Supplementary Appendix for
Estimation of a Scale-Free Network Formation Model

Anton Kolotilin and Valentyn Panchenko

June, 2018

1 NLS Estimator

In Section 2, we derived the asymptotic degree distribution $P(d)$. Below, we approximate the asymptotic degree distribution of the CF model using an alternative method: mean-field approximation. The NLS estimator discussed below is based on the mean-field approximation of the asymptotic degree distribution.

1.1 Mean-Field Approximation of the Degree Distribution

Using the mean-field method of Barabasi and Albert (1999), we approximate the CF network formation process by a continuous time process such that

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

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

As $t\to\infty$, the differential equation asymptotes to

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

where $\eta\kappa$ is the expected number of edge endpoints added uniformly at random per vertex. The solution to this differential equation is:

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

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, vertices added at an earlier time period ("older" vertices) have a larger expected degree than vertices added at later periods ("younger" vertices). Thus, the cumulative distribution of expected degrees of vertices with the initial degree m can be approximated by (for $d \geq m$):

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

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

$$F^{\mathrm{MF}}(d) = \sum_{m=0}^{\min\{P,d\}} p_m F_t^m(d). \tag{1}$$

For a sufficiently large d and a constant K, the complementary cumulative distribution can be approximated by a power-law distribution:

$$1 - F^{\mathrm{tail}}(d) = Cd^{-1/\eta}.$$

This result is analogous to part 2 of Corollary 1 in Section 2.2, which shows that the asymptotic degree distribution $P(d)$ has a power-law tail.

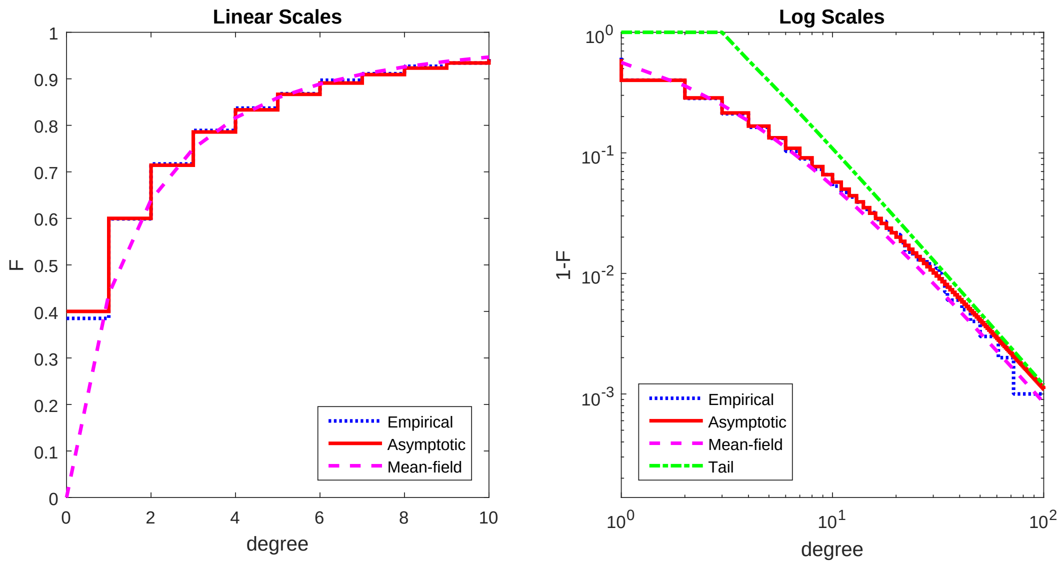
Now we can connect various approximations of the degree distribution to specific estimators: the PML and GMM estimators are derived from the asymptotic degree distribution $P(d)$, the NLS estimator is derived from the mean-field approximation, and the Hill estimator and other tail estimators are based on the power-law approximation in the tail.

Figure 1 compares the cumulative distributions (left panel, linear scale) and the complementary cumulative distributions (right panel, log scale) for the benchmark specification: $t = 1000$, $p_0 = 1$ ($m(t) = 0$), $q_1 = q_2 = 0.5$ ($\overline{M} = 1.5$), and $A_1 = B_1 = C_1 = 0.5$ ($\eta = 0.5$). The empirical cumulative distribution from a simulated network is much closer to the asymptotic approximation than to the mean-field and tail approximations, especially for small degrees d. This may explain inferior performance of the estimators based on the latter two approximations.

1.2 NLS Estimator

We now turn to the NLS estimator commonly used for scale-free network formation models (see Pennock et al., 2002; Jackson and Rogers, 2007; Jackson, 2008).

Figure 1: Degree distributions for a simulation of the CF model



In order to derive the NLS estimator, we need to fix m, that is, assume that $m(t) = m$.[1] Since most of the real networks have vertices with zero degree, we set $m(t) = 0$. Under this assumption, (1) can be expressed as

$$\ln(1 - F^{\mathrm{MF}}(d)) = 1/\eta \left(\ln\left(2\overline{M}(1/\eta - 1)\right) - \ln\left(d + 2\overline{M}(1/\eta - 1)\right)\right).$$

Moreover, $\overline{M}$ can be consistently estimated as[2]

$$\widehat{\overline{M}} = \frac{1}{2}\sum_{d=0}^{\infty} d\frac{D_t(d)}{t}.$$

Parameter η is then estimated by numerically minimizing the quadratic loss:

$$\widehat{\eta}^{\mathrm{NLS}} = \arg\min_{\eta} \sum_{d=0}^{\infty} \left(\ln(1 - \widehat{F}_t(d)) - 1/\eta\left(\ln(2\widehat{\overline{M}}(1/\eta - 1)) - \ln(d + 2\widehat{\overline{M}}(1/\eta - 1))\right)\right)^2,$$

where $\widehat{F}_t(d)$ is an empirical analogue of the cumulative distribution.

There are several alternatives of what can be used as degree observations for the NLS estimator: (i) *observed distinct* degrees (without repetition), (ii) *consecutive degrees in the range* $[d_{\mathrm{min}}, d_{\mathrm{max}}]$ where $d_{\mathrm{min}} = \min_v d(v,t)$ and $d_{\mathrm{max}} = \max_v d(v,t)$,[3] and (iii) *observed degrees with repetition* (Newman, 2005, Appendix A).

[1] Note that Pennock et al. (2002) assume $m(t) = 0$ and $M(t) = M$ for some constant M, whereas Jackson (2008) and Jackson and Rogers (2007) assume $m(t) = m$ and $M(t) = 0$ for some constant m.

[2] Hereafter, we assume $|V(t)| = t$ for notational simplicity.

[3] Our reconstruction suggests that this method in conjunction with the empirical cumulative distribution for $\widehat{F}_t(d)$ and removed d_{max} is used in Jackson (2008).

$\widehat{F}_t(d)$ can also be specified in several ways. The empirical cumulative distribution is a common candidate, $\widehat{F}_t(d) = \sum_{i=0}^{d} D_t(i)/t$. But, in this case, $\widehat{F}_t(d_{\max}) = 1$, and hence, $\ln(1-\widehat{F}_t(d_{\max}))$ is not defined. One way to overcome this issue is to remove the observation(s) with $d = d_{\max}$. Alternatively, $\widehat{F}_t(d)$ can be scaled with $1/(t+1)$ instead of $1/t$ (see, e.g., Beirlant et al., 2006, p. 5). Moreover, when the observed degrees with repetition are used, $1-\widehat{F}_t$ can be specified as ordinal ranks (scaled by $1/t$) (see, Newman, 2005, Appendix A). In this case, each observation (vertex) is assigned a distinct ordinal number from 1 to t according to its degree d in descending order. Hence, $\widehat{F}_t$ changes in discrete steps of $1/t$. Gabaix and Ibragimov (2011) advocate using rank $-$ 1/2 adjustment for improved performance.

The NLS estimators based on the above definitions and adjustments are compared in the Excel spreadsheet of this Supplement. It appears that using (i) observed distinct degrees and (ii) observed degrees with repetition in conjunction with removed $d_{\max}$ yield the best performance. These estimators are reported in the main text of the paper as (i) NLS_D and (ii) NLS_R.

2 Tail Estimators

There is a well-developed literature on tail estimators, starting from Hill (1975), Pickands (1975), and Smith (1987); see Beirlant et al. (2006) for a detailed analysis and references. These estimators rely on a specific behavior in the tail of the distribution. Since the CF model yields a degree distribution with a power-law tail, tail estimators based on Pareto-type models are appropriate for estimating parameter η, which determines power-law parameter $1+1/\eta$. Most tail estimators are designed for continuous independently identically distributed random variables, but degrees in the CF model are discrete valued, interdependent, and not identically distributed. Moreover, an appropriate choice of the number of observations in the tail, called a tail cutoff, $d_t^\dagger$, after which the tail approximation holds, is crucial for these estimators. We will first assume that $d_t^\dagger$ is known and then, after introducing the estimators, we will discuss various methods for selecting $d_t^\dagger$.

2.1 Pareto-Type Distribution

A simple and popular way to estimate the power-law parameter is to run a rank-degree regression in logs. Specifically, denote increasingly ordered degree observations by $d_1 \leq \cdots \leq d_t$. The regression is $\ln j = c - \frac{1}{\eta} \ln d_{t-j+1}$ for j such that $d_{t-j+1} > d_t^\dagger$. Gabaix and Ibragimov (2011) propose an important simple bias-reducing adjustment. They recommend using $\ln(j - 1/2)$ instead of $\ln j$ in the regression. We implement this estimator and refer to it as the GI estimator.

Hill (1975) estimator is the main tail estimator. It can be derived as a maximum likelihood estimator based on the two assumptions: (i) the tail of the distribution follows continuous Pareto distribution with density $f(d) = \frac{1}{\eta d_t^\dagger}\left(\frac{d}{d_t^\dagger}\right)^{-1-1/\eta}$ conditional on $d > d_t^\dagger$, and (ii) these tail observations are independent,[4]

$$\widehat{\eta}^{\text{Hill}} = \frac{1}{k_t} \sum_{v:d(v,t)>d_t^\dagger} \ln \frac{d(v,t)}{d_t^\dagger} = \frac{1}{k_t} \sum_{d=d_t^\dagger+1}^{\infty} D_t(d) \ln \frac{d}{d_t^\dagger},$$

where $k_t = \sum_{d=d_t^\dagger+1}^{\infty} D_t(d)$ is the number of vertices that have degree greater than $d_t^\dagger$.

For a discrete distribution, Clauset et al. (2009) propose a simple adjustment for the Hill estimator,

$$\widehat{\eta}_{\text{C}}^{\text{Hill}} = \frac{1}{k_t} \sum_{d=d_t^\dagger+1}^{\infty} D_t(d) \ln \frac{d}{d_t^\dagger + 1/2}.$$

The discrete counterpart of the Pareto distribution is zeta distribution. Assuming the zeta distribution for the tail, the probability that a vertex has degree d, for $d > d_t^\dagger$, is

$$P(d) = \frac{d^{-1-1/\eta}}{\zeta(1 + 1/\eta, d_t^\dagger + 1)},$$

where $\zeta(1 + 1/\eta, d_t^\dagger + 1) = \sum_{i=0}^{\infty}(i + d_t^\dagger + 1)^{-(1+1/\eta)}$ is the Hurwitz zeta function. Goldstein et al. (2004) and Bauke (2007) use a maximum likelihood tail estimator for discrete data,

$$\widehat{\eta}_{\text{G}}^{\text{Hill}} = \underset{\eta}{\text{argmax}} - \frac{1 + 1/\eta}{k_t} \sum_{d=d_t^\dagger+1}^{\infty} D_t(d) \ln d - \ln \zeta(1 + 1/\eta, d_t^\dagger + 1).$$

2.2 Other Distributions

The Hill estimator, together with its variants discussed above, is applicable for estimating the tail of Pareto-type distributions, and thus of the degree distribution of the CF model. We now introduce other tail estimators applicable for estimating the tails of distributions belonging to the Pareto, Weibull, and Gumbel classes.

Pickands (1975) proposes a tail estimator which is based on sample quantiles in the tails,

$$\widehat{\eta}^{\text{Pic}} = \frac{1}{\ln 2} \ln \frac{d_{t-[k_t/4]} - d_{t-[k_t/2]}}{d_{t-[k_t/2]} - d_{t-k_t}}.$$

[4]Because of degree interdependences in the CF model, this and related estimators, which ignore the interdependences, should formally be referred to as pseudo maximum likelihood estimators. In the main text of the paper, we prove consistency of the Hill estimator.

Dekkers et al. (1989) propose an estimator based on higher moments of the Hill estimator,

$$\widehat{\eta}^{\text{Dek}} = \widehat{\eta}^{\text{Hill}} + 1 - \frac{1}{2}\left(1 - \frac{\left(\widehat{\eta}^{\text{Hill}}\right)^2}{\frac{1}{k_t}\sum_{d=d_t^\dagger+1}^{\infty} D_t(d)\left(\ln \frac{d}{d_t^\dagger}\right)^2}\right)^{-1}.$$

Smith (1987) proposes a maximum likelihood tail estimator assuming generalized Pareto distribution with density $f(d) = \frac{1}{\sigma}\left(1 + \frac{\eta(d-d_t^\dagger)}{\sigma}\right)^{-1/\eta-1}$ conditional on $d > d_t^\dagger$,

$$\widehat{\eta}^{\text{Smith}} = \underset{\eta,\sigma}{\operatorname{argmax}} - \frac{1+1/\eta}{k_t}\sum_{d=d_t^\dagger+1}^{\infty} D_t(d)\ln\left(1 + \frac{\eta(d-d_t^\dagger)}{\sigma}\right) - \ln\sigma.$$

2.3 Selection of Tail Cutoff

Up to this point, we have treated $d_t^\dagger$ as given. Next, we discuss several methods for selecting $d_t^\dagger$ as this is a crucial step for any tail estimator. For the Hill estimator, multiple methods are proposed in the literature (Beirlant et al., 2006, Chapter 4.7). We use a popular analytical method, which we refer to as MS, aiming to balance the asymptotic bias and variance by selecting $d_t^\dagger$ such that it minimizes the asymptotic mean squared error (AMSE) of the estimator (Beirlant et al., 1996; Matthys and Beirlant, 2000), given by

$$\text{AMSE}\left(\widehat{\eta}^{\text{Hill}}_{d_t^\dagger}\right) = \text{ABias}^2\left(\widehat{\eta}^{\text{Hill}}_{d_t^\dagger}\right) + \text{AVar}\left(\widehat{\eta}^{\text{Hill}}_{d_t^\dagger}\right),$$

where $\text{AVar}\left(\widehat{\eta}^{\text{Hill}}_{d_t^\dagger}\right) = \eta^2/k_t$. Lower $d_t^\dagger$ yields higher k_t which, in turn, reduces the variance, but increases the bias. Estimating $\text{ABias}\left(\widehat{\eta}^{\text{Hill}}_{d_t^\dagger}\right)$ relies on the use of scaled log-spacing representation of the Hill estimator as in Beirlant et al. (2002). Define scaled log-spacing as $Z_j = j(\log d_{t-j+1} - \log d_{t-j})$, where $j = 1, \dots, k_t$. The asymptotic bias can be estimated as[5]

$$\widehat{\text{ABias}}\left(\widehat{\eta}^{\text{Hill}}_{d_t^\dagger}\right) = \frac{6}{k_t}\sum_{j=1}^{k_t}\left(\frac{j}{k_t+1} - \frac{1}{2}\right) Z_j.$$

For the other tail estimators, there are no equivalent methods, but for comparison we apply $d_t^\dagger$ selected by this method to other tail estimators as well.

Clauset et al. (2009) proposes a universal method for any tail estimator: to choose $d_t^\dagger$ so that the distance, D, between the theoretical cumulative distribution of the underlying power-law, $F_{\widehat{\eta}}(d) = 1 - (d/d_t^\dagger)^{-1/\widehat{\eta}}$, with estimated $\widehat{\eta}$ and the empirical cumulative distribution,

[5] For more details on deriving this expression see Chapter 4.5.1 of Beirlant et al. (2006). There $\text{ABias}\left(\widehat{\eta}^{\text{Hill}}_{d_t^\dagger}\right) = \frac{b}{1+\beta}$ and the least-squares estimator for b is given on p. 117. Difficulties of estimating β are discussed on the same page and based on this discussion we set $\beta = 1$.

$\widehat{F}_{k_t}(d) = \sum_{i=d_t^\dagger+1}^{d} D_t(i)/k_t$, is minimized for all $d > d_t^\dagger$. The authors suggest using the Kolmogorov-Smirnov (KS) distance,[6] so that the distance is given by

$$D_{\text{KS}} = \max_{d > d_t^\dagger} \left| F_{\widehat{\eta}}(d) - \widehat{F}_{k_t}(d) \right| .$$

2.4 Simulations

Our simulations (see Supplement, Excel spreadsheet) show that the performance in terms of the mean square error of various tail estimators is substantially better when the tail cutoff is selected using the MS method rather than the KS method.

Comparing the performance of all tail estimators, we find that the Smith estimator outperforms all tail estimators. The Hill estimator is among the best performing tail estimators; it shows a substantial bias for $t = 1000$, which reduces with the number of observations. The continuity correction suggested by Clauset et al. (2009) slightly helps in reducing the bias. The NLS estimators perform better than the tail estimators only in small samples with at most 1000 observations. The introduced PML and GMM estimators outperform both the NLS and tail estimators.

References

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

Bauke, Heiko (2007) “Parameter Estimation for Power-law Distributions by Maximum Likelihood Methods,” *The European Physical Journal B*, **58** (2), pp. 167–173.

Beirlant, Jan, Petra Vynckier, and Jozef L Teugels (1996) “Tail Index Estimation, Pareto Quantile Plots Regression Diagnostics,” *Journal of the American Statistical Association*, **91** (436), pp. 1659–1667.

Beirlant, J, G Dierckx, A Guillou, and C Staăricaă (2002) “On Exponential Representations of Log-Spacings of Extreme Order statistics,” *Extremes*, **5** (2), pp. 157–180.

[6]The first application of the KS distance in this context dates back to Pickands (1975). Clauset et al. (2009) discussed other distances, such as the Cramer-von-Misses (CM) and Anderson-Darling (AD) distance and modified KS distance to penalize tails. We tried these distances, but the results did not change substantially in comparison to the KS distance.

Beirlant, Jan, Yuri Goegebeur, Johan Segers, and Jozef Teugels (2006) *Statistics of Extremes: Theory and Applications*: John Wiley & Sons.

Clauset, Aaron, Cosma Rohilla Shalizi, and Mark EJ Newman (2009) "Power-Law Distributions in Empirical Data," *SIAM Review*, **51** (4), pp. 661–703.

Dekkers, Arnold LM, John HJ Einmahl, and Laurens De Haan (1989) "A Moment Estimator for the Index of an Extreme-Value Distribution," *Annals of Statistics*, **17** (4), pp. 1833–1855.

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.

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.

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.

Matthys, Gunther and Jan Beirlant (2000) "Adaptive Threshold Selection in Tail Index Estimation," in Paul Embrechts ed. *Extremes and Integrated Risk Management*, pp. 37–49.

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

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," *Proceedings of the National Academy of Sciences*, **99** (8), pp. 5207–5211.

Pickands, III, James (1975) "Statistical Inference Using Extreme Order Statistics," *Annals of Statistics*, **3** (1), pp. 119–131.

Smith, Richard L. (1987) "Estimating Tails of Probability Distributions," *Annals of Statistics*, **15** (3), pp. 1174–1207.