Degree Distribution Redux

Econometric Methods for Networks,

SMU, May 29th & June 1st, 2017

Bryan S. Graham

University of California - Berkeley

Power Law Analysis

Following Barabási and Albert (1999), many researchers have found degree distributions, at least over some range, follow discrete Pareto or 'power law' distributions.

Specifically, the probability that a randomly sampled agent has Indegree d_+ is assumed to equal

$$\Pr\left(D_{+i} = d_+\right) = C d_+^{-\alpha}$$

with C the normalizing constant

$$C = \left[\sum_{j=0}^{\infty} \left(j + \underline{d}_+\right)^{-\alpha}\right]^{-1}$$

(i.e., inverse of the Hurwitz zeta function, $\zeta\left(\alpha, \underline{d}_+\right)$).

Power Law Analysis: Moments

The p^{th} moment of a random variable obeying a power law equals:

$$\mathbb{E}\left[D_{+i}^{p}\right] = \sum_{d_+=\underline{d}_+}^{\infty} d_+^p \,\mathsf{Pr}\left(D_{+i} = d_+\right) \simeq \lim_{y\to\infty} C \int_{\underline{d}_+}^{y} x^{p-\alpha}\mathsf{d}x$$

This integral converges if $p - \alpha + 1 \leq 0$ and diverges otherwise.

Therefore all moments which satisfy $p \leq \alpha - 1$ are finite...

...and all moments $p > \alpha - 1$ are infinite (sample moments will diverge as $N \to \infty$).

Power Law Analysis: Moments

Empirical evidence suggests that in many real world networks α lies between 2 and 3.

If accurate, this suggests we should observe greater variability in D_{+i} in larger networks.

In practice large networks do tend to have so-called ‘super hubs’.

Whether the power law description is accurate is mildly controversial.

Estimation

One approach to estimation of α is based upon the equality

$$\ln\left[\Pr\left(D_{+i} = d_{+}\right)\right] = \ln C - \alpha \ln d_{+}.$$

Which suggests an ordinary least squares approach (based upon estimates of $\ln\left[\Pr\left(D_{+i} = d_{+}\right)\right]$ for $i = 1 \ldots N$).

In practice this estimator works very poorly (cf., Gabaix, 2009).

Estimation

Recipe: Clauset, Shalizi, Newman (2009, *SIAM Review*) ... *over 5,000 Google Scholar citations!*

1. For a given $\underline{d}_+$ let $N_{\min} = \sum_{i=1}^{N} \mathbf{1}\left(D_{+i} \geq \underline{d}_+\right)$.

2. Estimate α by ML for discrete power law. This MLE is often well-approximated by the Hill (1975) estimate:

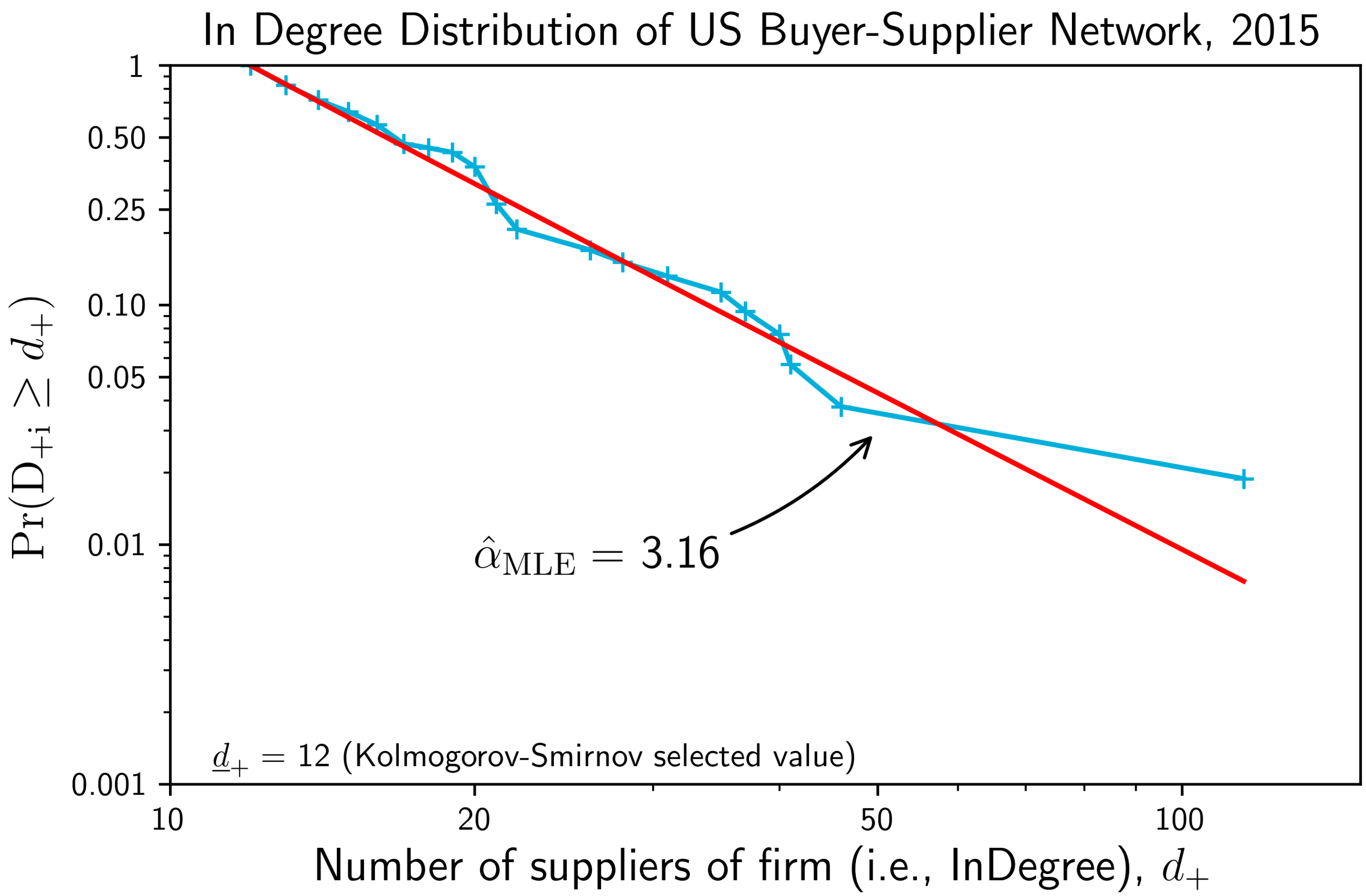
$$\widehat{\alpha} \simeq 1 + N_{\min}\left[\sum_{i\in\{D_{+i}\geq \underline{d}_+\}} \ln \frac{D_{+i}}{\underline{d}_+ - \frac{1}{2}}\right]^{-1}.$$

Estimation

3. Choose $\underline{d}_+$ to minimize the KS statistic

$$\max_{d_+ \geq \underline{d}_+} \left| \Pr\left(D_{+i} \leq d_+ \middle| D_{+i} \geq \underline{d}_+ \right) - P(d_+ \middle| \widehat{\alpha}, \underline{d}_+) \right|.$$

In this last step α is re-estimated for each possible value of $\underline{d}_+$ (i.e., Steps 1 and 2 above are repeated).



In Degree Distribution of US Buyer-Supplier Network, 2015
1
0.50
0.25
0.10
0.05
0.01
0.001
$\Pr(\mathrm{D}_{+\mathrm{i}} \geq d_+)$
$\hat{\alpha}_{\mathrm{MLE}} = 3.16$
$\underline{d}_+ = 12$ (Kolmogorov-Smirnov selected value)
10
20
50
100
Number of suppliers of firm (i.e., InDegree), d_+

Powerlaw Package

In Python the methods describe by Clauset, Shalizi and Newman (2009) have been implemented in the *powerlaw* package.

This package is described in Alstott, Bullmore and Plenz (2014, PLOS ONE).

This package was used to produce the figure shown above.

Inference?

Challenges to accurate inference:

- Likelihood derived under the assumption that $D_{+1}, \ldots, D_{+N_{\mathsf{min}}}$ are i.i.d. draws from discrete Pareto distribution (this is not true).

- Uncertainty associated with choosing/estimating $\underline{d}_+$ is not accounted for.

Wrap-Up

The powerlaw plot is a ubiquitous feature of empirical network analysis.

Appropriate inference procedures for $\widehat{\alpha}$ remains an open question.

Later we will connect (empirical) moments of the degree sequence to the (empirical) frequency of star subgraph configurations.