v,t)\geq D}\left((1+\xi)\ln d(v,t) + \ln\zeta(1+\xi, D)\right).$$

Therefore, the pseudo maximum likelihood estimate $\widehat{\xi}$ solves

$$s(\widehat{\xi}|D) = -\frac{1}{|V(t)|}\sum_{v:d(v,t)\geq D}\left(\ln d(v,t) + \frac{\zeta'(1+\widehat{\xi}, D)}{\zeta(1+\widehat{\xi}, D)}\right) = 0,$$

$$\frac{\zeta'(1+\widehat{\xi}, D)}{\zeta(1+\widehat{\xi}, D)} = -\frac{1}{t_{Hill}}\sum_{d\geq D} D_t(d)\ln d, \tag{17}$$

where the prime denotes the derivative with respect to the first argument. Note that the variance of this estimate cannot be approximated as $\mathcal{I}^{-1}(\widehat{\xi})$, where $\mathcal{I}(\widehat{\xi})$ is the Fisher information matrix, because vertex degrees are interdependent. Nevertheless many papers incorrectly use $\mathcal{I}^{-1}(\widehat{\xi})$ as an estimate of the variance of $\widehat{\xi}$.

In simulations, we choose D according to AIC or BIC criteria (see, e.g., Burnham and Anderson (2002)). Specifically, we assume that $\Pr(d=k) = \pi_k$ for $k < D$, where π_k are unknown parameters. Then the log-likelihood can be written as

$$\widetilde{L}_t(\pi,\xi) = \sum_{d=0}^{D-1}\frac{D_t(d)}{|V(t)|}\ln\pi_d + \left(1 - \sum_{d=0}^{D-1}\frac{D_t(d)}{|V(t)|}\right)\ln\left(1-\sum_{d=0}^{D-1}\pi_d\right) + L(\xi|D). \tag{18}$$

Handcock and Jones (2004) show that estimates $\widehat{\pi}_d$ and $\widehat{\xi}$ that maximize (18) are as follows: $\widehat{\pi}_d = \frac{D_t(d)}{t}$ for $d < D$ and $\widehat{\xi}$ solves (17). We choose D which minimizes AIC or BIC:

$$AIC = -2\widetilde{L}_t(\widehat{\pi},\widehat{\xi}) + (D+1)\frac{2}{|V(t)|},$$

$$BIC = -2\widetilde{L}_t(\widehat{\pi},\widehat{\xi}) + (D+1)\frac{\ln|V(t)|}{|V(t)|}.$$

5 Simulations

In this section, we use the GMM (including PML), NLLS, and Hill estimators to estimate networks simulated according to the CF model. In Section 5.1, we estimate the CF model for benchmark model parameters. Simulations suggest that the GMM estimators produce much less biased estimates than the NLLS and Hill estimators do. Moreover, the GMM estimates appear to be asymptotically normally distributed, whereas the NLLS and Hill estimates appear to be inconsistent. In Section 5.2, we carry out robustness checks. First, we investigate how the GMM and NLLS estimators perform for networks of different size. The GMM estimators give accurate estimates even for small networks, whereas the NLLS estimators give very biased estimates. Second, we estimate the CF model for different parameter values and show that the above comparison of the GMM and NLLS estimators holds for a wide range of parameter values. Finally, we show that the GMM estimators are not robust to misspecifications, whereas the NLLS estimators are relatively robust. This observation suggests that we can potentially test whether the model is correctly specified using the GMM estimators, but the NLLS estimators always produce the best fit regardless of whether the model is correctly specified.

Since parameters η (the expected share of edges added by preferential attachment) and $\overline{m}+\overline{M}$ (the expected number of edges added at each time step) have more intuitive meaning than ξ and κ, we report estimates and standard errors of η and $\overline{m}+\overline{M}$ in this section. Each reported result is based on 900 simulations of networks using the CF model.

As a benchmark, we consider a subclass of the CF model, which spans all possible values of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. First, we let $M(t)$ to be supported only on M and $M+1$ for some M; i.e., $q_i = 0$ for $i \neq M, M+1$. Second, we impose the same probability of preferential attachment for each type of edge endpoint: $A_1 = B_1 = C_1 = \alpha$. Third, similar to Bollobas et al. (2001), we assume that initial graph $G(1)$ consists of one vertex and a random number $\max\{m(1)+M(1), 1\}$ of loops. Parameters ξ and κ in this subclass are given by:[16]

$$\begin{aligned} \tfrac{1}{\xi} &= \left(1 - \tfrac{\overline{m}}{2(\overline{m}+\overline{M})}\right)\alpha, \\ \kappa &= 2(\overline{m}+\overline{M})\tfrac{1-\alpha}{\alpha}. \end{aligned} \tag{19}$$

5.1 Benchmark Parameters

In this section, we estimate the CF model for the following benchmark parameters: $t = 900$, $\eta = 0.5$, $p_0 = 1$ ($m(t) = 0$), and $q_1 = q_2 = 1/2$ ($\overline{M} = 1.5$). Figure 1 shows degree distributions $\widehat{F}_t(d)$, $F(d) = \sum_{\widetilde{d}=0}^{d} P(\widetilde{d})$, $F_M^{MF}(d)$, and $F^{\lim}(d) = 1 - \sum_{\widetilde{d}>d} P^{\lim}(d)$ on linear and log-log scales.[17] As we can see, the empirical distribution $\widehat{F}_t(d)$ is close to the asymptotic

[16] Although the considered subclass spans all possible values in $\boldsymbol{\Theta}$, the variance of estimates $\widehat{\boldsymbol{\theta}}$ depends on the chosen subclass as discussed in Appendix A. For example, the variance of $\widehat{\overline{m}+\overline{M}}$ depends not only on $\boldsymbol{\theta}$ but also on the distribution $\boldsymbol{q}$ of $M(t)$ (we restrict $q_i = 0$ for $i \neq M, M+1$).

[17] $\widehat{F}_t(d) \equiv \sum_{v=1}^{t} \mathbf{1}(d(v,t) \leq d)/|V(t)|$ is the empirical degree distribution obtained from one simulation of a network. $P(d)$ is the asymptotic degree distribution given by (6). $F^{MF}(d)$ the mean-field approximation of the asymptotic degree distribution given by (2). $P^{\lim}(d)$ is the limit degree distribution given by (15).

distribution $F(d)$ but is far, for small d, from the mean-field approximation $F^{MF}(d)$ and the limit distribution $F^{\lim}(d)$.

<<Place Figure 1 about here>>

For each simulated network, we estimate model parameters using the PML, GMM, NLLS, and Hill estimators. For any parameter β, the statistics reported in all tables are defined as follows: $mean\,(\beta)$ and $std\,(\beta)$ are the sample mean value and standard deviation of an estimate $\widehat{\beta}$; $bias\,(\beta) = (mean\,(\beta) - \beta_0)/\beta_0$ is the sample bias of an estimate $\widehat{\beta}$, measured in percentages; $D^{\beta}_{\max}$ is the Kolmogorov-Smirnov statistic, which is used to test normality of $\widehat{\beta}$.[18]

Table 1 presents the PML and GMM estimation results. For Columns 1-3, we use estimators with $P = 0$, and for Column 4, we use an estimator with $P = 1$. Columns 1 and 2 (PML^0 and PML^0_P) report the PML and plug-in PML estimates given by (8) and (9). Column 3 ($\text{GMM}^0_{-\kappa}$) reports the GMM estimates for the moment function vector $g(d;\boldsymbol{\theta}) = (s_\xi(d;\boldsymbol{\theta}),\ d - 2\left(\overline{m} + \overline{M}\right))'$, where $s_\xi(d;\boldsymbol{\theta})$ is the ξ-component of the score function $s(d;\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \ln P(d;\boldsymbol{\theta})$. Finally, Column 4 ($\text{GMM}^1_{-\kappa}$) reports the GMM estimates for the moment function vector $g(d;\boldsymbol{\theta}) = (s_\xi(d;\boldsymbol{\theta}),\ s_{p_0}(d;\boldsymbol{\theta}),\ d - 2\left(\overline{m} + \overline{M}\right))'$, where $s_{p_0}(d;\boldsymbol{\theta})$ is the p_0-component of $s(d;\boldsymbol{\theta})$. As we can see from Table 1, all estimators are effectively unbiased; so we can compare estimators based on their standard deviations. The $\text{GMM}^1_{-\kappa}$ estimator, which estimates more parameters than required ($P = 1$ instead of $P = 0$), is not worse than the PML and GMM estimators that estimate the correct number of parameters. Moreover, the plug-in PML estimator is better than the PML estimator, and both are worse than the other GMM estimators.

<<Place Table 1 about here>>

Table 2 presents the NLLS and Hill estimation results. Columns 1 and 2 (NLLS and NLLS_R) report the NLLS estimates with observations being all observed distinct degrees and all degrees in the range $[d_{\min}, d_{\max}]$. In both cases, we use $\ln(1 - \widehat{F}_t(d))$ as the dependent variable and drop observations with $d = d_{\max}$. Columns 3 and 4 (Hill_{AIC} and Hill_{BIC}) report Hill estimates with the threshold degree D chosen according to AIC and BIC criteria. As we can see from Tables 1 and 2, all PML and GMM estimators are much less biased than the

[18] Formally, $D^{\beta}_{\max} = \max|\widehat{F}(x) - F^*(x)|$, where $\widehat{F}(x)$ is the empirical distribution of $\widehat{\beta}$ and $F^*(x)$ is the cumulative normal distribution whose mean and variance are estimated from the sample.

NLLS estimators, which are in turn are less biased than Hill estimators.[19] Moreover, the NLLS and Hill estimators appear to be inconsistent; the bias does not decrease considerably as we increase the size of network t. The NLLS estimator is less biased than the NLLS$_R$ estimator; the other versions of the NLLS estimator discussed in Section 4.3 are also inconsistent and do not perform better. We get the same qualitative comparisons of all estimators for other parameters of the CF model.

<<Place Table 2 about here>>

Back-of-the-envelope calculations suggest asymptotic normality of the PML and GMM estimators, except the GMM$^1_{-\kappa}$ estimator, which has the boundary parameter value $p_0 = 1$. For the normal distribution with the sample size 900 (our number of simulations) upper 10% and 1% quantiles for $D_{\max}$ are 0.028 and 0.035 respectively (see Lilliefors (1967) and Dallal and Wilkinson (1986)). Tables 1 shows that $D^{\eta}_{\max}$ is smaller than 0.028 for the PML and GMM estimators. Thus, based on 900 simulations, the hypothesis that the PML and GMM estimates of η are normally distributed cannot be rejected even on 10% significance level.

5.2 Robustness Checks

Table 3 explores how statistical properties of estimates depend on the number of nodes in a network. We report estimation results for the GMM$^0_{-\kappa}$ and NLLS estimators when t is equal to 25, 100, and 400 (results for $t = 900$ are presented in Tables 1 and 2). As we can see, the GMM$^0_{-\kappa}$ estimates are considerably less biased than the NLLS estimates. The GMM$^0_{-\kappa}$ estimate of η has a small bias even for small networks, e.g., the bias of $\widehat{\eta}$ is just 6% when the number of vertices of the network is 25. Moreover, the standard deviation of the GMM$^0_{-\kappa}$ estimate of η appears to decrease with the rate $\sqrt{t}$, which suggests that the GMM estimates of η are $\sqrt{t}$-consistent.

<<Place Table 3 about here>>

Table 4 explores how results change with parameters. Columns 1 through 4 present results for different values of η keeping all other parameters fixed at the benchmark level. Again,

[19] The Hill estimator with D chosen by AIC criterion outperforms the Hill estimator with D chosen by BIC criterion. It might be possible to improve Hill estimator by changing the choice of D, but we do not pursue this goal in the paper. Simulations suggest that as we increase D the bias of $\widehat{\eta}$ decreases but the standard deviation of $\widehat{\eta}$ decreases. Since AIC has a smaller "penalty term" for a number of parameters, it chooses higher D and thus it gives a less biased estimate of $\widehat{\eta}$.

the $\mathrm{GMM}^0_{-\kappa}$ estimator is much less biased than the NLLS estimator. Columns 5, 6, and 7 present results for the benchmark parameters with one modification: $\boldsymbol{p} = (0.5, 0.5)$; i.e., $m(t)$ is either 1 or 0 with equal probabilities. In this case, the NLLS and $\mathrm{GMM}^0_{-\kappa}$ estimators are misspecified, but the $\mathrm{GMM}^1_{-\kappa}$ estimator is correctly specified. The (misspecified) $\mathrm{GMM}^0_{-\kappa}$ estimate of η has a larger bias than the NLLS estimate of η has. This observation suggests that the GMM estimators are not robust to misspecification. The PML and other GMM estimators are not robust to misspecification too. When we use the PML and GMM estimators, misspecifications can be visually detected on the log-log plot of empirical and fitted asymptotic distributions, because empirical and fitted asymptotic distributions will have different asymptotic power-law decays. But the NLLS estimators are meant to produce the best possible visual fit between the mean-field approximated and empirical distributions on the log-log scale, regardless of whether the model is correctly specified or not. The (correctly specified) $\mathrm{GMM}^1_{-\kappa}$ estimator gives an accurate estimate of η and $p_0 = \Pr(m(\tau) = 0)$ with small biases and standard deviations.

<<Place Table 4 about here>>

Table 5 presents robustness checks to different choices of subclasses of the CF model for the benchmark parameters. Columns 1 and 2 present results for the case when the initial graph $G(1)$ is a fully connected graph with $2\left\lfloor \overline{m} + \overline{M} \right\rfloor + 1$ vertices instead of a vertex with $\max\{m(1) + M(1), 1\}$ loops. Columns 3 and 4 present results for the case when $q_0 = q_3 = 1/2$ ($M(t)$ is either 0 or 3 with equal probabilities) instead of $q_1 = q_2 = 1/2$ ($M(t)$ is either 1 or 2 with equal probabilities). Finally, columns 5 and 6 present results for the case when $B_1 = 0$ and $C_1 = 1$ instead of $B_1 = C_1 = 1/2$; i.e., the initial vertex of each edge is chosen uniformly at random and the terminal vertex of each edge is chosen by preferential attachment. Table 4 shows that in all of these variations, the $\mathrm{GMM}^0_{-\kappa}$ estimator is much less biased than the NLLS estimator. Moreover, biases and standard deviations of the $\mathrm{GMM}^0_{-\kappa}$ estimator for all these variations have similar magnitudes.

<<Place Table 5 about here>>

6 Conclusion

We estimate a general model of scale-free network formation using the GMM, NLLS, and Hill estimators. We prove consistency of the GMM estimators. The GMM estimators produce considerably better estimates than the NLLS and Hill estimators if the model is correctly specified. However, the GMM estimators are less robust to misspecification compared to the NLLS estimator. This observation implies that we can potentially test whether the model is correctly specified using the GMM estimators, but the NLLS estimators produce the best fit, regardless of whether the model is correctly specified.

It is a pioneering paper, which is meant to show that rigorous econometric theory can be and should be used to estimate parameters of a network formation process. Estimating such processes allows us to estimate economically relevant parameters, such as the power-law parameter of the degree distribution.

The results of this paper are useful for a new and growing literature on estimation of networks. For example, Atalay et al. (2011) and Atalay (2013) use the PML estimator to estimate similar network formation models. Using the methodology developed in this paper, one can prove consistency of the PML estimator for these models. Moreover, our simulations suggest that some GMM estimators yield two times smaller standard errors of estimates than the PML estimator does; so one can estimate model parameters more precisely using our general class of GMM estimators, which includes the PML estimator as a special case.

We conjecture that the GMM estimates are asymptotically normally distributed. This conjecture is supported by simulations and an informal argument. We hope that further research will formally establish asymptotic normality and will show how the asymptotic variance of the GMM estimates depends on model parameters.

Appendix A: Asymptotic Distribution

For the expositional purpose we work with the case $m(t) = 0$; i.e., $\boldsymbol{\theta} = (\xi, \kappa)$. Generalization to the case $\boldsymbol{p} \in \Delta^P$ is straightforward but technically difficult. Since the focus is on asymptotic distribution, we can assume that $|V(1)| = 1$. Proposition 4 gives sufficient conditions for asymptotic normality of the GMM estimators.

Proposition 4 *Let* $\overline{\boldsymbol{\Theta}} \subset \boldsymbol{\Theta}$ *be compact such that* $\boldsymbol{\theta}_0 \in interior(\overline{\boldsymbol{\Theta}})$. *Suppose that* $\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \overline{\boldsymbol{\Theta}}} \widehat{Q}_t(\boldsymbol{\theta})$, *where* $\widehat{Q}_t(\boldsymbol{\theta})$ *is given by (13),* $\widehat{W} \xrightarrow{P} W$, *and the moment function vector is* $g(d; \boldsymbol{\theta}) = \left(\nabla_{\boldsymbol{\theta}} \ln P(d; \boldsymbol{\theta}),\ d - \frac{\kappa}{\xi - 1}\right)'$. *If (i)* W *is positive semi-definite and* $W \sum_{d=0}^{\infty} g(d; \boldsymbol{\theta}) P(d; \boldsymbol{\theta}_0) = 0$ *only if* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, *(ii)* $G'WG$ *is nonsingular for* $G = \sum_{d=0}^{\infty} \nabla_{\boldsymbol{\theta}} g(d; \boldsymbol{\theta}_0) P(d; \boldsymbol{\theta}_0)$, *and (iii)* $\sqrt{t} \sum_{d=0}^{\infty} g(d; \boldsymbol{\theta}_0) \frac{D_t(d)}{t} \rightsquigarrow N(0, \Sigma)$, *then* $\sqrt{t}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \rightsquigarrow N\left[0, (G'WG)^{-1} G'W\Sigma WG\,(G'WG)^{-1}\right]$.

Condition (i) is an identification condition required for consistency: $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$. Note that the GMM estimators $\widehat{\boldsymbol{\theta}}^1$ and $\widehat{\boldsymbol{\theta}}^2$ given by (8) and (9) are consistent by Proposition 2, so condition (i) can be dropped for them. Condition (ii) is a strict local identification condition. At the end of the proof of Proposition 4 we derive G; so for given $\boldsymbol{\theta}_0$ and W, we can easily verify condition (ii) by showing that $\det(G'WG) \neq 0$.

Condition (iii) is an asymptotic normality condition for a sample average of $g(d;\boldsymbol{\theta}_0)$, which usually follows from a central limit theorem. Simulations suggest that this condition holds, but we cannot prove it formally because vertex degrees have nonstandard interdependencies and standard central limit theorems do not apply. Informally, asymptotic normality is suggested by Lemma 3 of Cooper (2006), which shows that most vertex degrees are asymptotically independently distributed according to a negative binomial distribution.

Lemma 3 *Let* $d^*(v,t;\eta) = m + \min\{t^\eta/\ln^4 t, v^{1/2}/\ln^4 t, t^{1/2}(v/t)^\eta/\ln^4 t\}$, *$K$ be any positive constant, and*

$$\pi_{d,m}(v,t;\eta,\nu) = \left(1 + O\left(\frac{1}{\ln^2 t}\right)\right)\binom{d+\frac{\nu}{\eta}-1}{d-m}\left(\frac{v}{t}\right)^{\eta\left(m+\frac{\nu}{\eta}\right)}\left(1-\left(\frac{v}{t}\right)^{\eta}\left(1+O\left(\frac{1}{\ln^2 t}\right)\right)\right)^{d-m},$$

where $\binom{x+y}{x}$ *is the generalized binomial coefficient:*

$$\binom{x+y}{x} = \frac{\Gamma(x+y+1)}{\Gamma(x+1)\Gamma(y+1)}.$$

1. *For* $\ln^8 t \le v \le (t - t/\ln t)$ *and* $m \le d \le d^*(v,t;\eta)$ *or for* $(t-t/\ln t) < v < t$ *and* $d = m, m+1$.

$$\Pr(d(v,t) = d | d(v,v) = m) = \pi_{d,m}(v,t;\eta,\nu) + O(v^{-K}).$$

2. *For* $\ln^8 t \le v < v' < (t - t/\ln t)$ *and* $d \le d^*(v,t;\eta)$, $d' \le d^*(v',t;\eta)$,

$$\Pr(d(v,t) = d, d(v',t) = d' | d(v,v) = m, d(v',v') = m') = \pi_{d,m}(v,t;\eta,\nu)\pi_{d',m'}(v',t;\eta,\nu) + O(v^{-K}).$$

It is straightforward to verify condition (v) and find Σ if we assume that asymptotic results given by Lemma 3 hold in a finite sample:

Condition 1 *All vertex degrees* $d(1,t),\ldots,d(v,t),\ldots,d(t,t)$ *are independently distributed.*

Condition 2 *Each vertex* $v \in \{1,..,t\}$ *has the negative binomial distribution with parameters* κ *and* $(v/t)^{1/\xi}$ *(NegBin($\kappa,(v/t)^{1/\xi}$)), i.e.* $P(d(v,t) = d) = f\left(d;\kappa,(v/t)^{1/\xi}\right)$, *where* $f(.)$ *is:*

$$f(l;r,p) = \binom{l+r-1}{l} p^r (1-p)^l. \tag{20}$$

Simulations suggest that Condition 2 is valid in that it does not affect the asymptotic variance Σ, but Condition 1 is not valid. To illustrate this point, we derive the pseudo asymptotic distribution of $\sqrt{t}\left(\sum_{d=0}^{\infty} d\frac{D_t(d)}{t} - \frac{\kappa}{\xi-1}\right)$ assuming Conditions 1 and 2 and then compare it with the correct asymptotic distribution $\sqrt{t}\left(\sum_{d=0}^{\infty} d\frac{D_t(d)}{t} - \frac{\kappa}{\xi-1}\right) \rightsquigarrow N(0, 4Var(M(t)))$, which follows from independence of $M(t)$ across t.

Claim 2 *Suppose Conditions 1 and 2 are satisfied. Then*

1. *if* $\xi > 2$
$$\sqrt{t}\left(\sum_{d=0}^{\infty} d\frac{D_t(d)}{t} - \frac{\kappa}{\xi-1}\right) \rightsquigarrow N\left(0, \frac{\kappa\xi}{(\xi-1)(\xi-2)}\right), \tag{21}$$

2. *if* $\xi \le 2$
$$Var\left(\sqrt{t}\sum_{d=0}^{\infty} d\frac{D_t(d)}{t}\right) \to \infty. \tag{22}$$

Claim 2 implies that the pseudo asymptotic variance of the average degree is finite if and only if $\xi > 2$ and depends only on $\boldsymbol{\theta}$, but the true asymptotic variance $4Var(M(t))$ is always finite and depends not only on $\boldsymbol{\theta}$ but also on $\boldsymbol{q}$. Furthermore, the true and pseudo asymptotic variances cannot be compared. Indeed, let $q_0 = 1 - \overline{M}/Q$ and $q_Q = \overline{M}/Q$ for some integer $Q \ge \overline{M} = \kappa/(\xi-1)$, so that $q_Q Q = \overline{M}$. Then $Var(M(t)) = \overline{M}\left(Q - \overline{M}\right)$ depends on $\boldsymbol{\theta}$ and $\boldsymbol{q}$ and can be made arbitrary large as we increase Q but keep $\boldsymbol{\theta}$ constant. This brings us to an important point that the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$ depends not only on $\boldsymbol{\theta}$ but also on other parameters of the model even though the asymptotic vertex degree distribution depends only on $\boldsymbol{\theta}$ (Lemma 3 part 1). Therefore, to analyze how the true asymptotic variance of $\widehat{\boldsymbol{\theta}}$ depends on parameters of the model, we need to refine results on the interdependence of vertex degrees (extension of Lemma 3 part 2), which is beyond the scope of this paper.[20]

To get a consistent estimate of the asymptotic variance of $\widehat{\boldsymbol{\theta}}$, we need to find consistent estimates of G and Σ.[21] $\left(G'WG\right)^{-1} G'W\Sigma WG \left(G'WG\right)^{-1}$A consistent estimate of G can be obtained by $\widehat{G} = \sum_{d=0}^{\infty} g(d;\widehat{\boldsymbol{\theta}})\frac{D_t(d)}{t}$,[22] but it is difficult to get a consistent estimate of Σ,

[20]All points made in this paragraph remain valid, if instead of Conditions 1 and 2 we follow the literature and incorrectly assume that each vertex degree is i.i.d. according to $n_0\left(d;\xi,\kappa\right)$. In this case, the pseudo asymptotic variance is $\frac{\kappa\xi(\xi+\kappa-1)}{(\xi-2)(\xi-1)^2}$ if $\xi > 2$ and infinity otherwise.

[21]By hypothesis of Proposition 4 $\widehat{W} \xrightarrow{P} W$ and $G'WG$ is nonsingular. If in addition $\widehat{G} \xrightarrow{P} G$ and $\widehat{\Sigma} \xrightarrow{P} \Sigma$, then by continuous mapping theorem, $\left(\widehat{G}'\widehat{W}\widehat{G}\right)^{-1} \widehat{G}'\widehat{W}\widehat{\Sigma}\widehat{W}\widehat{G}\left(\widehat{G}'\widehat{W}\widehat{G}\right)^{-1} \to \left(G'WG\right)^{-1} G'W\Sigma WG \left(G'WG\right)^{-1}$.

[22]Consistency, continuity, and uniform convergence imply: $\left\|\widehat{G} - G\right\| \le \left\|\widehat{G} - G(\widehat{\boldsymbol{\theta}})\right\| + \left\|G(\widehat{\boldsymbol{\theta}}) - G\right\| \le \sup_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}} \left\|\sum_{d=0}^{\infty} g(d;\widehat{\boldsymbol{\theta}})\frac{D_t(d)}{t} - G(\boldsymbol{\theta})\right\| + \left\|G(\widehat{\boldsymbol{\theta}}) - G\right\| \xrightarrow{P} 0$.

because vertex degrees are interdependent and not identically distributed. Indeed, Σ is given by

$$\Sigma = \lim_{t\to\infty} \frac{1}{t} Var\left(\sum_{v=1}^{t} g(d(v,t);\boldsymbol{\theta}_0)\right) = \Sigma_1 - \Sigma_2 + \Sigma_3,$$

where

$$\Sigma_1 = \lim_{t\to\infty} \frac{1}{t} \sum_{v=1}^{t} \mathbb{E}\left[g(d(v,t);\boldsymbol{\theta}_0) g(d(v,t);\boldsymbol{\theta}_0)'\right],$$

$$\Sigma_2 = \lim_{t\to\infty} \frac{1}{t} \sum_{v=1}^{t} \mathbb{E}\left(g(d(v,t);\boldsymbol{\theta}_0)\right) \mathbb{E}\left(g(d(v,t);\boldsymbol{\theta}_0)\right)',$$

$$\Sigma_3 = \lim_{t\to\infty} \frac{1}{t} \sum_{v,v':v\neq v'} \mathbb{E}\left[\left(g(d(v,t);\boldsymbol{\theta}_0) - \mathbb{E}g(d(v,t);\boldsymbol{\theta}_0)\right) g\left((d(v',t);\boldsymbol{\theta}_0) - \mathbb{E}g(d(v',t);\boldsymbol{\theta}_0)\right)'\right].$$

It is easy to construct a consistent estimate of Σ_1:

$$\widehat{\Sigma}_1 = \sum_{v=1}^{t} g(d;\widehat{\boldsymbol{\theta}}) g(d;\widehat{\boldsymbol{\theta}})' \frac{D_t(d)}{t}.$$

We can get an estimate of Σ_2 if we approximate the distribution of $d(v,t)$ by its asymptotic distribution (Condition 2). Simulations suggest that this estimate of Σ_2 is consistent. However, we cannot get a consistent estimate of Σ_3 (Claim 2 shows that $\Sigma_3 \neq 0$; i.e. we cannot use Condition 1). Σ_2 is positive semi definite as the sum of positive semi definite matrices, thus $\widehat{\Sigma}_1$ can serve as a conservative estimate of $\Sigma_1 - \Sigma_2$. Moreover, if Σ_3 is negative semi definite or relatively small, we can use $\widehat{\Sigma}_1$ as a conservative estimate of Σ.

Theorem 5.2 of Newey and McFadden (1994) implies that the GMM estimator with $\widehat{W} \xrightarrow{P} \Sigma^{-1}$ is asymptotically efficient in the class of GMM estimators with the moment function $g(d;\boldsymbol{\theta}) = \left(\nabla_{\boldsymbol{\theta}} \ln P(d;\boldsymbol{\theta}),\ d - \frac{\kappa}{\xi-1}\right)'$. However, to be able to apply this result, we need to get a consistent estimate of Σ. More generally, one might want to find an asymptotically efficient estimator in the class of all GMM estimators, which allows other moment functions. This problem seems to be unsolvable without further results on the structure of vertex degree interdependencies. Moreover, it is not clear why we should look for an efficient estimator in the class of GMM estimators. If Conditions 1 and 2 were valid, we could write the log-likelihood as

$$\frac{1}{t} \log\left(\sum_{\widetilde{\boldsymbol{d}}\in\mathcal{P}(\boldsymbol{d})} \prod_{v=1}^{t} f\left(\widetilde{d}(v,t);\kappa,(v/t)^{1/\xi}\right)\right), \tag{23}$$

where $\mathcal{P}(\boldsymbol{d})$ denotes the set of all distinct permutations of $\boldsymbol{d} = (d(1,t),...,d(v,t))$. The estimator that maximizes (23) does not belong to the class of GMM estimators. This argument suggests that the estimator that maximizes the true likelihood, which takes into account interdependencies of vertex degrees, does not belong to the class of GMM estimators either. Even if we could set up the true likelihood or its asymptotic limit, it would seem to be computationally infeasible to calculate it for large networks because we would have to sum up probabilities over all possible permutations, which number grows very fast (the number of permutations is given by $t!/\prod D_t(d)$).

Appendix B: Proofs

Proof of Lemma 1. See the proof of Theorem 2.1 in Cooper (2006). ∎

Proof of Proposition 1. Summing up expressions from part 1 of Lemma 1 gives part 1 of this proposition. The following sequence of inequalities establishes part 2:

$$\begin{aligned}
\Pr\left(|D_t(d)-\mathbb{E}D_t(d)|\geq \tfrac{\mathbb{E}D_t(d)}{\sqrt{\ln t}}\right) &= \Pr\left(\left|\textstyle\sum_{m=0}^{\min\{P,d\}}\left(D_t(d,m)-\mathbb{E}D_t(d,m)\right)\right|\geq \tfrac{\sum_{m=0}^{\min\{P,d\}}\mathbb{E}D_t(d,m)}{\sqrt{\ln t}}\right)\\
&\leq \Pr\left(\textstyle\sum_{m=0}^{\min\{P,d\}}|D_t(d,m)-\mathbb{E}D_t(d,m)|\geq \tfrac{\sum_{m=0}^{\min\{P,d\}}\mathbb{E}D_t(d,m)}{\sqrt{\ln t}}\right)\\
&\leq \Pr\left(\exists m: |D_t(d,m)-\mathbb{E}D_t(d,m)|\geq \tfrac{\mathbb{E}D_t(d,m)}{\sqrt{\ln t}}\right)\\
&\leq \textstyle\sum_{m=0}^{\min\{P,d\}}\Pr\left(|D_t(d,m)-\mathbb{E}D_t(d,m)|\geq \tfrac{\mathbb{E}D_t(d,m)}{\sqrt{\ln t}}\right)=O\left(\tfrac{1}{\ln t}\right).
\end{aligned}$$

∎

Proof of Corollary 1. The following sequence of equalities and inequalities proves the corollary:

$$\begin{aligned}
&\Pr\left(|D_t(d)-\mathbb{E}D_t(d)|\geq \tfrac{\mathbb{E}D_t(d)}{\sqrt{\ln t}}\right)\\
&=\Pr\left(\left|\tfrac{D_t(d)}{|V(t)|}-P(d;\eta,\nu,\boldsymbol{p})\tfrac{t}{t+|V(1)|-1}\left(1+O\left(\tfrac{1}{\ln t}\right)\right)\right|\geq \tfrac{P(d;\eta,\nu,\boldsymbol{p})}{\sqrt{\ln t}}\tfrac{t}{|V(t)|}\left(1+O\left(\tfrac{1}{\ln t}\right)\right)\right)\\
&=\Pr\left(\left|\tfrac{D_t(d)}{|V(t)|}-P(d;\eta,\nu,\boldsymbol{p})\right|\geq \tfrac{P(d;\eta,\nu,\boldsymbol{p})}{\sqrt{\ln t}}\left(1+O\left(\tfrac{1}{\ln t}\right)\right)\right)\\
&\geq \Pr\left(\left|\tfrac{D_t(d)}{|V(t)|}-P(d;\eta,\nu,\boldsymbol{p})\right|\geq M\tfrac{P(d;\eta,\nu,\boldsymbol{p})}{\sqrt{\ln t}}\right),
\end{aligned}$$

for any $M>1$ and sufficiently large t. ∎

Proof of Claim 1. To establish part 1 we note that $d^*(t;\eta)\to\infty$ as $t\to\infty$, thus $d\leq d^*(t;\eta)$ and Corollary 1 applies.

Part 2 follows from the well-known result $\frac{\Gamma(z+\alpha)}{\Gamma(z+\beta)}=z^{\alpha-\beta}(1+O\left(\frac{1}{z}\right))$ applied to $\frac{\Gamma\left(d+\frac{\nu}{\eta}\right)}{\Gamma\left(d+\frac{\nu+1}{\eta}+1\right)}$ (see, e.g., Palumbo (1998)).

Part 3 follows from the representation

$$n_m(d;\eta,\nu)=\frac{\left(d+\frac{\nu}{\eta}-1\right)\ldots\left(m+\frac{\nu}{\eta}\right)}{\eta\left(d+\frac{\nu+1}{\eta}\right)\ldots\left(m+\frac{\nu+1}{\eta}\right)}\underset{\eta\to 0}{\to}\frac{\left(\frac{\nu}{\eta}\right)^{d-m}}{\eta\left(\frac{\nu+1}{\eta}\right)^{d-m+1}}=\frac{1}{\nu+1}\left(\frac{\nu}{\nu+1}\right)^{d-m}.$$

∎

Proof of Lemma 2. Part (i). We first prove that $G_0^n(\boldsymbol{\theta}) = \sum_{d=0}^{n} a(d;\boldsymbol{\theta})P(d;\boldsymbol{\theta}_0)$ converges uniformly on $\overline{\boldsymbol{\Theta}}$ to $G_0(\boldsymbol{\theta})$. Palumbo (1998) shows that

$$\tfrac{\Gamma(k+\lambda)}{\Gamma(k+1)} > \quad (k+1)^{\lambda-1} \quad \text{for } \lambda > 2 \text{ and } k \geq 0.$$

Thus for $d \geq 1-\kappa_0$

$$P(d;\boldsymbol{\theta}_0) = C(\boldsymbol{\theta}_0)\tfrac{\Gamma(d+\kappa_0)}{\Gamma(d+\kappa_0+\xi_0+1)} < C(\boldsymbol{\theta}_0)\,(d+\kappa_0)^{-1-\xi_0}\,, \tag{24}$$

where $C(\boldsymbol{\theta}_0) = \displaystyle\sum_{m=0}^{P} p_{m0}\tfrac{\xi_0\Gamma(m+\kappa_0+\xi_0)}{\Gamma(m+\kappa_0)}$. Thus

$$\|a(d;\boldsymbol{\theta})P(d;\boldsymbol{\theta}_0)\| < C(\boldsymbol{\theta}_0)\,(d+\kappa_0)^{-1-\xi_0}\,C(d+1) = J_d,$$

for some C. Evidently, $\sum_{d=0}^{\infty} J_d < \infty$. Thus $G_0^n(\boldsymbol{\theta})$ converges uniformly on $\overline{\boldsymbol{\Theta}}$ to $G_0(\boldsymbol{\theta})$ (Theorem 7.10 of Rudin (1976)). Moreover, since $G_0^n(\boldsymbol{\theta})$ is continuous on $\overline{\boldsymbol{\Theta}}$, $G_0(\boldsymbol{\theta})$ is also continuous on $\overline{\boldsymbol{\Theta}}$ (Theorem 7.12 of Rudin (1976)).

Part (ii). Clearly, there exists $\widetilde{d}(t)$ such that $\widetilde{d}(t) \leq d^*(t;\eta)$, $\widetilde{d}(t) \to \infty$, and $\widetilde{d}(t)/\ln t \to 0$ as $t \to \infty$. Then,

$$\begin{aligned}\left\|\widehat{G}_t(\boldsymbol{\theta}) - G_0(\boldsymbol{\theta})\right\| &= \left\|\sum_{d=0}^{\widetilde{d}(t)-1} a(d;\boldsymbol{\theta})\left(\tfrac{D_t(d)}{|V(t)|} - P(d;\boldsymbol{\theta}_0)\right) + \sum_{d=\widetilde{d}(t)}^{\infty} a(d;\boldsymbol{\theta})\left(\tfrac{D_t(d)}{|V(t)|} - P(d;\boldsymbol{\theta}_0)\right)\right\| \\ &\leq \underbrace{\left\|\sum_{d=\widetilde{d}(t)}^{\infty} a(d;\boldsymbol{\theta})P(d;\boldsymbol{\theta}_0)\right\|}_{\widehat{S}_1(\boldsymbol{\theta})} + \underbrace{\left\|\sum_{d=0}^{\widetilde{d}(t)-1} a(d;\boldsymbol{\theta})\left|\frac{D_t(d)}{|V(t)|} - P(d;\boldsymbol{\theta}_0)\right|\right\|}_{\widehat{S}_2(\boldsymbol{\theta})} + \underbrace{\left\|\sum_{d=\widetilde{d}(t)}^{\infty} a(d;\boldsymbol{\theta})\frac{D_t(d)}{|V(t)|}\right\|}_{\widehat{S}_3(\boldsymbol{\theta})}.\end{aligned}$$

To prove $\sup_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}}\left\|\widehat{G}_t(\boldsymbol{\theta}) - G_0(\boldsymbol{\theta})\right\| \xrightarrow{P} 0$, it suffices to show that $\sup_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}}\widehat{S}_1(\boldsymbol{\theta}) \xrightarrow{P} 0$, $\sup_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}}\widehat{S}_2(\boldsymbol{\theta}) \xrightarrow{P} 0$, and $\sup_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}}\widehat{S}_3(\boldsymbol{\theta}) \xrightarrow{P} 0$.

1. Because $G_0^n(\boldsymbol{\theta})$ uniformly converges to $G_0(\boldsymbol{\theta})$ on $\overline{\boldsymbol{\Theta}}$, we have $\sup_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}}\widehat{S}_1(\boldsymbol{\theta}) \xrightarrow{P} 0$.

2. Corollary 1 implies that for any $M > 1$ there exists $N(\boldsymbol{\theta}_0)$ such that for $0 \leq d \leq d^*(t;\eta)$, we have:

   $$\Pr\left(\left|\tfrac{D_t(d)}{|V(t)|} - P(d;\boldsymbol{\theta}_0)\right| \geq M\tfrac{P(d;\boldsymbol{\theta}_0)}{\sqrt{\ln t}}\right) \leq \tfrac{N(\boldsymbol{\theta}_0)}{\ln t}.$$

   Therefore, by definition of $\widetilde{d}(t)$, we have

   $$\Pr\left(\exists d \leq \widetilde{d}(t) : \left|\tfrac{D_t(d)}{|V(t)|} - P(d;\boldsymbol{\theta}_0)\right| \geq M\tfrac{P(d;\boldsymbol{\theta}_0)}{\sqrt{\ln t}}\right) \leq \tfrac{N(\boldsymbol{\theta}_0)\widetilde{d}(t)}{\ln t} \to 0. \tag{25}$$

   Thus, with probability approaching one:

   $$\widehat{S}_2(\boldsymbol{\theta}) \leq \left\|\sum_{d=0}^{\widetilde{d}(t)-1} a(d;\boldsymbol{\theta})M\tfrac{P(d;\boldsymbol{\theta}_0)}{\sqrt{\ln t}}\right\| < C_1\tfrac{M\|G_0(\boldsymbol{\theta})\|}{\sqrt{\ln t}},$$

for some C_1. The last inequality follows from the uniform convergence of $G_0^n(\boldsymbol{\theta})$ on $\overline{\boldsymbol{\Theta}}$. Since $G_0(\boldsymbol{\theta})$ is continuous on a compact set $\overline{\boldsymbol{\Theta}}$, $\|G_0(\boldsymbol{\theta})\|$ is bounded on $\overline{\boldsymbol{\Theta}}$; so $\sup_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}} \widehat{S}_2(\boldsymbol{\theta}) \xrightarrow{P} 0$.

3. Since $\|a(d;\boldsymbol{\theta})\| < C(d+1)$, showing $\sum_{d=\widetilde{d}(t)}^{\infty} d\frac{D_t(d)}{|V(t)|} \xrightarrow{P} 0$ is sufficient for $\sup_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}} \widehat{S}_3(\boldsymbol{\theta}) \xrightarrow{P} 0$. By definition of $n_m(d;\eta,\nu)$,

$$\textstyle\sum_{d=m}^{n} dn_m(d;\eta,\nu) = \frac{\kappa+\xi m}{\xi-1} - n\frac{B(n+\kappa+1,\xi)}{B(m+\kappa,\xi)} - \frac{B(n+\kappa+1,\xi-1)}{B(m+\kappa,\xi)} \underset{n\to\infty}{\longrightarrow} \frac{\kappa+\xi m}{\xi-1},$$

which gives

$$\textstyle\sum_{d=0}^{\infty} dP(d;\boldsymbol{\theta})) = \frac{\kappa+\xi\overline{m}}{\xi-1} = 2\left(\overline{m}+\overline{M}\right). \tag{26}$$

Since $m(t)+M(t)$ are i.i.d., the standard law of large numbers gives:

$$\textstyle\sum_{d=0}^{\infty} d\frac{D_t(d)}{|V(t)|} \xrightarrow{P} 2(\overline{m}_0+\overline{M}_0) = \sum_{d=0}^{\infty} dP(d;\boldsymbol{\theta}_0)).$$

From (25) it follows that

$$\textstyle\sum_{d=0}^{\widetilde{d}(t)-1} d\frac{D_t(d)}{|V(t)|} = \left(\sum_{d=0}^{\widetilde{d}(t)-1} dP(d;\boldsymbol{\theta}_0)\right)\left(1+O_P\left(\frac{1}{\sqrt{\ln t}}\right)\right) \xrightarrow{P} \sum_{d=0}^{\widetilde{d}(t)-1} dP(d;\boldsymbol{\theta}_0).$$

Thus,

$$\textstyle\sum_{d=\widetilde{d}(t)}^{\infty} d\frac{D_t(d)}{|V(t)|} = \sum_{d=0}^{\infty} d\frac{D_t(d)}{|V(t)|} - \sum_{d=0}^{\widetilde{d}(t)-1} d\frac{D_t(d)}{|V(t)|} \xrightarrow{P} \sum_{d=\widetilde{d}(t)}^{\infty} dP(d;\boldsymbol{\theta}_0) \xrightarrow{P} 0,$$

where the last implication follows from inequality (24):

$$\begin{aligned}\textstyle\sum_{d=\widetilde{d}(t)}^{\infty} dP(d;\boldsymbol{\theta}_0) &< \textstyle\sum_{d=\widetilde{d}(t)}^{\infty} dC(\boldsymbol{\theta}_0)\left(d+\kappa_0\right)^{-1-\xi_0} < C(\boldsymbol{\theta}_0)\int_{\widetilde{d}(t)-1}^{\infty}(x+\kappa_0)^{-\xi_0}dx \\ &= \frac{C(\boldsymbol{\theta}_0)}{\xi_0}\left(\widetilde{d}(t)+\kappa_0-1\right)^{1-\xi_0} \to 0,\end{aligned}$$

which completes the proof of $\sup_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}} \widehat{S}_3(\boldsymbol{\theta}) \xrightarrow{P} 0$.

■

Proof of Proposition 2. Denote $\overline{\xi} = \max_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}} \xi$ and $\overline{\kappa} = \max_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}} \kappa$. Palumbo (1998) shows that

$$\frac{\Gamma(k+\lambda)}{\Gamma(k+1)} < \left(k+\frac{\lambda}{2}\right)^{\lambda-1} \qquad \text{for } \lambda > 2 \text{ and } k \geq 0.$$

Thus,

$$|\ln P(d;\boldsymbol{\theta})| = -\ln P(d;\boldsymbol{\theta}) \leq \ln\left(\frac{\Gamma(d+\kappa+\xi+1)}{\underline{C}\Gamma(d+\kappa)}\right) < -\ln\underline{C} + \left(1+\overline{\xi}\right)\ln\left(d+\overline{\kappa}+\frac{\overline{\xi}}{2}\right),$$

where $\underline{C} = \min\limits_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}} \sum\limits_{m=0}^{P} p_m\xi\frac{\Gamma(m+\kappa+\xi)}{\Gamma(m+\kappa)} > 0$. Thus there is C such that $|\log P(d;\boldsymbol{\theta})| < C\ln(d+2)$ and Lemma 2 applies; i.e., $\sup_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}} \left|\widehat{L}_t(\boldsymbol{\theta}) - L_0(\boldsymbol{\theta})\right| \xrightarrow{P} 0$, where $L_0(\boldsymbol{\theta}) \equiv \sum_{d=0}^{\infty} P(d;\boldsymbol{\theta}_0)\ln P(d;\boldsymbol{\theta})$ is a continuous function.

$L_0(\boldsymbol{\theta})$ is uniquely maximized at $\boldsymbol{\theta}_0$ by information inequality. Indeed, it is clear that $\sum_{d=0}^{\infty} |\ln P(d;\boldsymbol{\theta})| \, P(d;\boldsymbol{\theta}_0) = -L_0(\boldsymbol{\theta}) < \infty$ for all $\boldsymbol{\theta} \in \overline{\boldsymbol{\Theta}}$. Moreover, if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, then there exists d such that $P(d;\boldsymbol{\theta}) \neq P(d;\boldsymbol{\theta}_0)$ and thus by the strict version of Jensen's inequality:

$$L_0(\boldsymbol{\theta}_0) - L_0(\boldsymbol{\theta}) = -\sum_{d=0}^{\infty} \ln \tfrac{P(d;\boldsymbol{\theta})}{P(d;\boldsymbol{\theta}_0)} P(d;\boldsymbol{\theta}_0) < \ln \left(\sum_{d=0}^{\infty} \tfrac{P(d;\boldsymbol{\theta})}{P(d;\boldsymbol{\theta}_0)} P(d;\boldsymbol{\theta}_0) \right) = \ln \left(\sum_{d=0}^{\infty} P(d;\boldsymbol{\theta}) \right) = 0. \tag{27}$$

Thus, if $\widehat{L}_t(\widehat{\boldsymbol{\theta}}) \geq \max_{\boldsymbol{\theta} \in \overline{\boldsymbol{\Theta}}} \widehat{L}_t(\boldsymbol{\theta}) + o_P(1)$, then all conditions of Theorem 2.1 of Newey and McFadden (1994) are satisfied and thus $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$. By definition $\widehat{L}_t(\widehat{\boldsymbol{\theta}}^1) = \max_{\boldsymbol{\theta} \in \overline{\boldsymbol{\Theta}}} L_t(\boldsymbol{\theta})$; so $\widehat{\boldsymbol{\theta}}^1 \xrightarrow{P} \boldsymbol{\theta}_0$. To solve for an estimate $\widehat{\boldsymbol{\theta}}^2$, we can substitute $\kappa(.) = 2(\widehat{\overline{m} + \overline{M}}) \, (\xi - 1) - \xi \overline{m}$ in $\widehat{L}_t(.)$ and maximize $\widehat{L}_t(\xi, \kappa(.), \boldsymbol{p})$ over ξ and $\boldsymbol{p}$. Since $\kappa(.)$ is continuous, $\widehat{\overline{m} + \overline{M}} \xrightarrow{P} \overline{m}_0 + \overline{M}_0$, and $\widehat{L}_t(\boldsymbol{\theta})$ uniformly converges to a continuous function $L_0(\boldsymbol{\theta})$, it follows that $\widehat{L}_t(\widehat{\boldsymbol{\theta}}^2) \geq \widehat{L}_t(\widehat{\xi}^1, \kappa(.), \widehat{\boldsymbol{p}}^1) = L_t(\widehat{\boldsymbol{\theta}}^1) + o_P(1)$, which implies that $\widehat{\boldsymbol{\theta}}^2 \xrightarrow{P} \boldsymbol{\theta}_0$. ∎

Proof of Proposition 3. See the proof of Theorem 2.6 in Newey and McFadden (1994) and replace Lemma 2.4 with Lemma 2 in the argument. ∎

Proof of Proposition 4. Consistency. We start by verifying conditions of Proposition 3 to make sure that $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$. Conditions (i) and (ii) hold by assumption and condition (iii) by inspection of $g\,(d, \theta)$.

From Section 3.2 for the case of $m(t) = 0$, we have

$$\ln P(d;\boldsymbol{\theta}) = \ln n_0(d;\boldsymbol{\theta}) = \ln \xi + \ln \Gamma\,(\kappa + \xi) - \ln \Gamma\,(\kappa) + \ln \Gamma\,(d + \kappa) - \ln \Gamma\,(d + \kappa + \xi + 1)\,.$$

The score function $s(d;\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \ln P(d;\boldsymbol{\theta})$ is:

$$s_\xi(d;\boldsymbol{\theta}) = \tfrac{1}{\xi} + \psi\,(\kappa + \xi) - \psi\,(d + \kappa + \xi + 1)\,, \tag{28}$$

$$s_\kappa(d;\boldsymbol{\theta}) = \psi\,(\kappa + \xi) - \psi\,(\kappa) + \psi\,(d + \kappa) - \psi\,(d + \kappa + \xi + 1)\,, \tag{29}$$

where $\psi(.)$ is a polygamma function. Polygamma function of order m is defined as $\psi^{(m)}(z) = \left(\frac{d}{dz}\right)^{m+1} \ln \Gamma(z)$ with $\psi(z) = \psi^{(0)}(z)$. Qi et al. (2005) shows that for $x > 0$:

$$\tfrac{1}{2x} - \tfrac{1}{12x^2} < \psi\,(x + 1) - \ln x < \tfrac{1}{2x},$$

which implies that there is C such that

$$\| g(d;\boldsymbol{\theta}) \| < C\,(d + 1)\,, \tag{30}$$

for all $d \geq 0$ and all $\boldsymbol{\theta} \in \overline{\boldsymbol{\Theta}}$; so condition (iv) of Proposition 3 holds. Finally, equation (12) holds for $g_d\,(d;\boldsymbol{\theta}) = d - \frac{\kappa}{\xi - 1}$ by (26) and for $s\,(d;\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \ln P(d;\boldsymbol{\theta})$ by (27) and interchangeability of summation and differentiation (see Theorems 7.10 and 7.17 of Rudin (1976)).

Asymptotic normality. To prove Proposition 4, we verify all conditions of Theorem 3.2 in Newey and McFadden (1994). $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$ is shown above, conditions (i), (iii), and (v) are satisfied by assumption, and condition (ii) is satisfied because $\frac{\kappa}{\xi-1}$ and $\widehat{L}_t(\boldsymbol{\theta})$ are twice continuously differentiable. Thus, we only need to check condition (iv) that there is a continuous $G(\boldsymbol{\theta})$ and $\sup_{\boldsymbol{\theta}\in\overline{\boldsymbol{\Theta}}} \left\| \sum_{d=0}^{\infty} \nabla_{\boldsymbol{\theta}} g(d;\boldsymbol{\theta}_0) \frac{D_t(d)}{t} - G(\boldsymbol{\theta}) \right\| \xrightarrow{P} 0$.

Denote $G(d;\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} g(d;\boldsymbol{\theta})$, then the last row of $G(d;\boldsymbol{\theta})$ is given by

$$\nabla_{\boldsymbol{\theta}} \left(d - \frac{\kappa}{\xi-1} \right) = \begin{pmatrix} \frac{\kappa}{(\xi-1)^2} & -\frac{1}{\xi-1} \end{pmatrix}.$$

Let us now calculate $h(d;\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta\theta}} \ln P(d;\boldsymbol{\theta})$

$$h_{\xi\xi}(d;\boldsymbol{\theta}) = -\frac{1}{\xi^2} + \psi^{(1)}(\kappa+\xi) - \psi^{(1)}(d+\kappa+\xi+1), \tag{31}$$

$$h_{\kappa\kappa}(d;\boldsymbol{\theta}) = \psi^{(1)}(\kappa+\xi) - \psi^{(1)}(\kappa) + \psi^{(1)}(d+\kappa) - \psi^{(1)}(d+\kappa+\xi+1), \tag{32}$$

$$h_{\xi\kappa}(d;\boldsymbol{\theta}) = \psi^{(1)}(\kappa+\xi) - \psi^{(1)}(d+\kappa+\xi+1). \tag{33}$$

Qi et al. (2005) shows that for $x > 0$

$$\frac{1}{2x^2} - \frac{1}{6x^3} < \frac{1}{x} - \psi^{(1)}(x+1) < \frac{1}{2x^2} - \frac{1}{6x^3} + \frac{1}{30x^5},$$

which implies that there is C such that

$$\|G(d;\boldsymbol{\theta})\| < C, \tag{34}$$

for all $d \geq 0$ and all $\boldsymbol{\theta} \in \overline{\boldsymbol{\Theta}}$. In addition $G(d;\boldsymbol{\theta})$ is continuous; so Lemma 2 applies. Therefore, condition (iv) of Theorem 3.2 in Newey and McFadden (1994) holds and

$$G(\boldsymbol{\theta}) = \sum_{d=0}^{\infty} \begin{pmatrix} h_{\xi\xi}(d;\boldsymbol{\theta}) & h_{\xi\kappa}(d;\boldsymbol{\theta}) \\ h_{\xi_k}(d;\boldsymbol{\theta}) & h_{\kappa\kappa}(d;\boldsymbol{\theta}) \\ \frac{\kappa}{(\xi-1)^2} & -\frac{1}{\xi-1} \end{pmatrix} n_0(d;\boldsymbol{\theta}_0)$$

with $h_{\xi\xi}(d;\boldsymbol{\theta})$, $h_{\kappa\kappa}(d;\boldsymbol{\theta})$, and $h_{\xi\kappa}(d;\boldsymbol{\theta})$ given by (31), (32), and (33) respectively. Again, we interchange the order of summation and differentiation using Theorems 7.10 and 7.17 of Rudin (1976). ■

Proof of Lemma 3. See the proof of Theorem 3.1 in Cooper (2006). ■

Proof of Claim 2. **Part 1.** To prove (21) we use Lyapounov's central limit theorem (Theorem 27.3 of Billingsley (1995)). It is sufficient to show that

$$\sqrt{t} \left(\mathbb{E} \left[\sum_{d=0}^{\infty} d \frac{D_t(d)}{t} \right] - \frac{\kappa}{\xi-1} \right) \to 0. \tag{35}$$

$$\frac{1}{t}\sum_{v=1}^{t} Var\left(d(v,t)\right) \to \frac{r\xi}{(\xi-1)(\xi-2)}, \tag{36}$$

and there exists $\delta > 0$ such that $\left|d(v,t) - \mathbb{E}\left[d(v,t)\right]\right|^{2+\delta}$ are integrable for all $v = 1, ..., t$ and that Lyapounov's condition holds

$$\lim_{t\to\infty} \sum_{v=1}^{t} \frac{\mathbb{E}\left(\left|d(v,t)-\mathbb{E}[d(v,t)]\right|^{2+\delta}\right)}{\left(t\frac{r\xi}{(\xi-1)(\xi-2)}\right)^{\frac{2+\delta}{2}}} = 0. \tag{37}$$

To show (35), (36) and (37), we rely on simple summation-integration inequalities: $\int_a^{b+1} f(x)dx \leq \sum_{x=a}^{b} f(x) \leq \int_{a-1}^{b} f(x)dx$ for a decreasing function $f(x)$ and $\int_{a-1}^{b} f(x)dx \leq \sum_{x=a}^{b} f(x) \leq \int_a^{b+1} f(x)dx$ for an increasing function $f(x)$.

By definition of $D_t(d)$ and the fact that the mean of $X \sim$NegBin(r,p) is given by $\mathbb{E}\left[X\right] = r\left(1-p\right)/p$, we have:

$$\mathbb{E}\left[\sum_{d=0}^{\infty} d\frac{D_t(d)}{t}\right] = \mathbb{E}\left[\sum_{v=1}^{t} \frac{d(v,t)}{t}\right] = \frac{1}{t}\sum_{v=1}^{t} \kappa\frac{1-(v/t)^{1/\xi}}{(v/t)^{1/\xi}}. \tag{38}$$

Since $\left(1-(v/t)^{1/\xi}\right)/\left(v/t\right)^{1/\xi}$ is monotonically decreasing in v we get the following bounds:

$$\mathbb{E}\left[\sum_{d=0}^{\infty} d\frac{D_t(d)}{t}\right] \leq \frac{1}{t}\int_0^t \kappa\frac{1-(v/t)^{1/\xi}}{(v/t)^{1/\xi}}dv = \kappa\xi\int_0^1 p^{\xi-2}(1-p)dp = \frac{\kappa}{\xi-1},$$
$$\mathbb{E}\left[\sum_{d=0}^{\infty} d\frac{D_t(d)}{t}\right] \geq \frac{1}{t}\int_1^t \kappa\frac{1-(v/t)^{1/\xi}}{(v/t)^{1/\xi}}dv = \kappa\xi\int_{(1/t)^{1/\xi}}^1 p^{\xi-2}(1-p)dp = \frac{\kappa}{\xi-1} + O\left(t^{\frac{1}{\xi}-1}\right).$$

If $\xi > 2$ these bounds imply (35):

$$\sqrt{t}\left(\mathbb{E}\left[\sum_{d=0}^{\infty} d\frac{D_t(d)}{t}\right] - \frac{\kappa}{\xi-1}\right) = O\left(t^{\frac{1}{\xi}-\frac{1}{2}}\right) \to 0.$$

Using summation-integration inequalities and the fact that the variance of $X \sim$NegBin(r,p) is given by $Var\left[X\right] = r\left(1-p\right)/p^2$, we establish (36):

$$\frac{1}{t}\sum_{v=1}^{t} Var\left(d(v,t)\right) = \frac{1}{t}\sum_{v=1}^{t} \kappa\frac{1-(v/t)^{1/\xi}}{(v/t)^{2/\xi}},$$
$$\frac{1}{t}\sum_{v=1}^{t} Var\left(d(v,t)\right) - \frac{\kappa\xi}{(\xi-1)(\xi-2)} = O\left(t^{\frac{2}{\xi}-1}\right) \to 0.$$

Finally we need to check Lyapounov's condition. First, we notice that $\left|d(v,t) - \mathbb{E}\left[d(v,t)\right]\right|^{2+\delta}$ are integrable, because

$$\sum_{d=0}^{\infty} \left|d - \mathbb{E}\left[d(v,t)\right]\right|^{2+\delta} f(d;\kappa,(v/t)^{1/\xi})$$

converges for all v since $\left|d - \mathbb{E}\left[d(v,t)\right]\right|^{2+\delta} f(d;\kappa,(v/t)^{1/\xi}) = O\left(d^{1+\delta+\kappa}\left(\frac{v}{t}\right)^{\frac{\kappa}{\xi}}\left(1-\left(\frac{v}{t}\right)^{1/\xi}\right)^d\right)$.

Second, using a generalized mean inequality ($\left(\frac{1}{2}\left(x^p+y^p\right)\right)^{\frac{1}{p}} \geq \frac{1}{2}\left(x+y\right)$ for $p \geq 1$ and positive x and y) we get that

$$\left|d(v,t) - \mathbb{E}\left[d(v,t)\right]\right|^{2+\delta} \leq \left|d(v,t) + \mathbb{E}\left[d(v,t)\right]\right|^{2+\delta} \leq 2^{1+\delta}\left(d(v,t)^{2+\delta} + \mathbb{E}\left[d(v,t)\right]^{2+\delta}\right).$$

Third, Jensen's inequality yields $\frac{1}{t}\sum_{v=1}^{t}\mathbb{E}\left[d(v,t)\right]^{2+\delta} \leq \mathbb{E}\left(\frac{1}{t}\sum_{v=1}^{t} d(v,t)^{2+\delta}\right)$. Thus, to establish Lyapounov's condition, it is sufficient to show that there exists $\delta > 0$ such that $\mathbb{E}\left(\frac{1}{t}\sum_{v=1}^{t} d(v,t)^{2+\delta}\right)$ is bounded in t. Because $d(v,t)^{2+\delta}$ are integrable, we have

$$\mathbb{E}\left(\frac{1}{t}\sum_{v=1}^{t} d(v,t)^{2+\delta}\right) = \mathbb{E}\left(\sum_{d=0}^{\infty} d^{2+\delta}\frac{D_t(d)}{t}\right) = \sum_{d=0}^{\infty} d^{2+\delta}\mathbb{E}\left(\frac{D_t(d)}{t}\right).$$

Using definition of $D_t(d)$, we calculate

$$\mathbb{E}\left(\frac{D_t(d)}{t}\right) = \frac{1}{t}\sum_{v=1}^{t} f\left(d;\kappa,(v/t)^{1/\xi}\right).$$

To obtain an upper bound on $\mathbb{E}\left(D_t(d)/t\right)$, we find regions on which $f\left(d;\kappa,(v/t)^{1/\xi}\right)$ is monotone in v and use summation-integration inequalities. Since $p=(v/t)^{1/\xi}$ is a monotone transformation of v it is sufficient to find regions on which $f(d;\kappa,p)$ is monotone in p

$$\frac{df(d;\kappa,p)}{dp} = \binom{d+\kappa-1}{d}p^{\kappa-1}(1-p)^{d-1}\left[\kappa-(\kappa+d)p\right].$$

Thus, $f(0;\kappa,p)$ is increasing in p for all $p\in[0,1]$, and $f(d;\kappa,p)$ is increasing in p on the interval $[0,\kappa/\left(\kappa+d\right))$ and decreasing in p on the interval $(\kappa/\left(\kappa+d\right),1]$, which leads to

$$\mathbb{E}\left(\frac{D_t(d)}{t}\right) \leq \frac{1}{t}\left[\int_{v=0}^{t} f\left(d;\kappa,(v/t)^{1/\xi}\right)dv + \max_{v\in\{1,..,t\}} f\left(d;\kappa,(v/t)^{1/\xi}\right)\right].$$

The integral simplifies to:

$$\begin{aligned}\frac{1}{t}\int_{v=0}^{t} f\left(d;\kappa,(v/t)^{1/\xi}\right)dv &= \binom{d+\kappa-1}{d}\int_{v=0}^{t}\xi p^{\kappa+\xi-1}(1-p)^{d}dp = \binom{d+\kappa-1}{d}\xi B(\kappa+\xi,d+1)\\ &= \xi\frac{\Gamma(d+\kappa)}{\Gamma(d+1)\Gamma(\kappa)}\frac{\Gamma(\kappa+\xi)\Gamma(d+1)}{\Gamma(\kappa+\xi+d+1)} = n_0(d;\xi,\kappa),\end{aligned}$$

where $B(x,y)$ is the beta function given by (5).

Since $n_0(d;\xi,\kappa)=O(d^{-1-\xi})$ (Claim 1 part 2), $\sum_{d=0}^{\infty} d^{2+\delta}n_0(d;\xi,\kappa)$ converges for $\delta<\xi-2$. Thus, to complete the proof, we need to show that

$$\mathbb{E}\left(\frac{1}{t}\sum_{d=0}^{\infty} d^{2+\delta}\max_{v\in\{1,..,t\}} f\left(d;\kappa,(v/t)^{1/\xi}\right)\right)$$

is bounded in t for some $\delta\in(0,\xi-2)$. Since $f\left(d;\kappa,p\right)$ is maximized at $p=\kappa/(\kappa+d)$, we have $\arg\max_{v\in\{1,..,t\}} f\left(d;\kappa,(v/t)^{1/\xi}\right)=1$ for all sufficiently large d. Finally, $d^{2+\delta}f\left(d;\kappa,(1/t)^{1/\xi}\right) = O\left(d^{1+\kappa+\delta}t^{-\kappa/\xi}e^{-dt^{-1/\xi}}\right)$ and $\int_0^{\infty} x^{1+\kappa+\delta}e^{-x}dx<\infty$ imply

$$\mathbb{E}\left(\frac{1}{t}\sum_{d=0}^{\infty} d^{2+\delta}\max_{v\in\{1,..,t\}} f\left(d;\kappa,(v/t)^{1/\xi}\right)\right) = O\left(t^{-1-\frac{\kappa}{\xi}+\frac{2+\kappa+\delta}{\xi}}\right) = O\left(t^{\frac{2+\delta}{\xi}-1}\right),$$

which converges to 0 as $t\to\infty$ if $\delta\in(0,\xi-2)$.

Part 2. Using summation-integration inequalities and the fact that the variance of X ~NegBin(r, p) is given by $Var\left[X\right] = r\left(1-p\right)/p^2$, we get:

$$Var\left(\sqrt{t}\sum_{d=0}^{\infty} d\frac{D_t(d)}{t}\right) = \frac{1}{t}\sum_{v=1}^{t} Var\left(d(v,t)\right) = \frac{1}{t}\sum_{v=1}^{t}\kappa\frac{1-(v/t)^{1/\xi}}{(v/t)^{2/\xi}},$$
$$\frac{1}{t}\sum_{v=1}^{t} Var\left(d(v,t)\right) \geq \frac{1}{t}\int_1^t \kappa\frac{1-(v/t)^{1/\xi}}{(v/t)^{2/\xi}}dv = \xi\kappa\int_{(1/t)^{1/\xi}}^{1} p^{\xi-3}(1-p)dp.$$

If $\xi = 2$, then as $t \to \infty$:

$$\int_{(1/t)^{1/2}}^{1} p^{-1}(1-p)dp = \left.(\ln p - p)\right|_{(1/t)^{1/2}}^{1} = -1 + \frac{1}{2}\ln t + \frac{1}{\sqrt{t}} \to \infty.$$

If $\xi < 2$, then as $t \to \infty$:

$$\int_{(1/t)^{1/\xi}}^{1} p^{\xi-3}(1-p)dp = \left.\left(-\frac{p^{\xi-2}}{2-\xi} - \frac{p^{\xi-1}}{\xi-1}\right)\right|_{(1/t)^{1/\xi}}^{1} = -\frac{1}{(2-\xi)(\xi-1)} + \frac{t^{\frac{2}{\xi}-1}}{2-\xi} + \frac{t^{\frac{1}{\xi}-1}}{\xi-1} \to \infty.$$

■

References

Albert, Reka and Albert-Laszlo Barabasi (2002) "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics*, **74** (1), pp. 47–97.

Atalay, Enghin (2013) "Sources of Variation in Social Networks," *Games and Economic Behavior*, **79**, pp. 106–131.

Atalay, Enghin, Ali Hortaçsu, James Roberts, and Chad Syverson (2011) "Network Structure of Production," *Proceedings of the National Academy of Sciences*, **108** (13), pp. 5199–5202.

Barabasi, Albert-Laszlo and Reka Albert (1999) "Emergence of Scaling in Random Networks," *Science*, **286** (5439), pp. 509–512.

Billingsley, Patrick (1995) *Probability and Measure*, New York: John Wiley and Sons.

Bollobas, Bela, Oliver Riordan, Joel Spencer, and Gabor Tusnady (2001) "The Degree Sequence of a Scale-Free Random Graph Process," *Random Structures and Algorithms*, **18** (3), pp. 279–290.

Bollobas, Bela and Oliver Riordan (2003) "Mathematical Results on Scale-Free Random Graphs," in Stefan Bornholdt and Heinz Georg Schuster eds. *Handbook of Graphs and Networks*, Berlin: Wiley, pp. 1–34.

Buckley, Pierce G. and Deryk Osthus (2004) "Popularity Based Random Graph Models Leading to a Scale-Free Degree Sequence," *Discrete Mathematics*, **282** (1-3), pp. 53–68.

Burnham, Kenneth P. and David R. Anderson (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, New York: Springer Verlag.

Chandrasekhar, Arun and Matthew Jackson (2013) "Tractable and Consistent Random Graph Models." Mimeo, Stanford University.

Christakis, Nicholas A, James H Fowler, Guido W Imbens, and Karthik Kalyanaraman (2010) "An Empirical Model for Strategic Network Formation." Mimeo, Harvard University.

Comola, Margherita and Marcel Fafchamps (2013) "Testing Unilateral and Bilateral Link Formation," *Economic Journal*, **forthcoming**.

Cooper, Colin (2006) "Distribution of Vertex Degree in Web-Graphs," *Combinatorics, Probability and Computing*, **15** (5), pp. 637–661.

Cooper, Colin and Alan Frieze (2003) "A General Model of Web Graphs," *Random Structures and Algorithms*, **22** (3), pp. 311–335.

Dallal, Gerard E. and Leland Wilkinson (1986) "An Analytic Approximation to the Distribution of Lilliefors's Test Statistic for Normality," *American Statistician*, **40** (4), pp. 294–296.

Dorogovtsev, Sergey N., Jose F.F. Mendes, and Alexandr N. Samukhin (2000) "Structure of Growing Networks with Preferential Linking," *Physical Review Letters*, **85** (21), pp. 4633–4636.

Dorogovtsev, Sergey N. and Jose F.F. Mendes (2002) "Evolution of Networks," *Advances in Physics*, **51** (4), pp. 1079–1187.

Gabaix, Xavier and Rustam Ibragimov (2011) "Rank- 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," *Journal of Business & Economic Statistics*, **29** (1), pp. 24–39.

Goldsmith-Pinkham, Paul and Guido W Imbens (2013) "Social Networks and the Identification of Peer Effects," *Journal of Business & Economic Statistics*, **31** (3), pp. 253–264.

Goldstein, Michel L., Steven A. Morris, and Gary G. Yen (2004) "Problems with Fitting to the Power-Law Distribution," *European Physical Journal B*, **41** (2), pp. 255–258.

Handcock, Mark S. and James Holland Jones (2004) "Likelihood-Based Inference for Stochastic Models of Sexual Network Formation," *Theoretical Population Biology*, **65** (4), pp. 413–422.

Hill, Bruce M. (1975) "A Simple General Approach to Inference about the Tail of a Distribution," *Annals of Statistics*, **3** (5), pp. 1163–1174.

Jackson, Matthew O. (2008) *Social and Economic Networks*, Princeton: Princeton University Press.

Jackson, Matthew O. and Brian W. Rogers (2007) "Meeting Strangers and Friends of Friends: How Random Are Social Networks?" *American Economic Review*, **97** (3), pp. 890–915.

Konig, Michael (2012) "The Formation of Networks with Local Spillovers and Limited Observability." Mimeo, University of Zurich.

Krapivsky, Paul L., Sidney Redner, and Francois Leyvraz (2000) "Connectivity of Growing Random Networks," *Physical Review Letters*, **85** (21), pp. 4629–4632.

Kumar, Ravi, Prabhakar Raghavan, Sridhar Rajagopalan, D. Sivakumar, Andrew Tomkins, and Eli Upfal (2000) "Stochastic Models for the Web Graph," in *41st Annual Symposium on Foundations of Computer Science*, pp. 57–65.

Lilliefors, Hubert W. (1967) "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," *Journal of the American Statistical Association*, **62** (318), pp. 399–402.

Mele, Angelo (2013) "A Structural Model of Segregation in Social Networks." Mimeo, Johns Hopkins University.

Newey, Whitney K. and Daniel McFadden (1994) "Large Sample Estimation and Hypothesis Testing," in Robert F. Engle and Daniel L. McFadden eds. *Handbook of Econometrics*, **4**, Amsterdam: Elsevier, pp. 2111–2245.

Newman, Mark EJ (2005) "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, **46** (5), pp. 323–351.

Palumbo, Biagio (1998) "A Generalization of Some Inequalities for the Gamma Function," *Journal of Computational and Applied Mathematics*, **88** (2), pp. 255–268.

Pennock, David M., Gary W. Flake, Steve Lawrence, Eric J. Glover, and C. Lee Giles (2002) “Winners Don’t Take All: Characterizing the Competition for Links on the Web,” *PNAS*, **99** (8), pp. 5207–5211.

Qi, Feng, Run-Qing Cui, Chao-Ping Chen, and Bai-Ni Guo (2005) “Some Completely Monotonic Functions Involving Polygamma Functions and an Application,” *Journal of Mathematical Analysis and Applications*, **310** (1), pp. 303–308.

Rudin, Walter (1976) *Principles of Mathematical Analysis*, New York: McGraw-Hill.

Sheng, Shuyang (2012) “Identification and Estimation of Network Formation Games.” Mimeo, University of Southern California.

Simon, Herbert A. (1955) “On a Class of Skew Distribution Functions,” *Biometrika*, **42** (3-4), pp. 425–440.

van der Vaart, Aad (2000) *Asymptotic Statistics*, New York: Cambridge University Press.

Figure 1: Degree distributions for a simulation of the CF model with benchmark parameters.

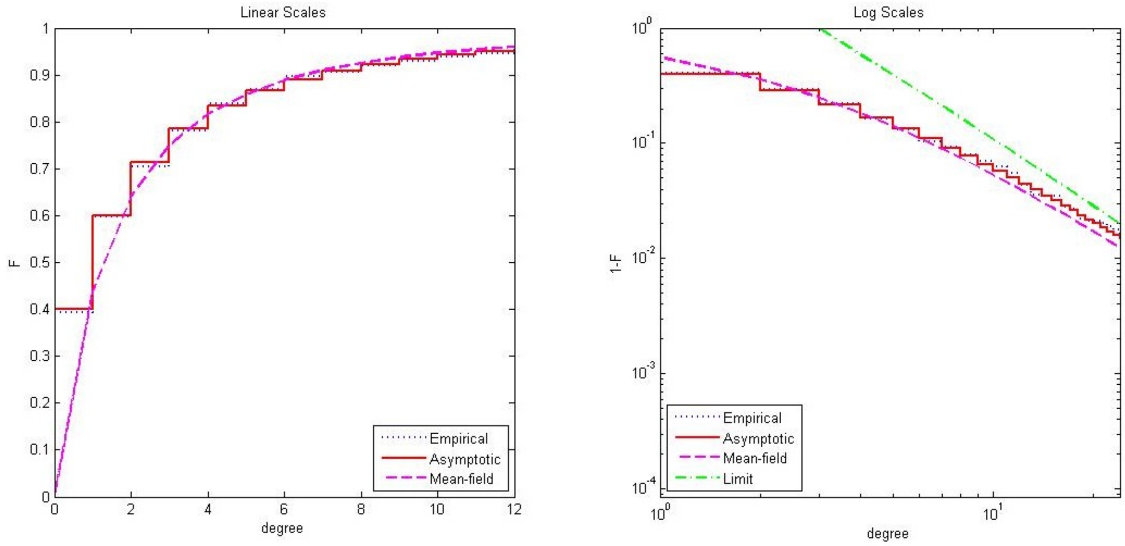

Table 1: The PML and GMM estimations for benchmark parameters

	PML^0	PML_P^0	$\text{GMM}_{-\kappa}^0$	$\text{GMM}_{-\kappa}^1$
$mean(\eta)$	0.501	0.500	0.501	0.498
$std(\eta)$	0.027	0.018	0.015	0.006
$bias(\eta)$	0%	0%	0%	0%
$D_{\max}^{\eta}$	0.016	0.023	0.026	0.217
$mean(\overline{m}+\overline{M})$	1.500	1.500	1.500	1.500
$std(\overline{m}+\overline{M})$	0.043	0.017	0.017	0.017
$mean(p_0)$				0.980
$std(p_0)$				0.027

Table 2: The NLLS and Hill estimations for benchmark parameters

	NLLS	NLLS_R	Hill AIC	Hill BIC
$mean(\eta)$	0.590	0.613	0.686	0.809
$std(\eta)$	0.049	0.078	0.077	0.070
$bias(\eta)$	18%	23%	37%	62%
$D_{\max}^{\eta}$	0.012	0.038	0.039	0.075
$mean(\overline{m}+\overline{M})$	1.500	1.500		
$std(\overline{m}+\overline{M})$	0.017	0.017		
$median(D)$			6	4
$std(D)$			2.67	0.77

Table 3: Different numbers of vertices

	$t = 400$		$t = 100$		$t = 25$	
	$\text{GMM}^0_{-\kappa}$	NLLS	$\text{GMM}^0_{-\kappa}$	NLLS	$\text{GMM}^0_{-\kappa}$	NLLS_M
$mean(\eta)$	0.499	0.637	0.496	0.661	0.472	0.509
$std(\eta)$	0.023	0.093	0.045	0.124	0.088	0.163
$bias(\eta)$	0%	27%	-1%	32%	-6%	2%
$D^\eta_{\max}$	0.028	0.036	0.020	0.242	0.039	0.052
$mean(\overline{m} + \overline{M})$	1.500	1.500	1.500	1.500	1.500	1.500
$std(\overline{m} + \overline{M})$	0.024	0.024	0.050	0.050	0.102	0.102

Table 4: Different parameter values

	$\eta = 0.2$		$\eta = 0.8$		$\boldsymbol{p} = (0.5, 0.5)$		
	$\text{GMM}^0_{-\kappa}$	NLLS	$\text{GMM}^0_{-\kappa}$	NLLS	$\text{GMM}^0_{-\kappa}$	$\text{GMM}^1_{-\kappa}$	NLLS
$mean(\eta)$	0.200	0.268	0.799	0.875	0.393	0.495	0.516
$std(\eta)$	0.002	0.032	0.010	0.033	0.019	0.036	0.052
$bias(\eta)$	0%	34%	0%	9%	-21%	-1%	3%
$D^\eta_{\max}$	0.024	0.018	0.020	0.214	0.022	0.350	0.033
$mean(\overline{m} + \overline{M})$	1.500	1.500	1.500	1.500	1.500	1.500	1.500
$std(\overline{m} + \overline{M})$	0.017	0.017	0.017	0.017	0.017	0.017	0.017
$mean(p_0)$						0.503	
$std(p_0)$						0.047	

Table 5: Different initial graphs and different distributions of $M(t)$

	$G^J(1)$		$q_0 = q_3 = 1/2$		$B_1 = 0,\ C_1 = 1$	
	$\text{GMM}^0_{-\kappa}$	NLLS_M	$\text{GMM}^0_{-\kappa}$	NLLS	$\text{GMM}^0_{-\kappa}$	NLLS
$mean(\eta)$	0.497	0.652	0.502	0.587	0.500	0.589
$std(\eta)$	0.015	0.038	0.018	0.054	0.012	0.045
$bias(\eta)$	-1%	30%	0%	17%	0%	18%
$D^\eta_{\max}$	0.028	0.018	0.021	0.020	0.023	0.033
$mean(\overline{m} + \overline{M})$	1.500	1.500	1.500	1.500	1.500	1.500
$std(\overline{m} + \overline{M})$	0.017	0.017	0.051	0.051	0.017	0.017