Network Formation Processes: Power-law degrees and Preferential Attachment
Next Time: New Topics

**Observations**
- Small diameter, Edge clustering
- Patterns of signed edge creation
- Viral Marketing, Blogosphere, Memetracking
- Scale-Free
- Densification power law, Shrinking diameters
- Strength of weak ties, Core-periphery

**Models**
- Erdös-Renyi model, Small-world model
- Structural balance, Theory of status
- Independent cascade model, Game theoretic model
- Preferential attachment, Copying model
- Microscopic model of evolving networks
- Kronecker Graphs

**Algorithms**
- Decentralized search
- Models for predicting edge signs
- Influence maximization, Outbreak detection, LIM
- PageRank, Hubs and authorities
- Link prediction, Supervised random walks
- Community detection: Girvan-Newman, Modularity
What do we observe that needs explaining?

- **Small-world model:**
  - Diameter
  - Clustering coefficient

- **Preferential Attachment:**
  - **Node degree distribution**
    - What fraction of nodes has degree $k$ (as a function of $k$)?
    - Prediction from simple random graph models: 
      $$p(k) = \text{exponential function of } k$$
    - **Observation:** Often a power-law: 
      $$p(k) \propto k^{-\alpha}$$
Degree Distributions

Expected based on $G_{np}$

$$P(k)$$

Found in data

$$P_{out}(k)$$

$$P(k) \propto k^{-\alpha}$$
Take a network, plot a histogram of $P(k)$ vs. $k$

- **Plot**: fraction of nodes with degree $k$:
  \[ p(k) = \frac{|\{u|d_u = k\}|}{N} \]

- **Flickr social network**: $n=584,207$, $m=3,555,115$
Plot the same data on log-log scale:

How to distinguish:

\[ P(k) \propto \exp(-k) \text{ vs. } P(k) \propto k^{-\alpha} \]?

Take logarithms:

if \( y = f(x) = e^{-x} \) then

\[ \log(y) = -x \]

If \( y = x^{-\alpha} \) then

\[ \log(y) = -\alpha \log(x) \]

So on log-log axis power-law looks like a straight line of slope \(-\alpha\)!
Node Degrees: Faloutsos

- Internet Autonomous Systems
  [Faloutsos, Faloutsos and Faloutsos, 1999]

Internet domain topology
Node Degrees: Web

- The World Wide Web [Broder et al., 2000]
Other Networks [Barabasi-Albert, 1999]
Above a certain $x$ value, the power law is always higher than the exponential!
Power-law vs. Exponential on log-log and semi-log (log-lin) scales

- $p(x) = cx^{-0.5}$ (log-log)
- $p(x) = cx^{-1}$ (log-log)
- $p(x) = c^{-x}$ (semi-log)

$X \ldots$ logarithmic axis
$Y \ldots$ logarithmic axis

$X \ldots$ linear
$Y \ldots$ logarithmic
Exponential vs. Power-Law

Bell Curve:
- Most nodes have the same number of links
- No highly connected nodes

Power Law Distribution:
- Very many nodes with only a few links
- A few hubs with large number of links
Power-Law Degree Exponents

- **Power-law degree exponent is typically** $2 < \alpha < 3$
  - **Web graph:**
    - $\alpha_{in} = 2.1$, $\alpha_{out} = 2.4$ [Broder et al. 00]
  - **Autonomous systems:**
    - $\alpha = 2.4$ [Faloutsos, 99]
  - **Actor-collaborations:**
    - $\alpha = 2.3$ [Barabasi-Albert 00]
  - **Citations to papers:**
    - $\alpha \approx 3$ [Redner 98]
  - **Online social networks:**
    - $\alpha \approx 2$ [Leskovec et al. 07]
Scale-Free Networks

- **Definition:**
  Networks with a power-law tail in their degree distribution are called “scale-free networks”

- **Where does the name come from?**
  - **Scale invariance:** There is no characteristic scale
    - Scale invariance is that laws do not change if scales of length, energy, or other variables, are multiplied by a common factor
  - **Scale-free function:** $f(ax) = a^\lambda f(x)$
    - Power-law function: $f(ax) = a^\lambda x^\lambda = a^\lambda f(x)$

  
  Log() or Exp() are not scale free!
  
  $f(ax) = \log(ax) = \log(a) + \log(x) = \log(a) + f(x)$
  
  $f(ax) = \exp(ax) = \exp(x)^a = f(x)^a$
Power-Laws are Everywhere

Many other quantities follow heavy-tailed distributions

[Clauset-Shalizi-Newman 2007]
Anatomy of the Long Tail

Online services carry far more inventory than traditional retailers. Rhapsody, for example, offers 19 times as many songs as Wal-Mart's stock of 39,000 tunes. The appetite for Rhapsody's more obscure tunes (charted below in yellow) makes up the so-called Long Tail. Meanwhile, even as consumers flock to mainstream books, music, and films (right), there is real demand for niche fare found only online.

Sources: Erik Brynjolfsson and Jeffrey Hu, MIT, and Michael Smith, Carnegie Mellon; Barnes & Noble; Netflix; RealNetworks

[Chris Anderson, Wired, 2004]
Not Everyone Likes Power-Laws 😊

CMU grad-students at the G20 meeting in Pittsburgh in Sept 2009
Mathematics of Power-Laws
Degrees are heavily skewed:

Distribution $P(X > x)$ is heavy tailed if:

$$\lim_{x \to \infty} \frac{P(X > x)}{e^{-\lambda x}} = \infty$$

Note:

- Normal PDF: $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Exponential PDF: $p(x) = \lambda e^{-\lambda x}$
- then $P(X > x) = 1 - P(X \leq x) = e^{-\lambda x}$ are not heavy tailed!
Various names, kinds and forms:

- Long tail, Heavy tail, Zipf’s law, Pareto’s law

Heavy tailed distributions:

- $P(x)$ is proportional to:

  \[
  P(x) \propto x^{-\alpha} \\
  x^{-\alpha} e^{-\lambda x} \\
  x^{\beta - 1} e^{-\lambda x^\beta} \\
  \frac{1}{x} \exp \left[ - \frac{(\ln x - \mu)^2}{2\sigma^2} \right]
  \]
What is the normalizing constant?

\[ p(x) = Z x^{-\alpha} \quad Z = ? \]

- \( p(x) \) is a distribution: \( \int p(x) \, dx = 1 \)

\[ 1 = \int_{x_m}^{\infty} p(x) \, dx = Z \int_{x_m}^{\infty} x^{-\alpha} \, dx \]

\[ = - \frac{Z}{\alpha-1} \left[ x^{-\alpha+1} \right]_{x_m}^{\infty} = - \frac{Z}{\alpha-1} \left[ \infty^{1-\alpha} - x_m^{1-\alpha} \right] \]

\[ \Rightarrow Z = (\alpha - 1) x_m^{\alpha-1} \]

\[ p(x) = \frac{\alpha - 1}{x_m} \left( \frac{x}{x_m} \right)^{-\alpha} \]

\[ \int (ax)^n = \frac{(ax)^{n+1}}{a(n+1)} \]
What’s the expected value of a power-law random variable $X$?

$$E[X] = \int_{x_m}^{\infty} x \, p(x) \, dx = Z \int_{x_m}^{\infty} x^{-\alpha+1} \, dx$$

$$= \frac{Z}{2-\alpha} \left[ x^{2-\alpha} \right]_{x_m}^{\infty} = \frac{(\alpha - 1)x_m^{\alpha - 1}}{-(\alpha - 2)} \left[ \infty^{2-\alpha} - x_m^{2-\alpha} \right]$$

$$\Rightarrow E[X] = \frac{\alpha - 1}{\alpha - 2} x_m$$

Need: $\alpha > 2$!

Power-law density:

$$p(x) = \frac{\alpha - 1}{x_m} \left( \frac{x}{x_m} \right)^{-\alpha}$$

$$Z = \frac{\alpha - 1}{x_m^{1-\alpha}}$$
Mathematics of Power-Laws

- **Power-laws have infinite moments!**
  \[ E[X] = \frac{\alpha - 1}{\alpha - 2} x_m \]
  - If \( \alpha \leq 2 \) : \( E[X] = \infty \)
  - If \( \alpha \leq 3 \) : \( \text{Var}[X] = \infty \)
    - Average is meaningless, as the variance is too high!

- **Consequence:** Sample average of \( n \) samples from a power-law with exponent \( \alpha \)

In real networks \( 2 < \alpha < 3 \) so:
\[ E[X] = \text{const} \]
\[ \text{Var}[X] = \infty \]
Estimating Power-law Exponent Alpha
Estimating Power-Law Exponent $\alpha$

Estimating $\alpha$ from data:

- **(1)** Fit a line on log-log axis using least squares:
  - Solve $\arg\min_{\alpha} (\log(y) - \alpha \log(x) + b)^2$
Estimating $\alpha$ from data:

- Plot Complementary CDF (CCDF) $P(X \geq x)$. Then the estimated $\alpha = 1 + \alpha'$ where $\alpha'$ is the slope of $P(X \geq x)$.

- **Fact:** If $p(x) = P(X = x) \propto x^{-\alpha}$ then $P(X \geq x) \propto x^{-(\alpha-1)}$

  - $P(X \geq x) = \sum_{j=x}^{\infty} p(j) \approx \int_{x}^{\infty} Z y^{-\alpha} dy = $
  - $= \frac{Z}{1-\alpha} [y^{1-\alpha}]_{x}^{\infty} = \frac{Z}{1-\alpha} x^{-(\alpha-1)}$
Estimating Power-Law Exponent $\alpha$

Estimating $\alpha$ from data:

- **Use maximum likelihood approach:**
  
  - The log-likelihood of observed data $d_i$:
    
    \[
    L(\alpha) = \ln(\prod_{i}^{n} p(d_i)) = \sum_{i}^{n} \ln p(d_i)
    \]
    
    \[
    = \sum_{i}^{n} \left( \ln(\alpha - 1) - \ln(x_m) - \alpha \ln \left( \frac{d_i}{x_m} \right) \right)
    \]

  - Want to find $\alpha$ that $\max L(\alpha)$: Set $\frac{dL(\alpha)}{d\alpha} = 0$
    
    \[
    \frac{dL(\alpha)}{d\alpha} = 0 \Rightarrow \frac{n}{\alpha - 1} - \sum_{i}^{n} \ln \left( \frac{d_i}{x_m} \right) = 0
    \]
    
    \[
    \Rightarrow \hat{\alpha} = 1 + n \left[ \sum_{i}^{n} \ln \left( \frac{d_i}{x_m} \right) \right]^{-1}
    \]

  - Power-law density:
    
    \[
    p(x) = \frac{\alpha - 1}{x_m} \left( \frac{x}{x_m} \right)^{-\alpha}
    \]
Flickr: Fitting Degree Exponent

Linear scale

Log scale, $\alpha=1.75$

CCDF, Log scale, $\alpha=1.75$

CCDF, Log scale, $\alpha=1.75$, exp. cutoff
Why are Power-Laws Surprising

- **Can not arise from sums of independent events!**
  - **Recall:** in $G_{np}$ each pair of nodes in connected independently with prob. $p$
    - $X$... degree of node $v$
    - $X_w$ ... event that $w$ links to $v$
    - $X = \sum_w X_w$
    - $E[X] = \sum_w E[X_w] = (n - 1)p$

- **Now, what is $P(X = k)$? Central limit theorem!**
  - $X_1, ..., X_n$: random vars with mean $\mu$, variance $\sigma^2$
  - $S_n = \sum_i X_i$: $E[S_n] = n\mu$, $\text{Var}[S_n] = n\sigma^2$, $\text{SD}[S_n] = \sigma\sqrt{n}$
  - $P(S_n = E[S_n] + x \cdot \text{SD}[S_n]) \sim \frac{1}{2\pi} e^{-\frac{x^2}{2}}$
Random network
(Erdos-Renyi random graph)

Degree distribution is Binomial

Scale-free (power-law) network

Degree distribution is Power-law

Consequence of Power-Law Degrees
How does network connectivity change as nodes get removed? [Albert et al. 00; Palmer et al. 01]

Nodes can be removed:

- Random failure:
  - Remove nodes uniformly at random

- Targeted attack:
  - Remove nodes in order of decreasing degree

This is important for robustness of the internet as well as epidemiology.
Network Resilience

- Networks with equal number of nodes and edges:
  - ER random graph
  - Scale-free network
- Study the properties of the network as an increasing fraction of nodes are removed
  - Node selection:
    - Random (this corresponds to random failures)
    - Nodes with largest degrees (corresponds to targeted attacks)
- Measures:
  - Fraction of nodes in the largest connected component
  - Average shortest path length between nodes in the largest component
Scale-free graphs are resilient to random attacks, but sensitive to targeted attacks:
What proportion of random nodes must be removed in order for the size \( (S) \) of the giant component to drop to 0?

- Infinite scale-free networks with \( \gamma < 3 \) never break down under random node failures

Source: Cohen et al., Resilience of the Internet to Random Breakdowns
- **Real networks are resilient to** random failures
- **$G_{np}$** has better resilience to targeted attacks
  - E.g., we need to remove all pages of degree >5 to disconnect the Web. But this is a very small fraction of all web pages!
Resilience in Real Networks

Source: Error and attack tolerance of complex networks. Réka Albert, Hawoong Jeong and Albert-László Barabási
The first few % of nodes removed:
- E: $G_{np}$
- SF: Scale-free

Notice how targeted attacks very quickly disconnect the network.
 Preferential Attachment Model
Model: Preferential attachment

- **Preferential attachment**
  [Price ‘65, Albert-Barabasi ‘99, Mitzenmacher ‘03]
  - Nodes arrive in order $1, 2, \ldots, n$
  - At step $j$, let $d_i$ be the degree of node $i < j$
  - A new node $j$ arrives and creates $m$ out-links
  - Prob. of $j$ linking to a previous node $i$ is proportional to degree $d_i$ of node $i$

\[ P(j \rightarrow i) = \frac{d_i}{\sum_{k} d_k} \]
Rich Get Richer

- New nodes are more likely to link to nodes that already have high degree

- Herbert Simon’s result:
  - Power-laws arise from “Rich get richer” (cumulative advantage)

- Examples
  - **Citations** [de Solla Price ‘65]: New citations to a paper are proportional to the number it already has
    - **Herding**: If a lot of people cite a paper, then it must be good, and therefore I should cite it too
  - **Sociology**: Matthew effect, [http://en.wikipedia.org/wiki/Matthew_effect](http://en.wikipedia.org/wiki/Matthew_effect)
    - “For whoever has will be given more, and they will have an abundance. Whoever does not have, even what they have will be taken from them.”
    - Eminent scientists often get more credit than a comparatively unknown researcher, even if their work is similar
We will analyze the following model:

- Nodes arrive in order 1, 2, 3, ..., \( n \)
- When node \( j \) is created it makes a **single out-link** to an earlier node \( i \) chosen:
  1) With prob. \( p \), \( j \) links to \( i \) chosen **uniformly at random** (from among all earlier nodes)
  2) With prob. \( 1 - p \), node \( j \) chooses \( i \) uniformly at random & links to a random node \( l \) that \( i \) points to
    - **This is same as saying:** With prob. \( 1 - p \), node \( j \) links to node \( l \) with prob. proportional to \( d_l \) (the in-degree of \( l \))
- **Our graph is directed:** Every node has out-degree 1
**Claim:** The described model generates networks where the fraction of nodes with in-degree $k$ scales as:

$$P(d_i = k) \propto k^{-(1+\frac{1}{q})}$$

where $q=1-p$

So we get power-law degree distribution with exponent:

$$\alpha = 1 + \frac{1}{1-p}$$
Continuous Approximation

- Consider deterministic and continuous approximation to the degree of node $i$ as a function of time $t$
  - $t$ is the number of nodes that have arrived so far
  - In-Degree $d_i(t)$ of node $i$ ($i = 1, 2, \ldots, n$) is a continuous quantity and it grows deterministically as a function of time $t$

- Plan: Analyze $d_i(t)$ – continuous in-degree of node $i$ at time $t$ ($t > i$)
  - Note: Node $i$ arrives to the graph at time $i$
Continuous Degree: What We Know

- **Initial condition:**
  - \( d_i(t) = 0 \), when \( t = i \) (node \( i \) just arrived)

- **Expected change of \( d_i(t) \) over time:**
  - Node \( i \) gains an in-link at step \( t + 1 \) only if a link from a newly created node \( t + 1 \) points to it

- **What’s the probability of this event?**
  - With prob. \( p \) node \( t + 1 \) links randomly:
    - Links to our node \( i \) with prob. \( 1/t \)
  - With prob. \( 1 – p \) node \( t + 1 \) links preferentially:
    - Links to our node \( i \) with prob. \( d_i(t)/t \)

- Prob. node \( t + 1 \) links to \( i \) is:
  \[
  p \frac{1}{t} + (1 – p) \frac{d_i(t)}{t}
  \]

Note: each node creates exactly 1 edge. So after \( t \) nodes/steps there are \( t \) edges in total.
At $t = 4$ node $i = 4$ comes. It has out-degree of 1 to deterministically share with other nodes:

<table>
<thead>
<tr>
<th>Node $i$</th>
<th>$d_i(t)$</th>
<th>$d_i(t+1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>$=0 + p \frac{1}{4} + (1 - p) \frac{0}{4}$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>$=2 + p \frac{1}{4} + (1 - p) \frac{2}{4}$</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>$=0 + p \frac{1}{4} + (1 - p) \frac{1}{4}$</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>$=1 + p \frac{1}{4} + (1 - p) \frac{1}{4}$</td>
</tr>
<tr>
<td>4</td>
<td>/</td>
<td>0</td>
</tr>
</tbody>
</table>

$d_i(t) - d_i(t - 1) = \frac{dd_i(t)}{dt} = p \frac{1}{t} + (1 - p) \frac{d_i(t)}{t}$

How does $d_i(t)$ evolve as $t \to \infty$?
What is the rate of growth of $d_i$?

- **Expected change of $d_i(t)$:**

  - $d_i(t + 1) - d_i(t) = p \frac{1}{t} + (1 - p) \frac{d_i(t)}{t}$

  - \[
  \frac{dd_i(t)}{dt} = p \frac{1}{t} + (1 - p) \frac{d_i(t)}{t} = \frac{p + q d_i(t)}{t} \]

  - \[
  \frac{1}{p + q d_i(t)} \frac{dd_i(t)}{dt} = \frac{1}{t} dt
  \]

  - \[
  \int \frac{1}{p + q d_i(t)} dd_i(t) = \int \frac{1}{t} dt
  \]

  - \[
  \frac{1}{q} \ln(p + q d_i(t)) = \ln t + c
  \]

  - \[
  p + q d_i(t) = e^{qc} t^q \quad \Rightarrow \quad d_i(t) = \frac{1}{q} ((At)^q - p)
  \]
What is the value of constant A?

- **We know:** \( d_i(i) = 0 \)

- **So:** \( d_i(i) = \frac{1}{q} ((Ai)^q - p) = 0 \)

- \( \Rightarrow A = \frac{p}{iq} \)

- And so \( \Rightarrow d_i(t) = \frac{p}{q} \left( \left( \frac{t}{i} \right)^q - 1 \right) \)

**Observation:** Old nodes (small i values) have higher in-degrees \( d_i(t) \)