Degrees, Power Laws and Popularity

Gonzalo Mateos
Dept. of ECE and Goergen Institute for Data Science
University of Rochester
gmateosb@ece.rochester.edu
http://www.hajim.rochester.edu/ece/sites/gmateos/

January 31, 2023

Descriptive analysis of network characterstics

- Given a network graph representation of a complex system
  $\Rightarrow$ Structural properties of G key to system-level understanding

Example

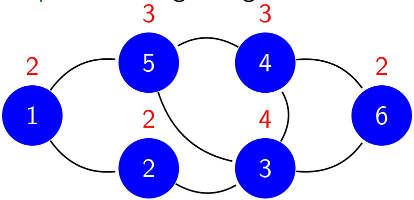
- Q1: Underpinning of various types of basic social dynamics?
  A: Study vertex triplets (triads) and patterns of ties among them
- Q2: How can we formalize the notion of 'importance' in a network?
  A: Define measures of individual vertex (or group) centrality
- Q3: Can we identify communities and cohesive subgroups?
  A: Formulate as a graph partitioning (clustering) problem

- Characterization of individual vertices/edges and network cohesion
  - Social network analysis, math, computer science, statistical physics

Degree

- **Def:** The degree d_v of vertex v is its number of incident edges

  $\Rightarrow$ Degree sequence arranges degrees in non-decreasing order



- In figure $\Rightarrow$ Vertex degrees shown in red, e.g., $d_1 = 2$ and $d_5 = 3$

  $\Rightarrow$ Graph's degree sequence is 2,2,2,3,3,4

- In general, the degree sequence does *not uniquely* specify the graph

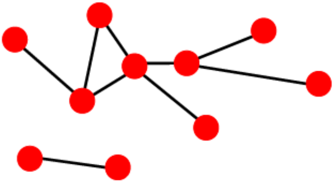
- High-degree vertices are likely to be influential, central, prominent

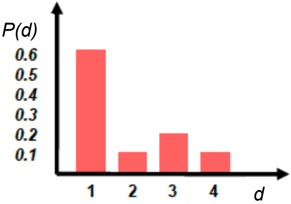
Degree distribution

- Let $N(d)$ denote the number of vertices with degree d

  $\Rightarrow$ Fraction of vertices with degree d is $\mathrm{P}(d) := \frac{N(d)}{N_v}$

- **Def:** The collection $\{\mathrm{P}(d)\}_{d \geq 0}$ is the degree distribution of G
  - Histogram formed from the degree sequence (bins of size one)



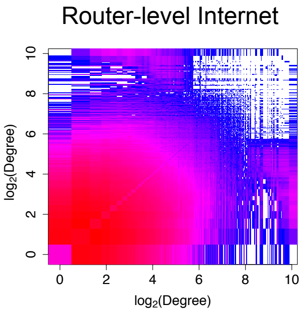


- $\mathrm{P}(d) =$ probability that randomly chosen node has degree d

  $\Rightarrow$ Summarizes the local connectivity in the network graph

Joint degree distribution

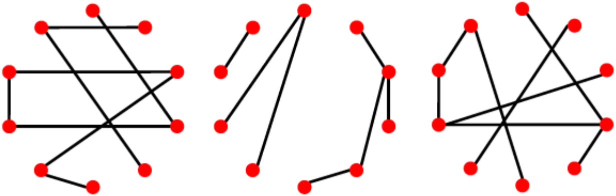
- Q: What about patterns of association among nodes of given degrees?
- A: Define the two-dimensional analogue of a degree distribution



- Prob. of random edge having incident vertices with degrees (d_1, d_2)

- **Def:** The Erdös-Renyi random graph model $G_{n,p}$
  - Undirected graph with n vertices, i.e., of order $N_v = n$
  - Edge (u, v) present with probability p, independent of other edges
- Simulation is easy: draw $\binom{n}{2}$ i.i.d. Bernoulli(p) RVs

Example



- Three realizations of $G_{10,\frac{1}{6}}$. The size N_e is a random variable

Degree distribution of $G_{n,p}$

- Q: Degree distribution $\mathrm{P}(d)$ of the Erdös-Renyi graph $G_{n,p}$?
- Define $\mathbb{I}\{(v,u)\} = 1$ if $(v,u) \in E$, and $\mathbb{I}\{(v,u)\} = 0$ otherwise.
  $\Rightarrow$ Fix v. For all $u \neq v$, the indicator RVs are i.i.d. Bernoulli(p)
- Let D_v be the (random) degree of vertex v. Hence,

$$D_v = \sum_{u \neq v} \mathbb{I}\{(v,u)\}$$

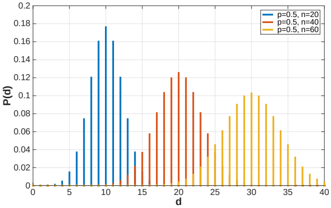
  $\Rightarrow$ D_v is binomial with parameters $(n-1, p)$ and

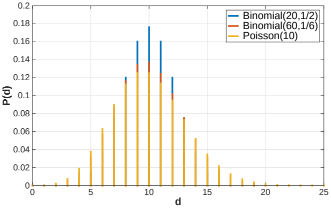
$$\mathrm{P}(d) = \mathrm{P}(D_v = d) = \binom{n-1}{d} p^d (1-p)^{(n-1)-d}$$

- In words, the probability of having exactly d edges incident to v
  $\Rightarrow$ Same for all $v \in V$, by independence of the $G_{n,p}$ model

- Q: How does the degree distribution look like for a large network?
- Recall D_v is a sum of $n-1$ i.i.d. Bernoulli(p) RVs

  $\Rightarrow$ Central Limit Theorem: $D_v \sim \mathcal{N}(np, np(1-p))$ for large n





- Makes most sense to increase n with fixed $\mathbb{E}[D_v] = (n-1)p = \mu$

  $\Rightarrow$ Law of rare events: $D_v \sim \text{Poisson}(\mu)$ for large n

Law of rare events

- Substituting $p = \mu/n$ in the binomial PMF yields

$$\mathrm{P}_n(d) = \frac{n!}{(n-d)!d!}\left(\frac{\mu}{n}\right)^d\left(1-\frac{\mu}{n}\right)^{n-d}$$
$$= \frac{n(n-1)\ldots(n-d+1)}{n^d}\,\frac{\mu^d}{d!}\,\frac{(1-\mu/n)^n}{(1-\mu/n)^d}$$

- In the limit, red term is $\lim\limits_{n\to\infty}(1-\mu/n)^n = e^{-\mu}$
- Black and blue terms converge to 1. Limit is the *Poisson* PMF

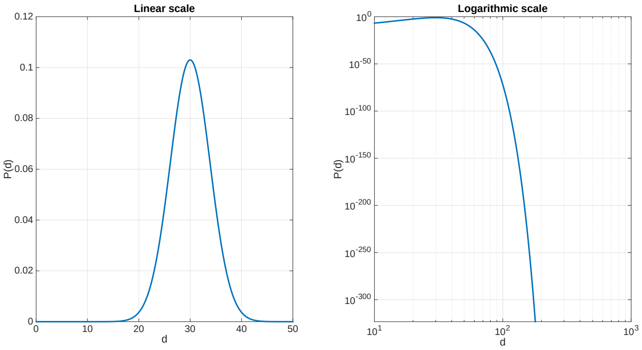
$$\lim_{n\to\infty}\mathrm{P}_n(d) = 1\,\frac{\mu^d}{d!}\,\frac{e^{-\mu}}{1} = e^{-\mu}\frac{\mu^d}{d!}$$

- Approximation usually called "law of rare events"
  - Individual edges happen with small probability $p = \mu/n$
  - The aggregate (degree, number of edges), though, need not be rare

The $G_{n,p}$ model and real-world networks

- For large graphs, $G_{n,p}$ suggests $\mathrm{P}(d)$ with an exponential tail

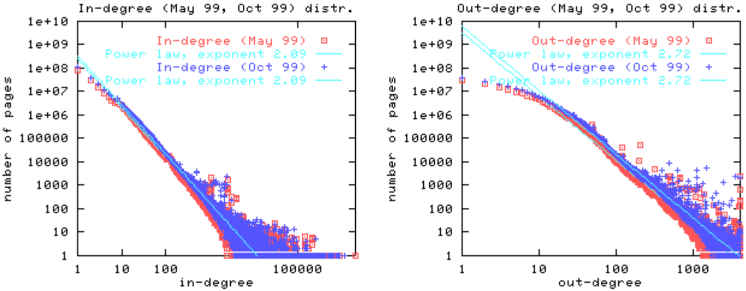
  $\Rightarrow$ Unlikely to see degrees spanning several orders of magnitude



- Concentrated distribution around the mean $\mathbb{E}[D_v] = (n-1)p$
- Q: Is this in agreement with real-world networks?

World Wide Web

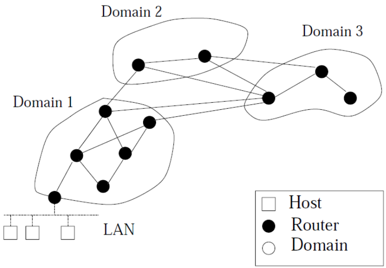
- Degree distributions of the WWW analyzed in [Broder et al '00]
  - $\Rightarrow$ Web a digraph, study both in- and out-degree distributions

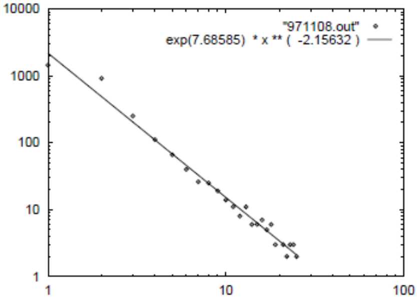


- Majority of vertices naturally have small degrees
  - $\Rightarrow$ Nontrivial amount with orders of magnitude higher degrees

Internet autonomous systems

- The topology of the AS-level Internet studied in [Faloutsos^3 '99]

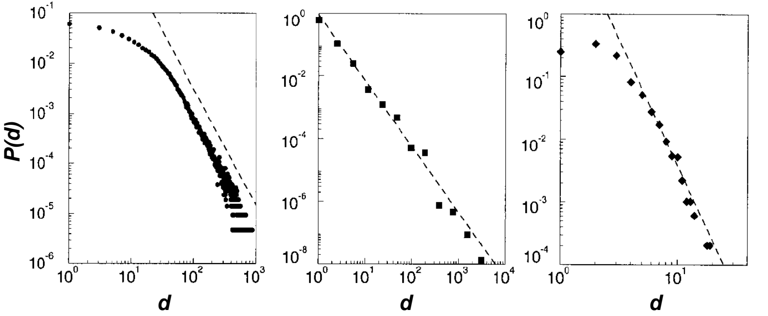




- Right-skewed degree distributions also found for router-level Internet

Seems to be a structural pattern

- More heavy-tailed degree distributions found in [Barabasi-Albert '99]



Author collaboration   Web graph   Power grid

- These heterogeneous, diffuse degree distributions are not exponential

Power laws

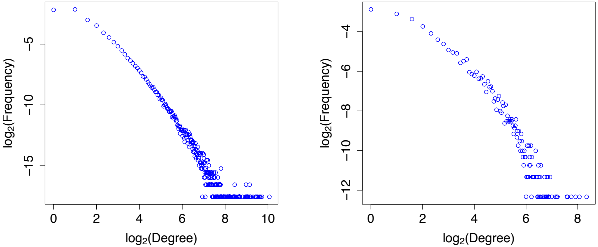
Degree distributions

Power-law degree distributions

Visualizing and fitting power laws

Popularity and preferential attachment

Power-law degree distributions



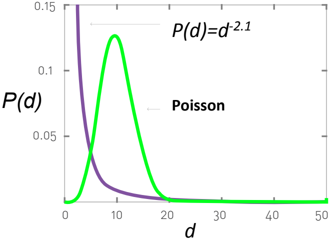
- Log-log plots show roughly a linear decay, suggesting the power law

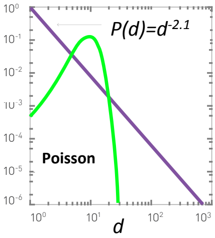
$$P(d) \propto d^{-\alpha} \Rightarrow \log P(d) = C - \alpha \log d$$

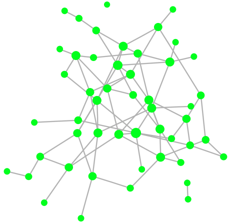
  - Power-law exponent (negative slope) is typically $\alpha \in [2, 3]$
  - Normalization constant C is mostly uninteresting

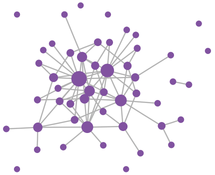
- Power laws often best followed in the tail, i.e., for $d \geq d_{\min}$

Power law and exponential degree distributions









- Erdös-Renyi's Poisson degree distribution exhibits a sharp cutoff

  $\Rightarrow$ Power laws upper bound exponential tails for large enough d

Scale-free networks

- Scale-free network: degree distribution with power-law tail
  - Name motivated for the scale-invariance property of power laws
- **Def:** A scale-free function $f(x)$ satisfies $f(ax) = bf(x)$, for $a, b \in \mathbb{R}$

Example

- Power-law functions $f(x) = x^{-\alpha}$ are scale-free since

$$f(ax) = (ax)^{-\alpha} = a^{-\alpha} f(x) = bf(x), \text{ where } b := a^{-\alpha}$$

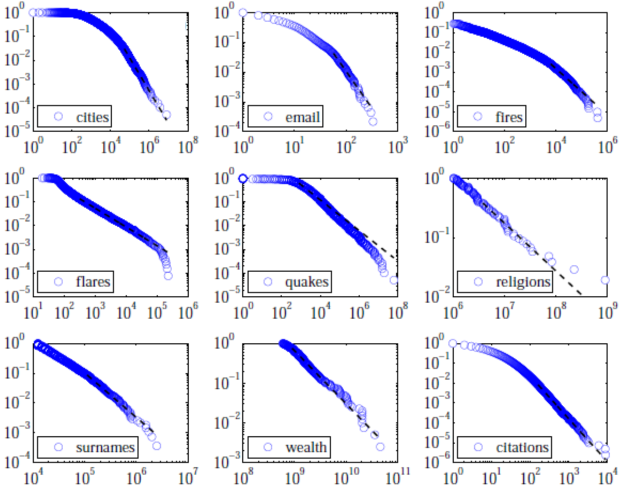
- Exponential functions $f(x) = c^x$ are not scale-free because

$$f(ax) = c^{ax} = (c^x)^a = f^a(x) \neq bf(x), \text{ except when } a = b = 1$$

- No 'characteristic scale' for the degrees. More soon

$\Rightarrow$ Functional form of the distribution is invariant to scale

Power-law distributions are ubiquitous



- Power-law distributions widespread beyond networks [Clauset et al '07]

Normalization

- The power-law degree distribution $\mathrm{P}(d) = Cd^{-\alpha}$ is a PMF, hence

$$1 = \sum_{d=0}^{\infty} \mathrm{P}(d) = \sum_{d=0}^{\infty} Cd^{-\alpha} \Rightarrow C = \frac{1}{\sum_{d=0}^{\infty} d^{-\alpha}}$$

- Often a power law is only valid for the tail $d \geq d_{\text{min}}$, hence

$$C = \frac{1}{\sum_{d=d_{\text{min}}}^{\infty} d^{-\alpha}} \approx \frac{1}{\int_{d_{\text{min}}}^{\infty} x^{-\alpha} dx} = (\alpha - 1) d_{\text{min}}^{\alpha-1}$$

$\Rightarrow$ Sound approximation since $\mathrm{P}(d)$ varies slowly for large d

- The normalized power-law degree distribution is

$$\mathrm{P}(d) = \frac{\alpha - 1}{d_{\text{min}}} \left(\frac{d}{d_{\text{min}}} \right)^{-\alpha}, \quad d \geq d_{\text{min}}$$

Power-law probability density function

- Often convenient to treat degrees as real valued, i.e., $d \in \mathbb{R}_+$
- Define a power-law PDF for the tail of the degree distribution as

$$p(d) = \frac{\alpha - 1}{d_{\min}} \left(\frac{d}{d_{\min}} \right)^{-\alpha}, \quad d \geq d_{\min}$$

$\Rightarrow$ A valid PDF, already showed that $\int_{d_{\min}}^{\infty} p(x)dx = 1$

$\Rightarrow$ Convergence of the integral requires $\alpha > 1$

- Ex: Probability that a random node has degree exceeding 100 is

$$\mathrm{P}\left(D_v > 100\right) = \int_{100}^{\infty} \frac{\alpha - 1}{d_{\min}} \left(\frac{x}{d_{\min}} \right)^{-\alpha} dx = \left(\frac{100}{d_{\min}} \right)^{1-\alpha}$$

Divergent moments

- Q: What is the m-th moment of a power-law distributed RV?
- From the definition of moment and the power-law PDF one has

$$\mathbb{E}\left[D_v^m\right] = \int_{d_{\min}}^{\infty} x^m p(x)dx = \frac{\alpha-1}{d_{\min}^{1-\alpha}}\left[\frac{x^{m+1-\alpha}}{m+1-\alpha}\right]_{d_{\min}}^{\infty}$$

$\Rightarrow$ Convergence of the integral requires $m+1<\alpha$

- For real-world networks, typically $\alpha \in (2,3)$ so

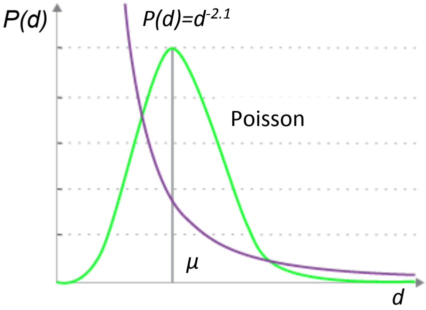
$$\mathbb{E}\left[D_v\right] = \left(\frac{\alpha-1}{\alpha-2}\right)d_{\min} < \infty \text{ and } \mathbb{E}\left[D_v^m\right] = \infty,\ m \geq 2$$

- In particular, the second moment and variance are infinite

$\Rightarrow$ Consistent with variability and heterogeneity of degrees

Revisiting the scale-free property

- A measure of scale of a RV is its standard deviation σ



Large random network $G_{n,p}$

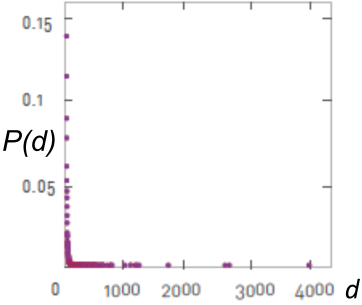
- Randomly chosen node has degree $d = \mu \pm \sqrt{\mu}$. The scale is μ

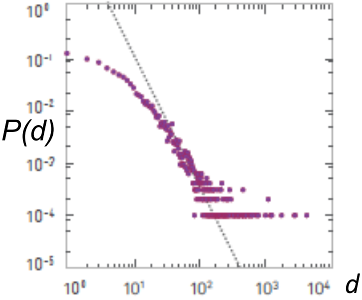
Scale-free network

- Randomly chosen node has degree $d = \mu \pm \infty$. There is no scale

Visualizing and fitting power laws

- A simple histogram may be problematic for visualizing $\mathrm{P}(d)$
  - $\Rightarrow$ Use log-log scale to warp probabilities and widespread degrees



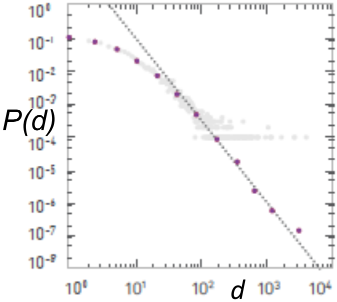


- Large statistical fluctuations ('noise') in the tail for large d
  - $\Rightarrow$ With bins of size one, high-degree counts are small
  - $\Rightarrow$ Makes sense to increase the bin size

Logarithmic binning

- Uniformly widening bins sacrifices resolution for small degrees

  $\Rightarrow$ Use bins of different sizes in different parts of the histogram



- Logarithmic binning is widely used. The n-th bin is

$$a^{n-1} \leq d < a^n, \quad n = 1, 2, \ldots$$

Ex: Common choice is $a = 2$, n-th bin has width $2^n - 2^{n-1} = 2^{n-1}$

- Normalize by the bin width. Wider bins will accrue higher counts

Complementary cumulative distribution function

- **Def:** The complementary cumulative distribution function (CCDF) is

$$\bar{F}(d) = \mathrm{P}\left(D_v \geq d\right)$$

$\Rightarrow$ Function $\bar{F}(d)$ is the fraction of vertices with degree at least d

- For a power-law PDF, the CCDF also obeys a power law since

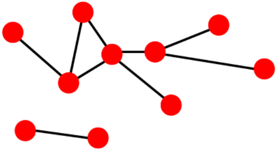
$$\mathrm{P}\left(D_v \geq d\right) = \int_d^\infty \frac{\alpha - 1}{d_{\min}} \left(\frac{x}{d_{\min}}\right)^{-\alpha} dx = \left(\frac{d}{d_{\min}}\right)^{-(\alpha-1)}$$

- If the PDF has exponent α, then CCDF $\bar{F}(d)$ has exponent $\alpha - 1$

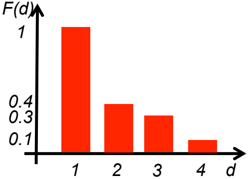
Step 1: List the degrees d_v in descending order

Step 2: Assign ranks r_v (from 1 to N_v) to vertices in that order

Step 3: The CCDF is the plot of r_v/N_v versus degree d_v



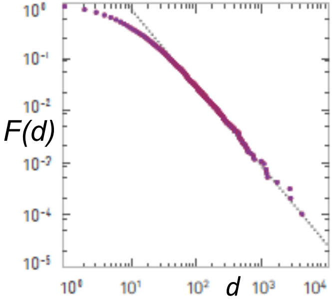
d_v	r_v	r_v/N_v	$F(d)$
4	1	0.1	0.1
3	2	0.2	
3	3	0.3	0.3
2	4	0.4	0.4
1	5	0.5	
1	6	0.6	
1	7	0.7	
1	8	0.8	
1	9	0.9	
1	10	1.0	1.0



- If degrees are repeated, CCDF is the largest value of r_v/N_v
- If d not observed, $\bar{F}(d)$ = value for next (larger) observed degree

Visualizing power laws with the CCDF

- Plot the CCDF in a log-log scale and look for a straight-line behavior



- Mitigates noise using cumulative frequencies (cf. raw frequencies)
- No binning needed $\Rightarrow$ Avoids information loss as bins widen

Fitting power-law distributions

- Basic, yet nontrivial task is to estimate the exponent α from data
- A power law implies the linear model $\log \mathrm{P}(d) = C - \alpha \log d + \epsilon$

  $\Rightarrow$ Natural to form the linear least-squares (LS) estimator

$$\{\hat{\alpha}, \hat{C}\} = \arg\min_{\alpha, C} \sum_i \left(\log \mathrm{P}(d_i) - C + \alpha \log d_i\right)^2$$

- Simple, very popular, but not advisable for at least three reasons:
  1) Extremely noisy high-degree data, where the counts are the lowest
  2) Estimates are biased. The log transform distorts unevenly the errors
  3) If the power law is only valid in the tail, need to hand pick $d_{\min}$

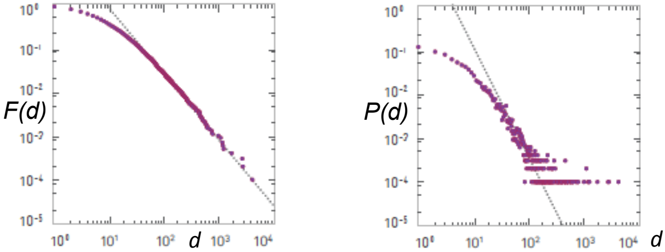
Linear regression inference on the CCDF

- A solution to the noise problem is to use the CCDF $\bar{F}(d)$

  $\Rightarrow$ Cumulative frequencies smoothen the noise

- Recall the CCDF follows a power law with exponent $\alpha - 1$

  $\Rightarrow$ Can use a linear regression-based approach to find $\hat{\alpha}$, but ...



- Successive points in the CCDF plot are not mutually independent

  $\Rightarrow$ (Ordinary) LS is not optimal for correlated errors

Maximum-likelihood estimator

- Suppose $\{d_i\}_{i=1}^{N_v}$ are independent and follow a power law. MLE of α?

  $\Rightarrow$ The data PDF is $f(d;\alpha) = \frac{\alpha-1}{d_{\min}}\left(\frac{d}{d_{\min}}\right)^{-\alpha}$, $d \geq d_{\min}$

- The log-likelihood function is (up to constants independent of α)

$$\ell_{N_v}(\alpha) = \sum_{i=1}^{N_v} \log f(d_i;\alpha) \propto N_v \log(\alpha-1) - \alpha \sum_{i=1}^{N_v} \log\left(\frac{d_i}{d_{\min}}\right)$$

- The MLE $\hat{\alpha}$ (a.k.a. Hill estimator) solves the equation

$$\left.\frac{\partial \ell_{N_v}(\alpha)}{\partial \alpha}\right|_{\alpha=\hat{\alpha}} = \frac{N_v}{\hat{\alpha}-1} - \sum_{i=1}^{N_v} \log\left(\frac{d_i}{d_{\min}}\right) = 0$$

- The solution is

$$\hat{\alpha} = 1 + \left[\frac{1}{N_v}\sum_{i=1}^{N_v} \log\left(\frac{d_i}{d_{\min}}\right)\right]^{-1}$$

Hill plot of ML estimates

- Q: How can we go around hand-picking the value of $d_{\min}$?

1) Rank-order degrees to obtain the sequence $d_{(1)} \leq \ldots \leq d_{(N_v)}$
2) For each $k \in \{1, \ldots, N_v - 1\}$ let $d_{\min} = d_{(N_v - k)}$. The MLEs are

$$\hat{\alpha}(k) = 1 + \left[\frac{1}{k} \sum_{i=0}^{k-1} \log \left(\frac{d_{(N_v - i)}}{d_{(N_v - k)}} \right) \right]^{-1}$$

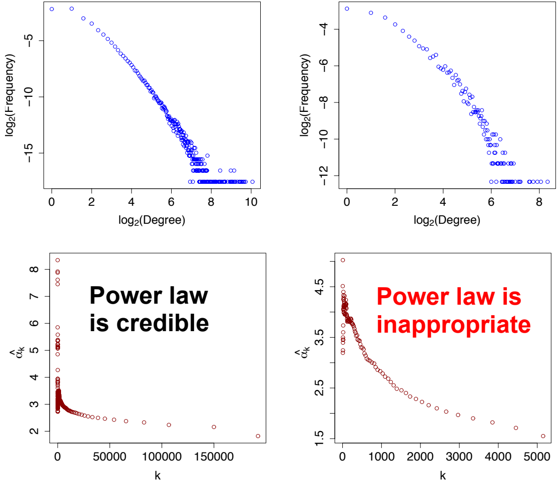
3) Draw and examine the Hill plot of $\hat{\alpha}(k)$ versus k

- If a power law is credible, the Hill plot should 'settle down'

  $\Rightarrow$ Identify stable $\hat{\alpha}$ for a wide range of (intermediate) k values

- Q: Why focus on values on the intermediate range?
  - Small k: Inaccurate estimation due to limited data
  - Large k: Bias if power law is only valid in the tail

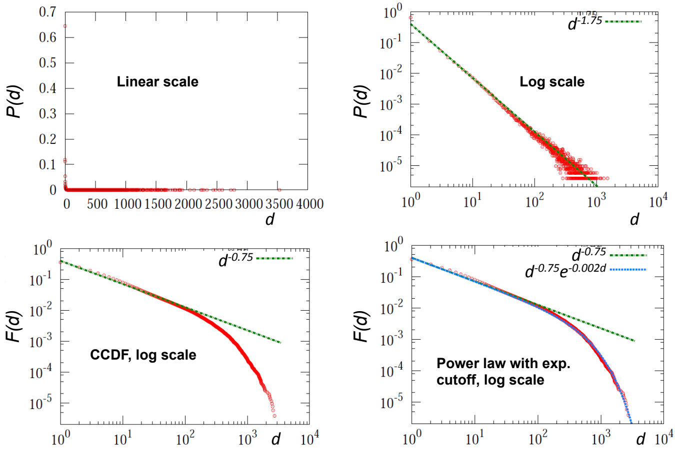
Example: Internet and protein interaction data



- Sharp decay in $\hat{\alpha}$ suggests a simple power-law model is inappropriate

Example: Flickr data

- Flickr social network: $N_v \approx 0.6M$, $N_e \approx 3.5M$ [Leskovec et al '08]



- Good fit to a power law with exponential cutoff $\bar{F}(d) \propto d^{-\alpha}e^{-\beta d}$

Popularity and preferential attachment

Degree distributions

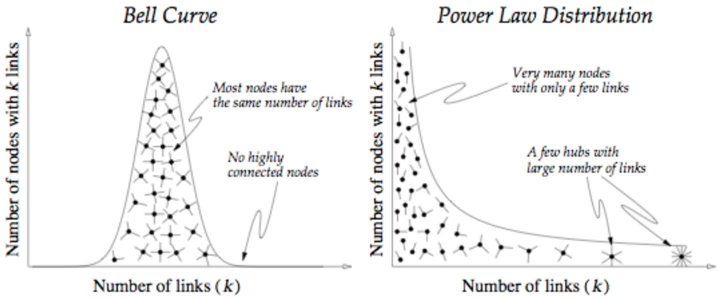
Power-law degree distributions

Visualizing and fitting power laws

Popularity and preferential attachment

Popularity as a network phenomenon

- Popularity is a phenomenon characterized by extreme imbalances
  - How can we quantify these imbalances? Why do they arise?



- Basic models of network behavior can be very insightful

  $\Rightarrow$ Result of coupled decisions, correlated behavior in a population

Preferential attachment model

- Simple model for the creation of e.g., links among Web pages
- Vertices are created one at a time, denoted $1, \ldots, N_v$
- When node j is created, it makes a single arc to i, $1 \leq i < j$
- Creation of (j, i) governed by a probabilistic rule:
  - With probability p, j links to i chosen uniformly at random
  - With probability $1 - p$, j links to i with probability $\propto d_i^{in}$
- The resulting graph is directed, each vertex has $d_v^{out} = 1$
- Preferential attachment model leads to "rich-gets-richer" dynamics

  $\Rightarrow$ Arcs formed preferentially to (currently) most popular nodes

  $\Rightarrow$ Prob. that i increases its popularity $\propto$ i's current popularity

Preferential attachment yields power laws

Theorem

The preferential attachment model gives rise to a power-law in-degree distribution with exponent $\alpha = 1 + \frac{1}{1-p}$, *i.e.,*

$$\mathrm{P}\left(d^{in} = d\right) \propto d^{-\left(1+\frac{1}{1-p}\right)}$$

- Key: "j links to i with probability $\propto d_i^{in}$" equivalent to copying, i.e., "j chooses k uniformly at random, and links to i if $(k, i) \in E$"
- Reflect: Copy other's decision vs. independent decisions in $G_{n,p}$
- As $p \to 0$ $\Rightarrow$ Copying more frequent $\Rightarrow$ Smaller $\alpha \to 2$
  - Intuitive: more likely to see extremely popular pages (heavier tail)

Continuous approximation

- In-degree $d_i^{in}(t)$ of node i at time $t \geq i$ is a RV. Two facts
  - **F1)** Initial condition: $d_i^{in}(i) = 0$ since node i just created at time $t = i$
  - **F2)** Dynamics of $d_i^{in}(t)$: Probability that new node $t+1 > i$ links to i is

$$\mathrm{P}\left((t+1, i) \in E\right) = p \times \frac{1}{t} + (1-p) \times \frac{d_i^{in}(t)}{t}$$

- Will study a deterministic, continuous approximation to the model
  - Continuous time $t \in [0, N_v]$
  - Continuous degrees $x_i^{in}(t) : [i, N_v] \mapsto \mathbb{R}_+$ are deterministic
- Require in-degrees to satisfy the following growth equation

$$\frac{dx_i^{in}(t)}{dt} = \frac{p}{t} + \frac{(1-p)x_i^{in}(t)}{t}, \quad x_i^{in}(i) = 0$$

Solving the differential equation

- Solve the first-order differential equation for $x_i^{in}(t)$ (let $q = 1 - p$)

$$\frac{dx_i^{in}}{dt} = \frac{p + qx_i^{in}}{t}$$

- Divide both sides by $p + qx_i^{in}(t)$ and integrate over t

$$\int \frac{1}{p + qx_i^{in}} \frac{dx_i^{in}}{dt} dt = \int \frac{1}{t} dt$$

- Solving the integrals, we obtain (c is a constant)

$$\ln\left(p + qx_i^{in}\right) = q \ln(t) + c$$

- Exponentiating and letting $K = e^c$ we find

$$\ln\left(p + q x_i^{in}(t)\right) = q \ln(t) + c \;\Rightarrow\; x_i^{in}(t) = \frac{1}{q}\left(K t^q - p\right)$$

- To determine the unknown constant K, use the initial condition

$$0 = x_i^{in}(i) = \frac{1}{q}\left(K i^q - p\right) \;\Rightarrow\; K = \frac{p}{i^q}$$

- Hence, the deterministic approximation of $d_i^{in}(t)$ evolves as

$$x_i^{in}(t) = \frac{1}{q}\left(\frac{p}{i^q} \times t^q - p\right) = \frac{p}{q}\left[\left(\frac{t}{i}\right)^q - 1\right]$$

Obtaining the degree distribution

- Q: At time t, what fraction $\bar{F}(d)$ of all nodes have in-degree $\geq d$?
  Approximation: What fraction of all functions $x_i^{in}(t) \geq d$ by time t?

$$x_i^{in}(t) = \frac{p}{q}\left[\left(\frac{t}{i}\right)^q - 1\right] \geq d$$

- Can be rewritten in terms of i as

$$i \leq t\left[\left(\frac{q}{p}\right)d + 1\right]^{-1/q}$$

- By time t there are exactly t nodes in the graph, so the fraction is

$$\bar{F}(d) = \left[\left(\frac{q}{p}\right)d + 1\right]^{-1/q} = 1 - F(d)$$

Identifying the power law

- The degree distribution is given by the PDF $p(d)$
- Recall that the PDF, CDF and CCDF are related, namely

$$p(x) = \frac{dF(x)}{dx} = -\frac{d\bar{F}(x)}{dx}$$

- Differentiating $\bar{F}(d) = \left[\left(\frac{q}{p}\right) d + 1\right]^{-1/q}$ yields

$$p(d) = \frac{1}{p}\left[\left(\frac{q}{p}\right) d + 1\right]^{-(1+\frac{1}{q})}$$

- Showed $p(d) \propto d^{-(1+1/q)}$, a power law with exponent $\alpha = 1 + \frac{1}{1-p}$
  - $\Rightarrow$ Disclaimer: Relied on heuristic arguments
  - $\Rightarrow$ Rigorous, probabilistic analysis possible

Glossary

- Degree distribution
- Erdös-Renyi model
- Binomial distribution
- Law of rare events
- Right-skewed distribution
- Logarithmic scale
- Power law
- Exponential and heavy tails
- Scale-free network
- Characteristic scale
- Logarithmic binning
- Cumulative frequencies
- Hill estimator and plot
- Exponential cutoff
- Coupled decisions
- Preferential attachment model
- Rich-gets-richer phenomena
- Growth equation