Simple rules govern finite-size effects in scale-free networks

Sara Cuenda and Juan A. Crespo

Dpto. Economía Cuantitativa
Facultad CC. Económicas y Empresariales
Universidad Autónoma de Madrid

GISC Workshop, February 2011

Scale-free networks.

Wikipedia:

- "A scale-free network is a network whose **degree distribution follows a power law, at least asymptotically**."

$$f(k) \sim ck^{-\gamma}$$

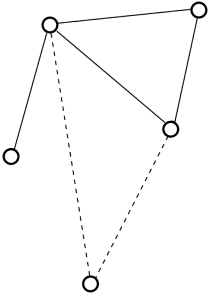
- "Scale-free networks are noteworthy because many empirically observed networks appear to be scale-free."

Most important model of scale-free networks:

- Barabási-Albert model (the first!).
- Most other models are based on BA.

Scale-free networks.

Barabási, Albert and Jeong, *Phys. A* **272**, p. 173 (1999).



- **Growth:** starting with t_0 nodes, at every time step we add a new node with m undirected new links.
- **Preferential attachment:** the probability that a node gets a new link is

$$\pi_{k,t} = \frac{mk}{\sum_i k_i} = \frac{k}{2(\mu t_0 + t)}$$

- **Mean-field approximation:** deterministic, continuum approximation in t and k so that

$$\frac{\partial k}{\partial t} = \pi_{k,t} \rightarrow k(t,\tau) = m\sqrt{\frac{\mu t_0 + t}{\mu t_0 + \tau}}$$

$\rightarrow$ For fixed t, k decreases with τ.
$\rightarrow$ Older nodes have higher degree.

Scale-free networks.

Barabási, Albert and Jeong, *Phys. A* **272**, p. 173 (1999).

- **Complementary cumulative degree distribution:**

$$F(k^*) = P(k \geq k^*) = \left(\frac{m}{k^*}\right)^2 \frac{\mu t_0 + t}{t} - \frac{\mu t_0}{t}$$

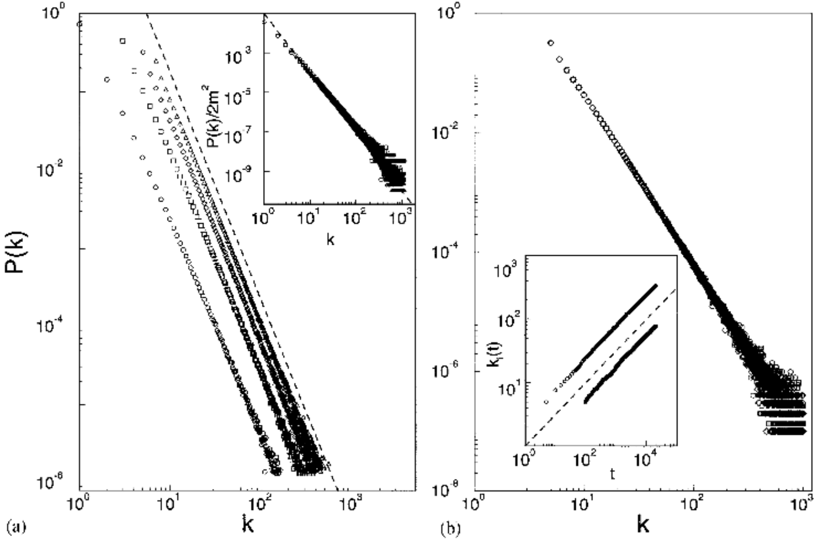
For $\frac{\mu t_0}{t} \simeq 0$:

$$F(k) \simeq \frac{\mu t_0 + t}{t} \left(\frac{m}{k}\right)^2$$

- **Power-law:** in $F(k)$ and $f(k)$! $\rightarrow$ scale-free.

Scale-free networks.

Barabási, Albert and Jeong, *Phys. A* **272**, p. 173 (1999).



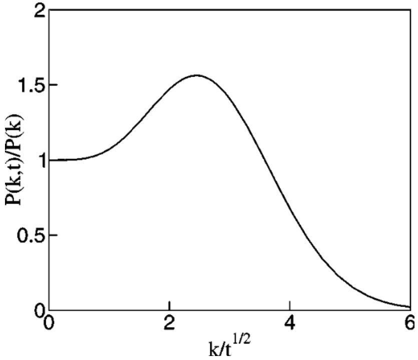
Finite-size effect.

However, further investigations showed:

- Dogorovtsev, Mendes and Samukhin, *Phys. Rev. E* **63**, p. 62101 (2001).
  For $t \gg k \gg 1$, the probability distribution verifies:

$$f(k,t) = f(k, t \to \infty) g(k/\sqrt{t})$$

Numerical simulations

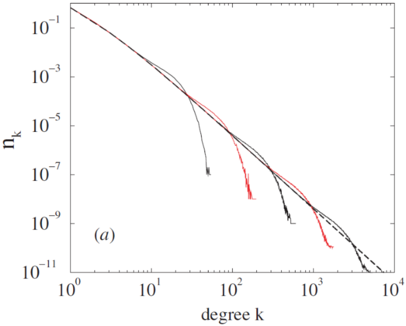


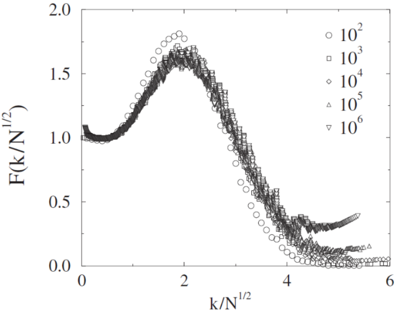
- Cut-off in the degree distribution.
- $g(k/\sqrt{t})$: cut-off function.
- The form of the hump depends on the initial configuration.

Finite-size effect.

However, further investigations showed:

- Kaprivsky and Redner, *J. Phys. A: Math. Gen.* **35**, p. 9517 (2002).



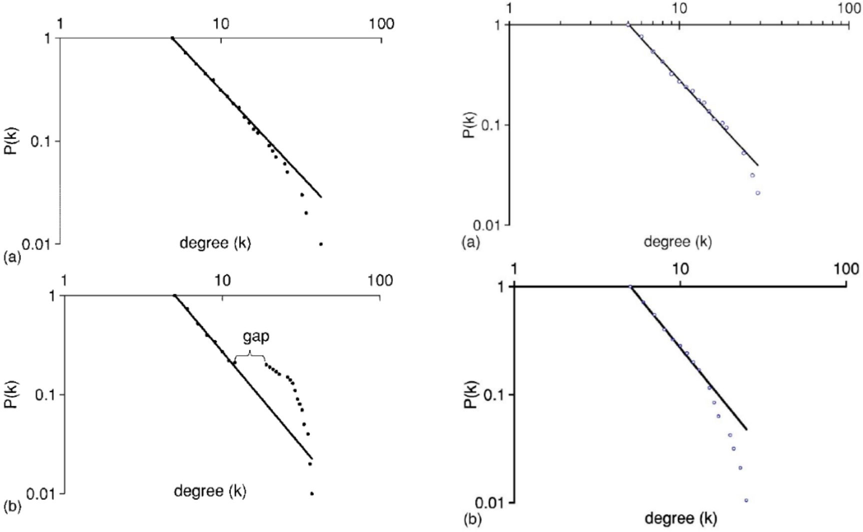


Exact resolution for $t_0 = 2$, $m = 1$.

Finite-size effect.

However, further investigations showed:

- Guimarães et al., *Phys. Rev. E* **71**, p. 37101 (2005).



Finite-size effect.

However, further investigations showed:

- Dogorovtsev, Mendes and Samukhin, *Phys. Rev. E* **63**, p. 62101 (2001).
- Kaprivsky and Redner, *J. Phys. A: Math. Gen.* **35**, p. 9517 (2002).
- Guimarães et al., *Phys. Rev. E* **71**, p. 37101 (2005).
- Waclaw and Sokolov, *Phys. Rev. E* **75**, p. 56114 (2007).
- Hassan, Hassan and Pavel, *arXiv:1101.4730v1* (2011).

"Mean-field is very rough: let's do something more accurate."

But no general explanation and/or prediction found!

Proper counting: Considering the initial nodes.

- **Mean-field approximation:**

  It may be rough, but please use it gently:

  $$k(t,\tau) = m\sqrt{\frac{\mu t_0 + t}{\mu t_0 + \tau}} \longrightarrow k(t,\tau,\kappa) = \kappa\sqrt{\frac{\mu t_0 + t}{\mu t_0 + \tau}}$$

  Value of κ:
  - For the added nodes: $\kappa = m$, $\tau > t_0$.
  - For the initial nodes: κ is a random variable, with degree distribution $F_0(\kappa)$, and $\tau = t_0$.

  Constructing $F(k,t)$:
  - Nodes that have degree k at time t, had degree $k_0(k,t) = k\sqrt{\frac{\mu t_0 + t_0}{\mu t_0 + t}}$ at time t_0.
  - Higher degree of an added node is $k_m(t) = m\sqrt{\frac{\mu t_0 + t}{\mu t_0 + t_0}}$.
  - Consider different ranges of the degree to calculate $F(k,t)$.

Proper counting: Considering the initial nodes.

- **Mean-field approximation:**

  Expression of $F(k,t)$:

$$F(k,t) = \begin{cases} 1 - \frac{t_0}{t}[1 - F_0(k_0(k,t))], & k \leq m \\ \frac{\mu t_0 + t}{t}\left(\frac{m}{k}\right)^2 - \frac{\mu t_0}{t} \\ \quad - \frac{t_0}{t}[1 - F_0(k_0(k,t))], & m < k \leq k_m(t) \\ \frac{t_0}{t} F_0(k_0(k,t)), & k > k_m(t) \end{cases}$$

  Cut-off function:

$$\frac{f(k,t)}{f(k,t\to\infty)} = \begin{cases} 1 + w\left(k\sqrt{\frac{\mu t_0 + t_0}{\mu t_0 + t}}\right), & k \leq k_m(t) \\ w\left(k\sqrt{\frac{\mu t_0 + t_0}{\mu t_0 + t}}\right), & k > k_m(t) \end{cases}$$

  where $w(x) = \frac{1}{2m^2(\mu+1)} x^3 f_0(x)$.

Proper counting: Considering the initial nodes.

- **Mean-field approximation:**

  Expression of $F(k,t)$:

$$F(k,t) = \begin{cases} 1 - \frac{t_0}{t}[1 - F_0(k_0(k,t))], & k \le m \\ \frac{\mu t_0 + t}{t}\left(\frac{m}{k}\right)^2 - \frac{\mu t_0}{t} \\ \quad -\frac{t_0}{t}[1 - F_0(k_0(k,t))], & m < k \le k_m(t) \\ \frac{t_0}{t} F_0(k_0(k,t)), & k > k_m(t) \end{cases}$$

  Cut-off function:

$$\frac{f(k,t)}{f(k,t\to\infty)} = \begin{cases} 1 + w\left(k\sqrt{\frac{\mu t_0 + t_0}{\mu t_0 + t}}\right), & k \le k_m(t) \\ w\left(k\sqrt{\frac{\mu t_0 + t_0}{\mu t_0 + t}}\right), & k > k_m(t) \end{cases}$$

  where $w(x) = \frac{1}{2m^2(\mu+1)} x^3 f_0(x)$.

  It is deterministic (there is no dispersion) :-(

Proper counting: Considering the initial nodes.

- **"Exact" calculation:**

$\pi_{k,t}$: probability of gaining a new link, independent from other nodes.

$$\pi_{k,t} = \frac{mk}{\sum_i k_i}$$

$P_{t,k}(\tau, \kappa)$: probability of a node having degree k at time t if it had degree κ at time τ.

$$P_{t,k} = P_{t-1,k}(1 - \pi_{k,t-1}) + P_{t-1,k-1}\pi_{k-1,t-1}.$$

$$P_{\tau,\kappa} = 1.$$

(You don't want to know the analytical expression.)

Proper counting: Considering the initial nodes.

- **"Exact" calculation:**

  Expected probability distribution of the degree of the network (pdf):

$$f(k,t) = \sum_{\kappa=0}^{k} \frac{t_0}{t} f_0(\kappa) P_{t,k}(t_0,\kappa) + \sum_{\tau=t_0+1}^{t} \frac{1}{t} P_{t,k}(\tau,m).$$

  Expected complementary cumulative degree distribution of the degree of the network (ccdf):

$$F(k,t) = \sum_{k^* \geq k} f(k^*,t)$$

Simulation of the BA model.

How to simulate the stochastic process with $\pi_{k,t} = \frac{mk}{\sum_i k_i}$?

- Linear in k.
- No correlation: $\pi_{k,t}$ only depends on the state of the node.

Equations that $\pi_{k,t}$ must satisfy:

- Expected number of total new links:

$$\sum_j \pi_{k_j,t} = m$$

- The sum of the probabilities of all possible configurations:

$$\sum_{i_1=1}^{t} \sum_{i_2<i_1} \cdots \sum_{i_m<i_{m-1}} \pi_{k_{i_1},t}\pi_{k_{i_2},t} \cdots \pi_{k_{i_m},t} = 1$$

Simulation of the BA model.

How to simulate the stochastic process with $\pi_{k,t} = \frac{mk}{\sum_i k_i}$?

- Linear in k.
- No correlation: $\pi_{k,t}$ only depends on the state of the node.

Equations that $\pi_{k,t}$ must satisfy:

- Expected number of total new links:

$$\sum_j \pi_{k_j,t} = m$$

- The sum of the probabilities of all possible configurations:

$$\sum_{i_1=1}^{t} \sum_{i_2<i_1} \cdots \sum_{i_m<i_{m-1}} \pi_{k_{i_1},t} \pi_{k_{i_2},t} \cdots \pi_{k_{i_m},t} = 1$$

True just for $m = 1$!!!

Simulation of the BA model.

Take a simulation algorithm that is well approximated by the model.

Algorithm:

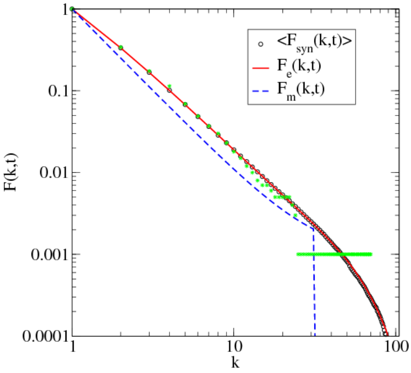
- m nodes chosen sequentially.
- Each one with probability $k / \sum_i k_i$.
- Avoid repetition.

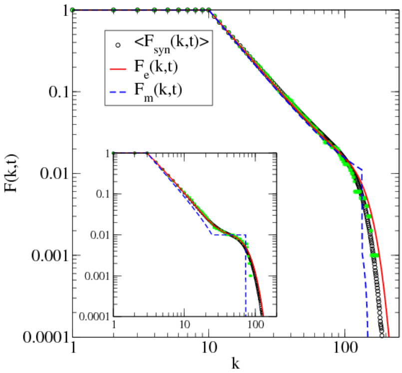
Model:

- Exact for $m = 1$.
- Exact for a delta distribution (all nodes with the same degree).
- Good approximation for $t \gg m$.
- High correlation for $t \gtrsim m$ (e.g., $\pi_{k,m}$ should be 1).

Simulation of the BA model.

Complementary, cumulative distribution functions:



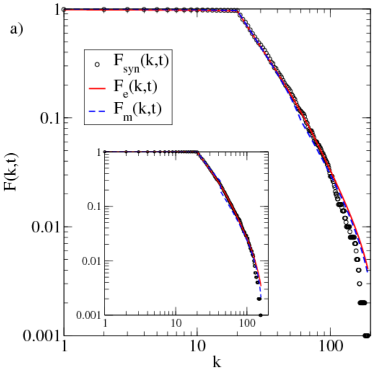


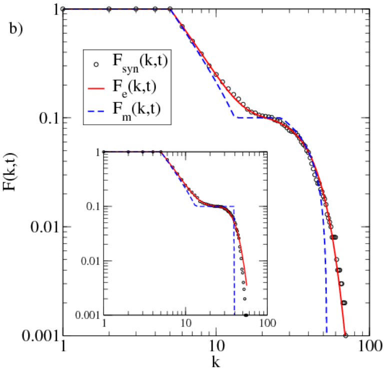
- $t_0 = 2$, $\delta_{1,k}$, $m = 1$, $t = 1000$.
- Notice the dispersion for large k.

- $t_0 = 10$, $\delta_{9,k}$, $m = 10$, $t = 1000$.
- The same, with $m = 3$.

Simulation of the BA model.

Complementary, cumulative distribution functions:





- $t_0 = 100$, $U(0,5)$, $m = 20$, $t = 1000$.
- inset: the same, for $\delta_{5,k}$.

- $t_0 = 100$, $U(10,20)$, $m = 5$, $t = 1000$.
- inset: the same, with $\delta_{15,k}$.

Conclusions

- The finite-size effect comes from the initial nodes in the network, and from the dispersion of the degree of nodes.
- The mean-field approximation is rough but it predicts a finite-size effect in the final degree distribution of nodes and explains the results of the cut-off function $g(k/\sqrt{t})$.
- The expected final degree distribution of the network calculated from the probability distribution of a single node is very accurate within the limits of the BA model.
- This methodology can be used for other uncorrelated models.
- For networks with $t \gtrsim m$ there is correlation, not considered by the BA model. However, as the network grows, the problem softens.
- Do you think now that this model is scale-free? Would you use it to model data that is scale-free?

THAN K YOU

(now let's go and have a drink!)