Scale-free Networks

Work of Albert-László Barabási, Réka Albert, and Hawoong Jeong
Presented by Jonathan Fink

February 13, 2007

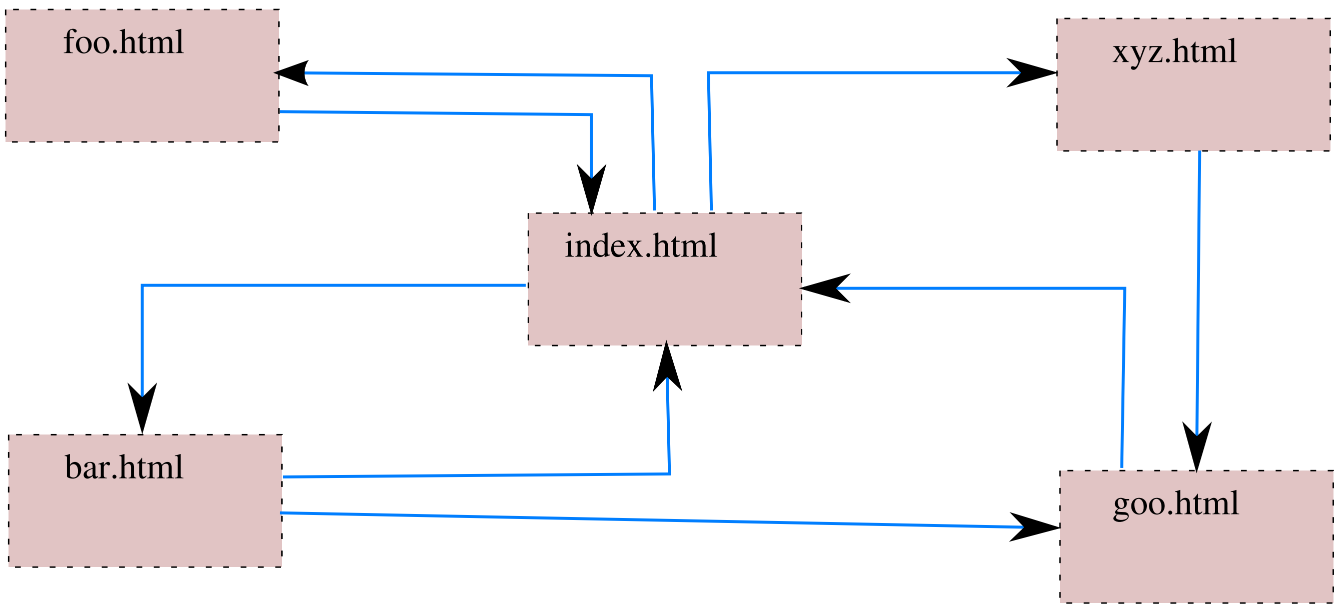
Outline

- Case study: The WWW
- Barabási-Albert Scale-free model
- Does it predict measured properties of the WWW?

Case Study: The WWW

Consider a graph where ...

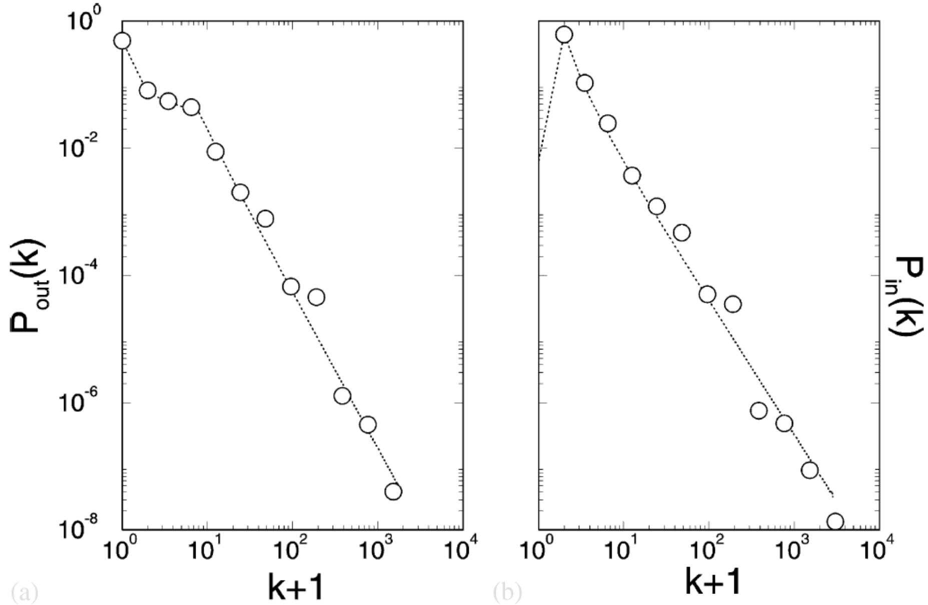
- HTML documents are nodes
- Links between HTML documents are edges



We are interested in average vertex connectivity as graph grows very large

Empirical measurements of the WWW

- Automated agent reads web-pages (in the nd.edu domain)
- Recursively follows links and builds graph of system
- Calculate link probabilities $P_{\text{out}}(k)$ and $P_{\text{in}}(k)$



$$P_{\text{out}}(k) \sim k^{\gamma_{\text{out}}} \quad P_{\text{in}}(k) \sim k^{\gamma_{\text{in}}}$$

ℓ: smallest number of links from one node to another

ℓ is computed by constructing a random graph with

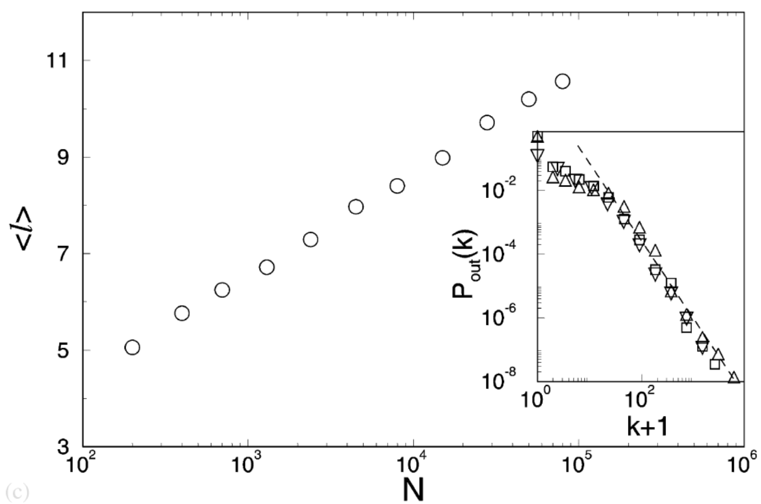
- N vertices
- k outgoing links from each vertex where k is drawn from empirically determined power-law distribution $P_{\text{out}}(k)$
- Destinations determined randomly, $P_{\text{in}}(k)$ satisfied for each vertex
- Compute $< \ell >$ to be the average across all vertex pairs

ℓ: smallest number of links from one node to another

ℓ is computed by constructing a random graph with

- N vertices
- k outgoing links from each vertex where k is drawn from empirically determined power-law distribution $P_{\text{out}}(k)$
- Destinations determined randomly, $P_{\text{in}}(k)$ satisfied for each vertex
- Compute $<\ell>$ to be the average across all vertex pairs

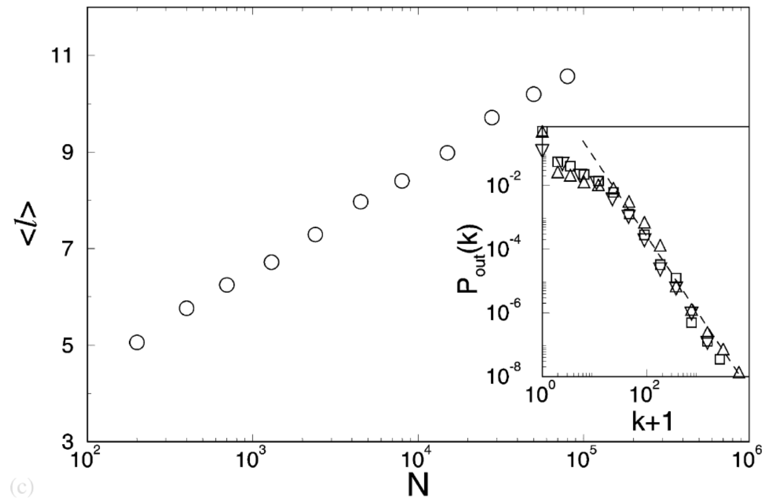
$$<\ell> \quad = \quad 0.35 + 2.06\log(N)$$



ℓ: smallest number of links from one node to another

ℓ is computed by constructing a random graph with

- N vertices
- k outgoing links from each vertex where k is drawn from empirically determined power-law distribution $P_{\text{out}}(k)$
- Destinations determined randomly, $P_{\text{in}}(k)$ satisfied for each vertex
- Compute $< \ell >$ to be the average across all vertex pairs



$$
\begin{aligned}
< \ell > &= 0.35 + 2.06 \log(N) \\
< \ell_{\text{nd.edu}} > &= 11.2 \quad \text{(data)} \\
< \ell_{3\times10^5} > &= 11.6 \quad \text{(model)} \\
< \ell_{8\times10^8} > &= 18.59 \quad \text{(2000 www)} \\
< \ell_{1.15\times10^{10}} > &= 21.07 \quad \text{(2005 www)}
\end{aligned}
$$

Earlier models

- Erdös-Rényi (ER) and Watts-Strogatz (WS) models predict $P(k)$ with exponential decay
- Assume networks with a fixed number of nodes N
- Links are random and uniform

Barabási-Albert Scale-free Model

Earlier models

- Erdös-Rényi (ER) and Watts-Strogatz (WS) models predict $P(k)$ with exponential decay
- Assume networks with a fixed number of nodes N
- Links are random and uniform

Barabási-Albert Model

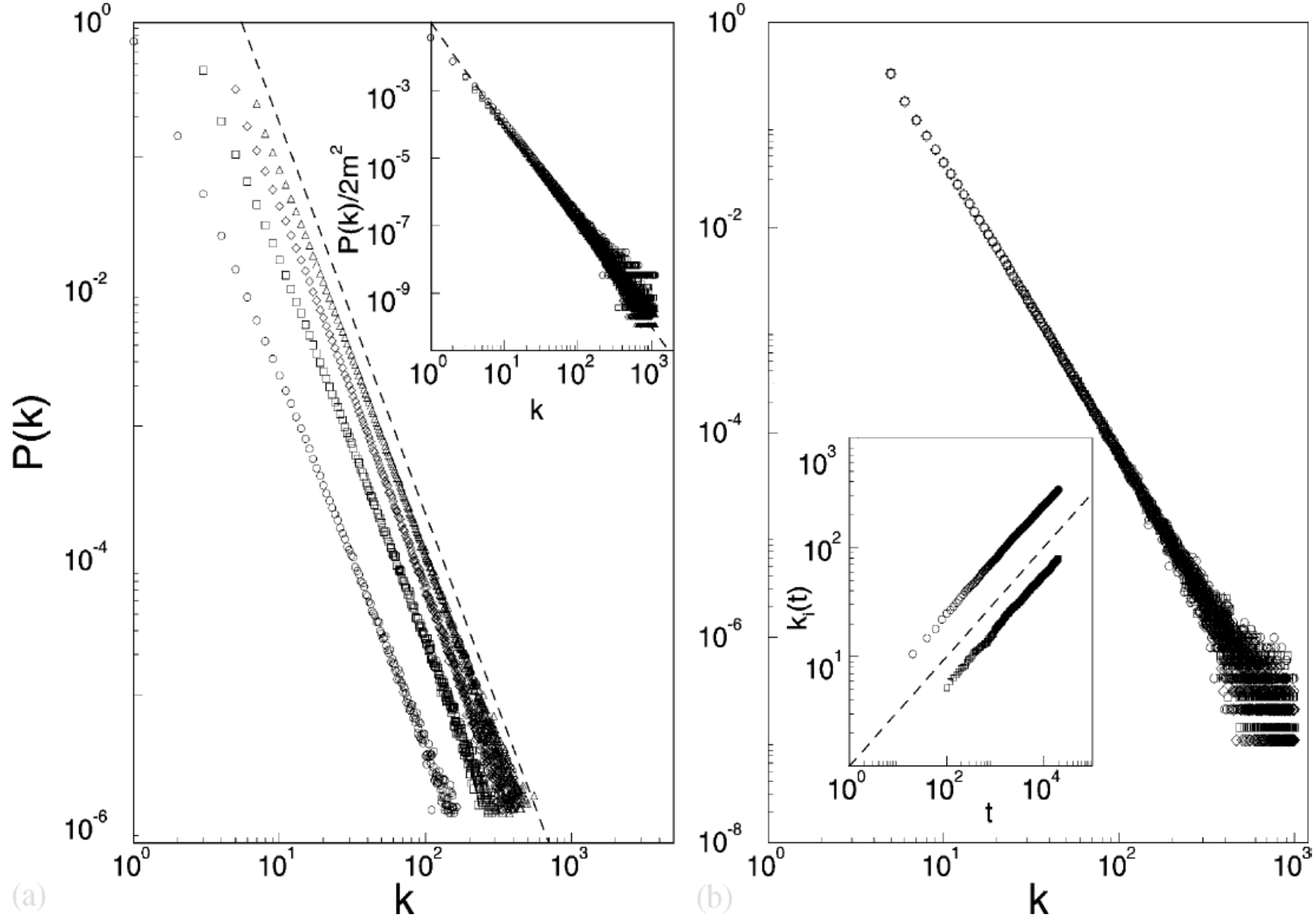
- *Growth*: Initialize with m_0 vertices and add a new vertex at every timestep with m edges to nodes already in the system ($m \leq m_0$)
- *Preferential Attachment*: Links for a new vertex are chosen according to

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

so that links to highly connected nodes are more likely.

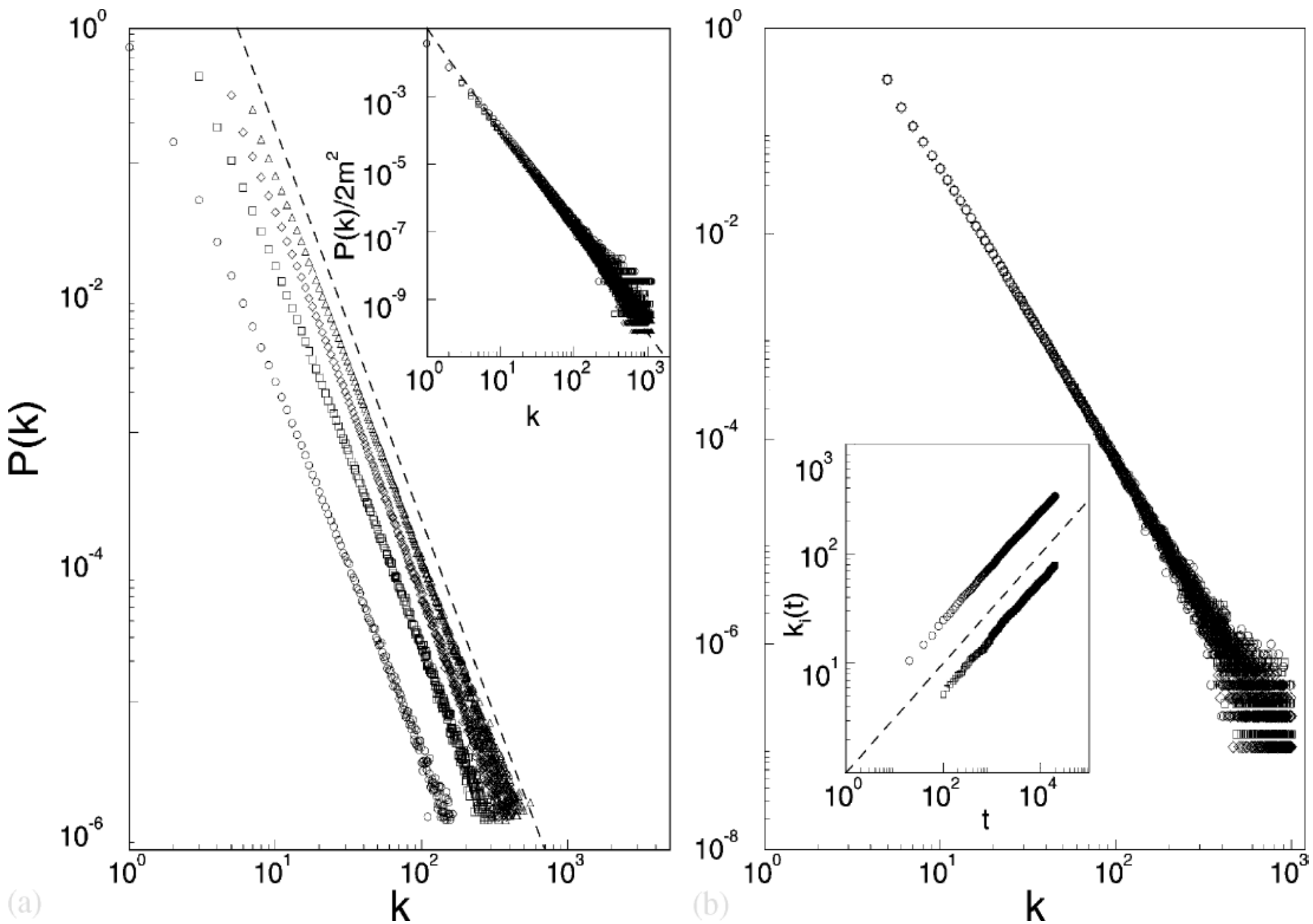
Scale-free Stationary State - Numeric

- $N = t + m_0$ vertices
- mt edges



Scale-free Stationary State - Numeric

- $N = t + m_0$ vertices
- mt edges
- $\gamma_{\text{model}} = 2.9 \pm 0.1$
- scale-free stationary state



Connectivity of vertex i

- Vertex i aquires edges at a rate $\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}$ with boundary condition $k(t_i) = m$.
- Solution: $k_i(t) = m(\frac{t}{t_i})^{0.5}$

Connectivity of vertex i

- Vertex i aquires edges at a rate $\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}$ with boundary condition $k(t_i) = m$.
- Solution: $k_i(t) = m(\frac{t}{t_i})^{0.5}$

Connectivity distribution

$$\begin{aligned} P[k_i(t) < k] = & \quad P[m(t/t_i)^{1/2} < k] \\ = & \quad P[t_i > (m/k)^2 t] \\ = & \quad 1 - P[t_i \leq (m/k)^2 t] \\ = & \quad 1 - \frac{m^2 t}{k^2(t+m_0)} \end{aligned}$$

Then,

$$P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \left(\frac{2m^2 t}{m_0 + t}\right)\left(\frac{1}{k^3}\right) \rightarrow 2m^2 k^{-3}$$

Which yields a stationary power-law distribution with $\gamma = 3$

Do we need Growth & Preferential Attachment?

- *Growth* Only
  - $\Pi(k) = 1/(m_0 + t - 1)$
  - $P(k) \sim \exp(-\beta k)$
- *Preferential Attachment* Only
  - Start with N vertices, no edges
  - $\Pi(k_i) = k_i / \sum_j k_j$
  - P(k) is not stationary - eventually all vertices are connected
- **Both** *Growth* and *Preferential Attachment* are necessary

Matching WWW with Barabási-Albert Model

Discrepancies

- $\gamma_{\text{model}} = 3$
- $\gamma_{\text{out}} = 2.45$, $\gamma_{\text{in}} = 2.1$

Matching WWW with Barabási-Albert Model

Discrepancies

- $\gamma_{\text{model}} = 3$
- $\gamma_{\text{out}} = 2.45$, $\gamma_{\text{in}} = 2.1$

Possible Reasons

- Graph is not simply growing
- Non-unique model (vertex ordering affects γ)
- Model of $\Pi(k)$ may be wrong
- Rewiring of links competes with growth

Summary

Barabási-Albert model is one step closer to representing systems like the WWW that have a power-law distribution on links.

In general, there may be several possible absorbing states:

- Scale-free state - growth sufficiently exceeds link reattachment
- Fully connected state - probability of adding links is too high
- Ripened state - popular vertices get *all* connections