

# Redes Complexas: Internet, Web e outras aplicações em computação

-Bloco #6-

2º semestre de 2008

“Scale Free Networks”

Virgílio A. F. Almeida

Outubro de 2008



Departamento de Ciência da Computação  
Universidade Federal de Minas Gerais

# Projeto: datasets disponiveis

- <http://www-personal.umich.edu/~mejn/netdata/>
- <http://tangra.si.umich.edu/clair/anthology>
- <http://www.theyrule.net/2004/tr2.php>
- <http://www.orgnet.com/SN.html>
- Datasets mentioned in [Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations](#), Jure Leskovec, Jon Kleinberg and Christos Faloutsos. ACM August 2005.  
Datasets: arXiv, Patents, Autonomous Systems, Affiliation network
- Patent Data: <http://iv.slis.indiana.edu/db/patents.html>
- arXiv Citation Data: <http://www.cs.cornell.edu/projects/kddcup/datasets.html>
- Enron Email Dataset: <http://www.cs.cmu.edu/~enron/>
- World-Wide-Web, Actor, Cellular, and Protein Interaction Data: <http://www.nd.edu/~networks/resources.htm>
- URV Email, Jazz musicians, PGP users, and C. Elegans Metabolic data: <http://deim.urv.cat/~aarenas/data/welcome.htm>
- Zachary's karate club: <http://www-personal.umich.edu/~mejn/netdata/karate.zip>

# Projetos: datasets disponiveis

- adjectives/nouns from David Copperfield <http://www-personal.umich.edu/~mejn/netdata/adjnoun.zip>
- American College Football <http://www-personal.umich.edu/~mejn/netdata/football.zip>
- Dolphin Social Network <http://www-personal.umich.edu/~mejn/netdata/dolphins.zip>
- Political Blogs <http://www-personal.umich.edu/~mejn/netdata/polblogs.zip>
- Power Grid <http://www-personal.umich.edu/~mejn/netdata/power.zip>
- Neural Networks of C. Elegans <http://www-personal.umich.edu/~mejn/netdata/celegansneural.zip>
- Condensed Matter Collaborations 2005 <http://www-personal.umich.edu/~mejn/netdata/cond-mat-2005.zip>
- Astrophysics Collaborations <http://www-personal.umich.edu/~mejn/netdata/astro-ph.zip>
- High-energy theory collaborations <http://www-personal.umich.edu/~mejn/netdata/hep-th.zip>
- Coauthorships in network science <http://www-personal.umich.edu/~mejn/netdata/netscience.zip>

# Projetos: datasets disponiveis

- Internet <http://www-personal.umich.edu/~mejn/netdata/as-22july06.zip>
- languages  
[http://www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/darwinBookInter\\_st.txt](http://www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/darwinBookInter_st.txt)  
[http://www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/frenchBookInter\\_st.txt](http://www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/frenchBookInter_st.txt),  
[http://www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/spanishBookInter\\_st.txt](http://www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/spanishBookInter_st.txt)  
[http://www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/japaneseBookInter\\_st.txt](http://www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/japaneseBookInter_st.txt)
- power grid [http://cdg.columbia.edu/uploads/datasets/power\\_unweighted](http://cdg.columbia.edu/uploads/datasets/power_unweighted)
- citations <http://www.cs.cornell.edu/projects/kddcup/download/hep-th-citations.tar.gz>
- comic book characters <http://bioinfo.uib.es/~joemiro/marvel/porgat.txt>
- Subset of the Maple Blog collection  
<http://tangra.si.umich.edu/clair/allnets/R1000/R1000-small.net>
- blog title [http://tangra.si.umich.edu/clair/allnets/blogtitles/cheney\\_all-small.net](http://tangra.si.umich.edu/clair/allnets/blogtitles/cheney_all-small.net)

# Leituras Essenciais

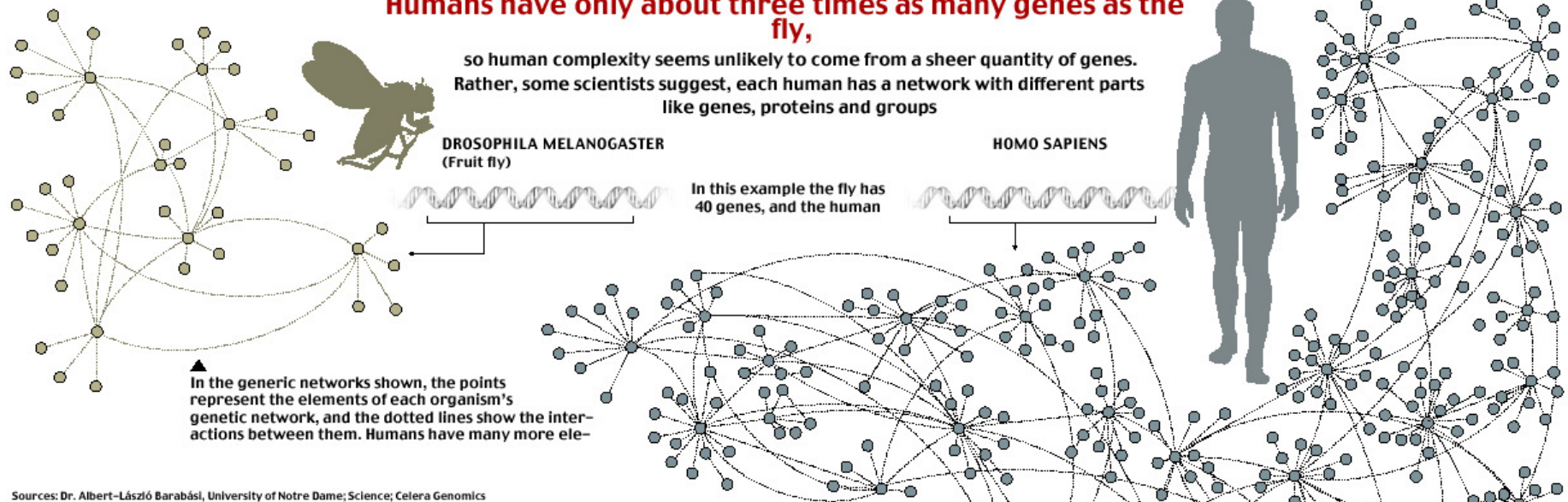
- A.-L. Barabási and E. Bonabeau, *Scale-Free Networks*, Scientific American 288, 60-69 (2003)
- Duncan J. Watts, "Beyond the Small World," Chapter 4 of D.J. Watts, *Six Degrees: The Science of a Connected Age* (New York & London: W.W. Norton & Company, 2002).
- Barabasi, A., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, vol. 286, pp. 509-512.
- Ebel, H., Mielsch, L.I., & Bornholdt, S. (2002). Scale-free topology of e-mail networks. *Physical Review E*, vol. 66, 035103.

# Tópicos

- Redes muito grandes (milhões ou bilhões de nodos e arestas)
- Ocorrência desse tipo de redes na natureza, sociedade, tecnologia, economia, etc.
- Evolução temporal.
- Exemplos: Internet and WWW

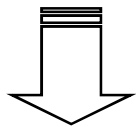
Humans have only about three times as many genes as the fly,

so human complexity seems unlikely to come from a sheer quantity of genes. Rather, some scientists suggest, each human has a network with different parts like genes, proteins and groups



## Sistemas Complexos

Feitos por  
muitos **elementos** não idênticos  
conectados por diversas **interações**.



**NETWORK**

New York Times

Steve Duenes/The New York Times

# Uma descoberta importante....

- Muitas redes na natureza, ecologia, economia, relações humanas e tecnológicas (Internet and WWW) tem a mesma estrutura topológica.
- São chamadas: **scale-free networks**
- Com propriedades idênticas, referentes a estrutura matemática e ao comportamento



## Problemas de pesquisa relacionadas a scale-free networks

- Pode a Internet funcionar bem se centenas de roteadores forem retirados propositalmente de operação?
- Quais partes da Internet são mais vulneráveis a ataques hostis?
- Como projetar algoritmos eficientes de busca para WWW, que façam uso dessas propriedades?

# Redes Randômicas

- Criadas por Paul Erdos and Alfred Renyi em 1959 e 1960.
- Hipóteses básicas:
  - Número fixo de nodos
  - Conectados por arestas aleatórias.
  - Nodos são “democraticos” i.e. Maior parte dos nodos tem aproximadamente um número igual de arestas incidentes.

## Introdução do conceito de *Scale-free networks*

- Introduzido por Barabasi e colaboradores em 1999.
- Em evolução e auto-organizadas.
- Duas regras fundamentais:
  - (a) crescimento no tempo pela adição de nodos e arestas.
  - (b) regra de ligação aos nodos, várias formas.

# Notação Matemática

- Grau do vértice/nodo – no. de arestas incidentes
- Distribuição de graus - característica probabilística da rede
- **$P(k, s, N)$**  - probabilidade que o vértice  $s$  na rede de tamanho  $N$  tenha  $k$  conexões (vizinhos mais próximos).
  
- Distribuição do grau total

$$P(k, N) = \frac{1}{N} \sum_{s=1}^N p(k, s, N)$$

# Distribuição de graus

- Definição:
- A probabilidade que um vértice tem  $k$  arestas.
- A distribuição de Poisson para as *random networks* de Erdos-Renyi.

$$P(k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

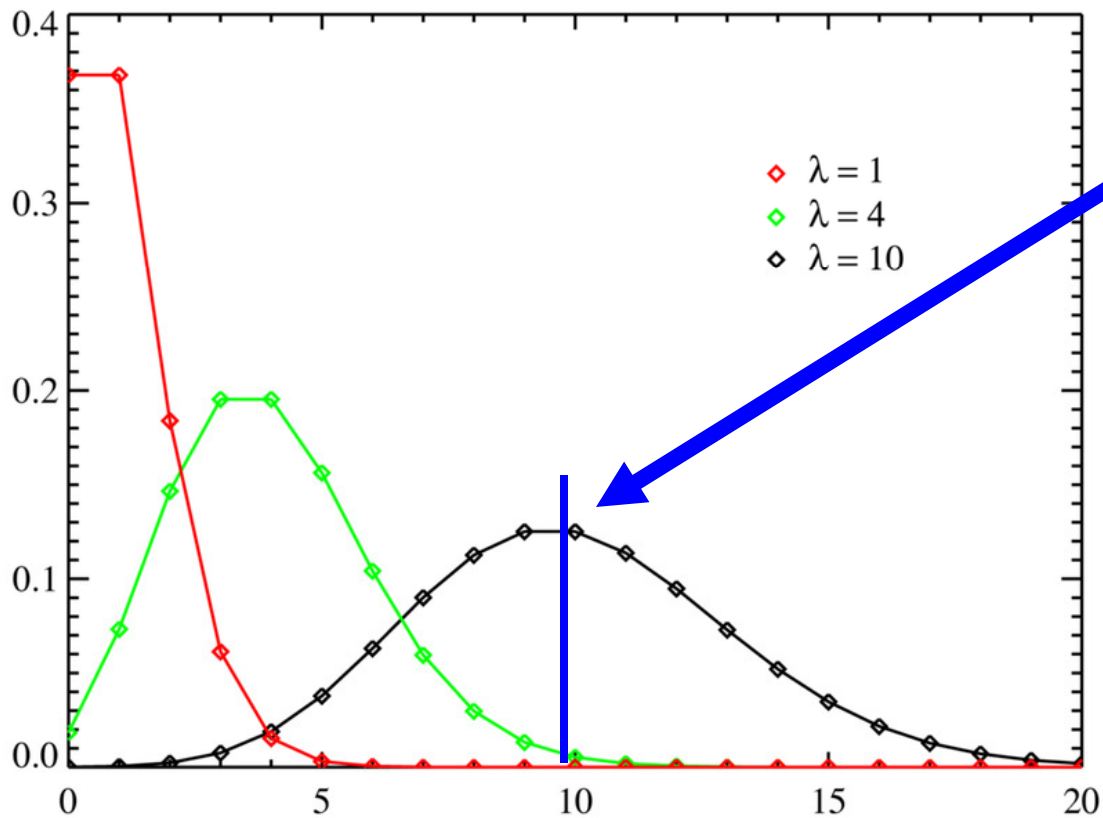
onde

$$\lambda = \sum_{k=0}^{\infty} kP(k)$$

# Distribuição de Poisson

- Distribuição de graus

Escala característica.  
Node médio típico



$$\lambda = \bar{k}$$

Distribuição Power law  
para as redes em evolução e auto-organizadas propostas por  
Barabasi e colaboradores

$$P(k) \propto k^{-\gamma}$$

$$2 < \gamma < 3$$

Intervalo típico

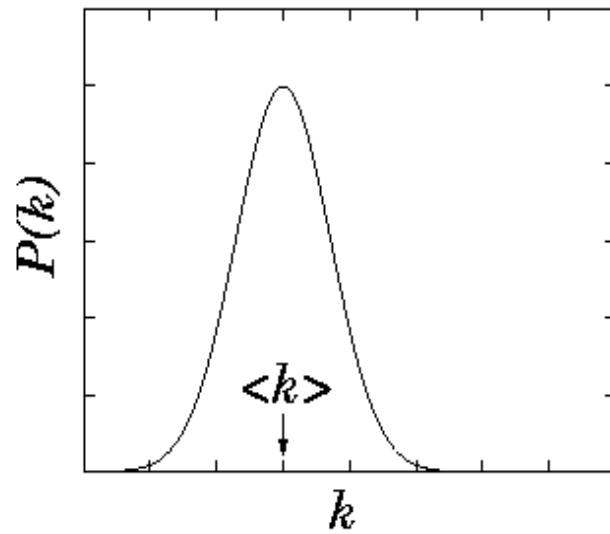
Essas redes não tem um número médio típico de arestas e são chamadas “scale-free”.

# Random vs. Scale Free

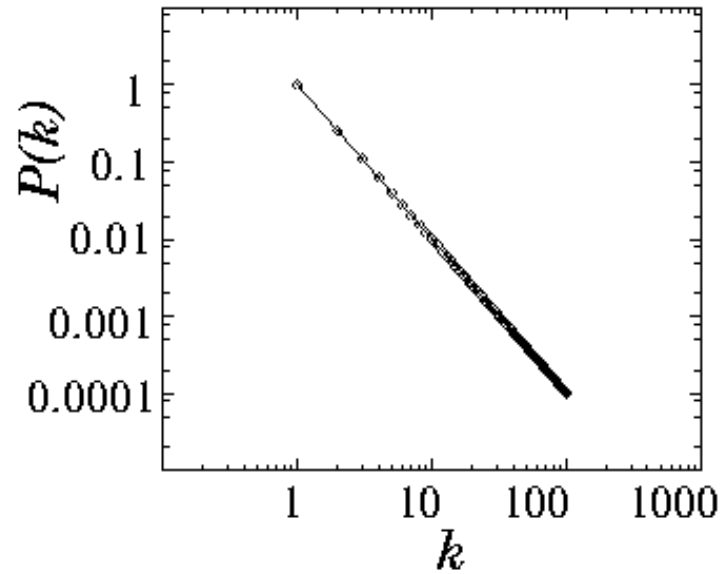
- Exemplos:
- Rede de estradas nos EUA
  - Próxima de uma random network, com uma distribuição em forma de “bell”
- Ao contrário, aeroportos nos EUA
  - Formam uma rede scale free network com vários hubs ligando um grande número de aeroportos.



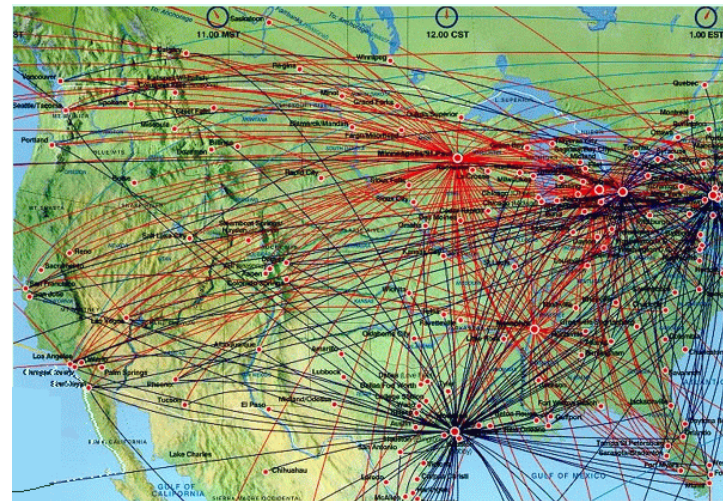
## Poisson distribution



## Power-law distribution



Random Network



Scale-free Network

## Scale free: outros exemplos reais

- Redes de Comunicação: Internet e WWW
- Biológicas : interações entre proteínas no corpo humano.
- Food webs,
- Redes Sociais, citações científicas, ....

# WWW

- Distribuição Power-law levemente modificada

$$P(k) \propto (k + c)^{-\gamma}$$

Home pages na Web

$\gamma$

C

Empresas

**2.05**

**193**

Universidades

**2.62**

**1370**

Cientistas da computação

**2.66**

**12**

A Web como um todo

**2.1**

**0**

# Redes do Tipo Scale-free

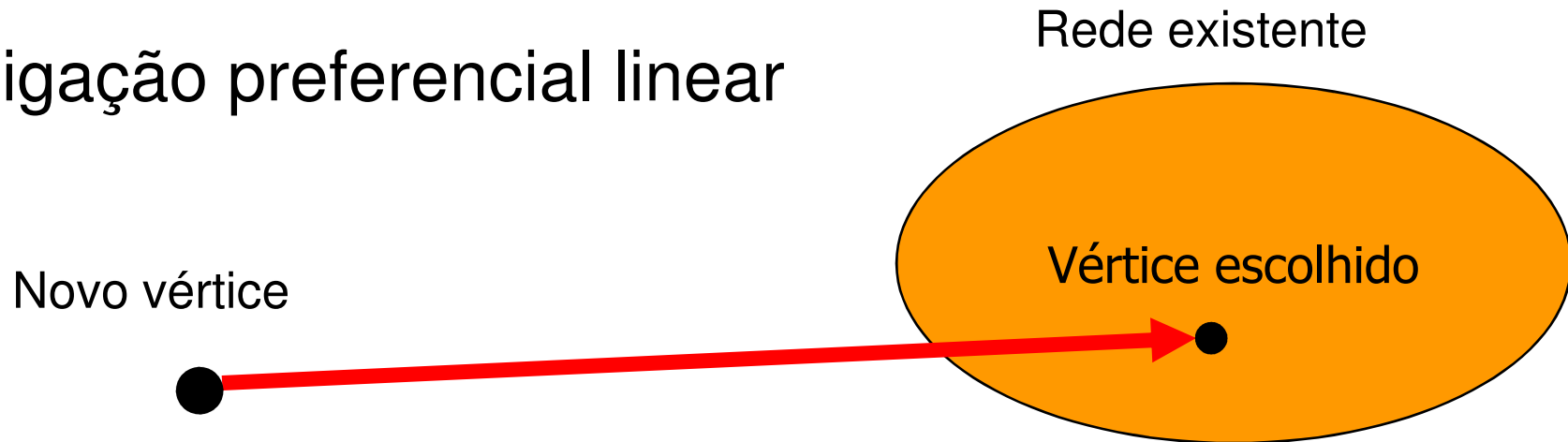
- Redes Scale-free são uma categoria importante das redes reais.
- Elas tem nodos muitos conectados (**hubs**) que representam um papel chave nas propriedades das redes.
- Redes Scale-free são um resultado direto da “self-organization”.
- Tipo especial de crescimento chamado “**the preferential linking**” ou “**preferential attachment**”.
- Enquanto a rede cresce seu novo nodo torna-se preferencialmente conectado a vertices com maiores números de conexões: “rich gets richer”.
- Como resultado dessa auto-organização HUBS são criados.

# Redes Scale-free

- A preferência no processo de crescimento da rede pode tomar várias formas.
- A mais natural é o tipo linear de preferência que resulta em redes **scale-free**.
- Exemplos dessa anexação preferencial inclui a Web, onde as paginas mais populares ganham mais links.
- Popularidade é atraente.

## Redes Scale-free

- Ligação preferencial linear



A probabilidade que uma nova aresta seja ligada a um vértice de grau  $k$  é proporcional a  $k$ .

Isso leva a uma rede scale-free com

$$\gamma = 3$$

Outras regras mais gerais de conexão são possíveis.

## Modelagem tradicional: rede como um grafo estático

Dada uma rede com  $N$  nodos e  $L$  links



Criar um grafo com topologia estatisticamente idêntica

**RESULTADO:** modelo de topologia de rede estática

**PROBLEMA:** **Redes reais são sistemas dinâmicos**

## Redes que Evoluem

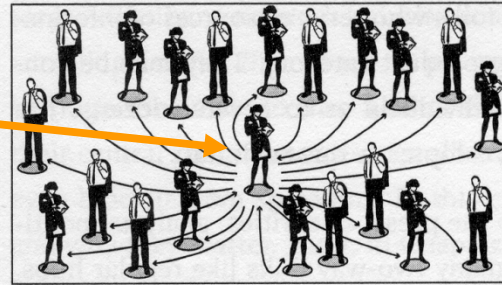
**OBJETIVOS:** capturar a dinâmica das redes

**MÉTODO:**  $\left[ \begin{array}{l} \bullet \text{ identificar os processos que contribuem para formação da topologia} \\ \bullet \text{ Desenvolver modelos dinâmicos que capturam esses processos} \end{array} \right.$

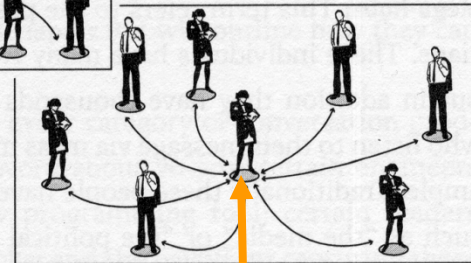


resultado: obtenha a topologia corretamente

# Sociedade

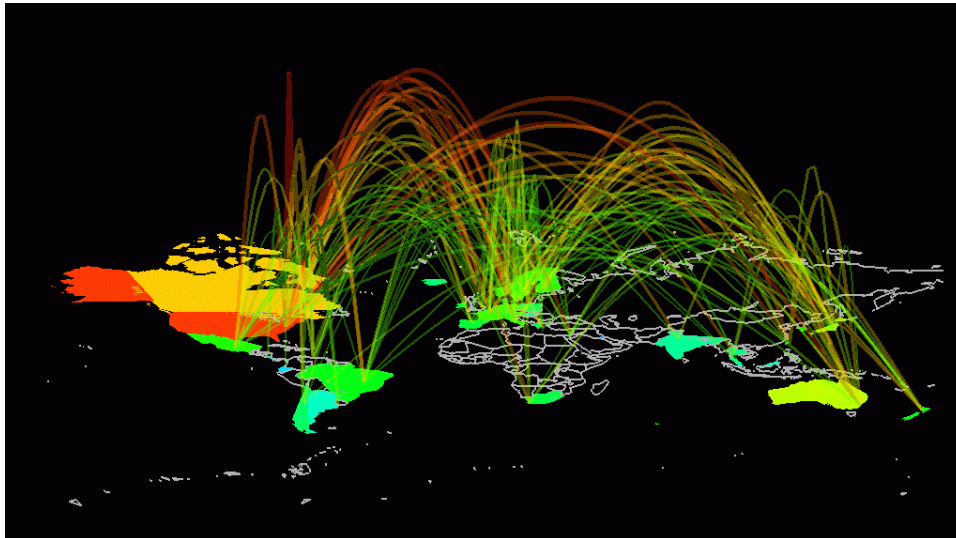


**Mega-Hub.** An MTV veejay spreads the word to thousands or millions of people through one-way links.



**Hub.** This undergraduate has spread the word to seven other people through two-way links.

# Internet





# Redes Complexas

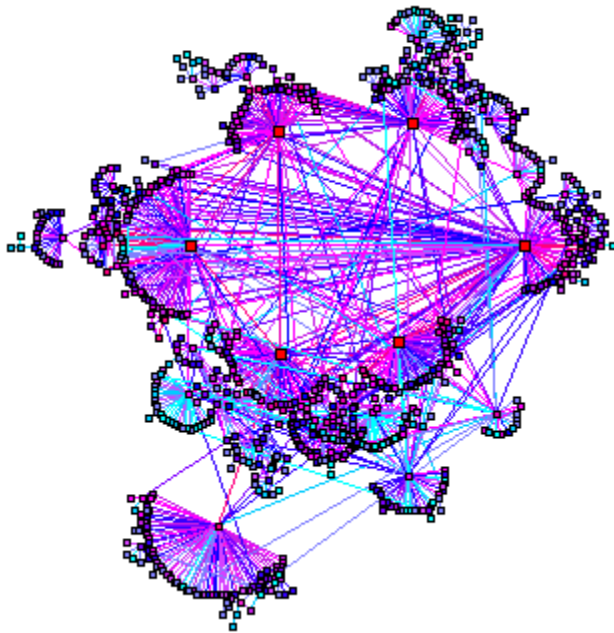
- As redes complexas são do tipo randomicas (random networks)?
- Redes tem finalidades!
- Como a topologia afeta a função/finalidade da rede
- Quais as propriedades comuns?

# World Wide Web

**Nodos:** documentos da Web

**Links:** URL links

800 milhões de documentos  
(Lawrence, 1999)

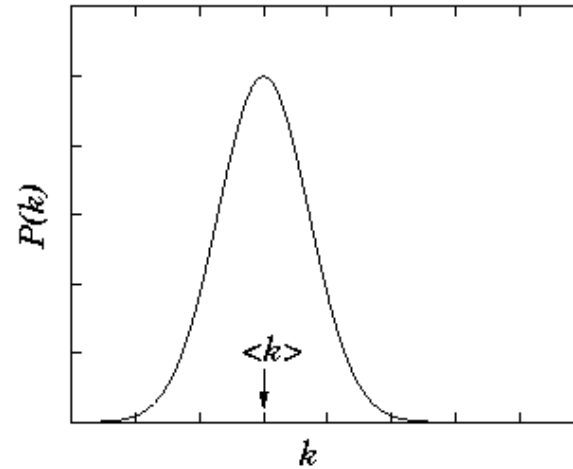


**ROBOT:** coletou  
todas URL's  
encontradas nos  
documentos e seguiu  
as recursivamente.

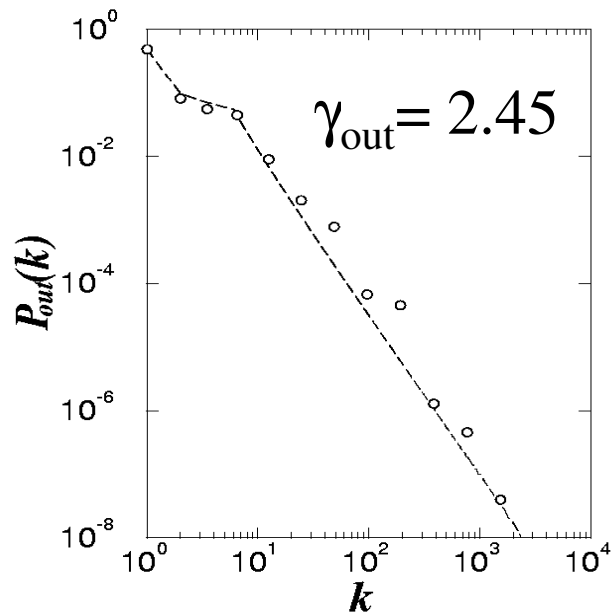
E aí?

R. Albert, H. Jeong, A-L Barabasi, Nature, **401** 130 (1999)

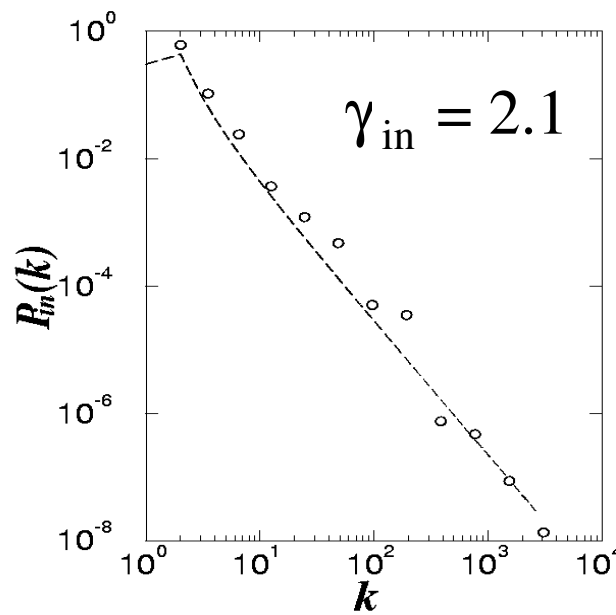
# O que era esperado?



O que foi encontrado:



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



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

# Origem das Redes SCALE-FREE

(1) O número de nodos (N) **NÃO** é fixo

Redes continuamente expandem pela adição de novos nodos e arestas

Exemplos:

**WWW** : adição de novos documentos

**Citação** : publicação de novos papers

(2) A anexação (attachment) **NÃO** é uniforme

Um nodo é ligado com maior probabilidade a um nodo que já tem um grande número de links.

Exemplos :

**WWW** : novos documentos se ligam a sites muito conhecidos (Google, CNN, Yahoo, Ebay, YouTube, NewYork Times, Citeseer, DLBP, etc)

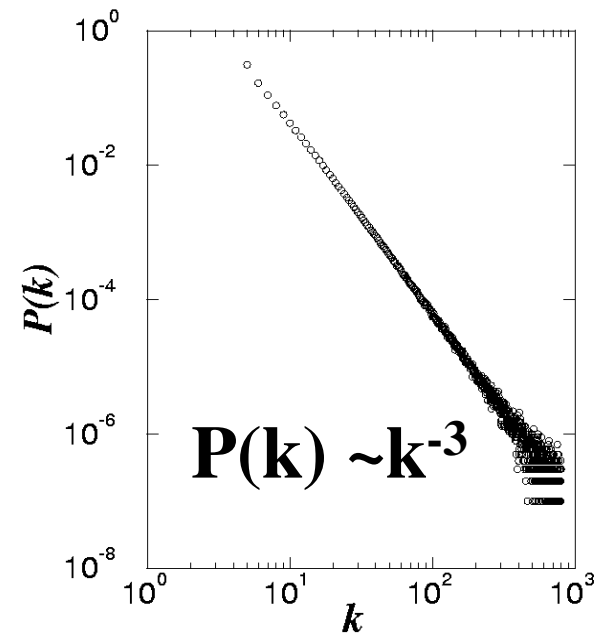
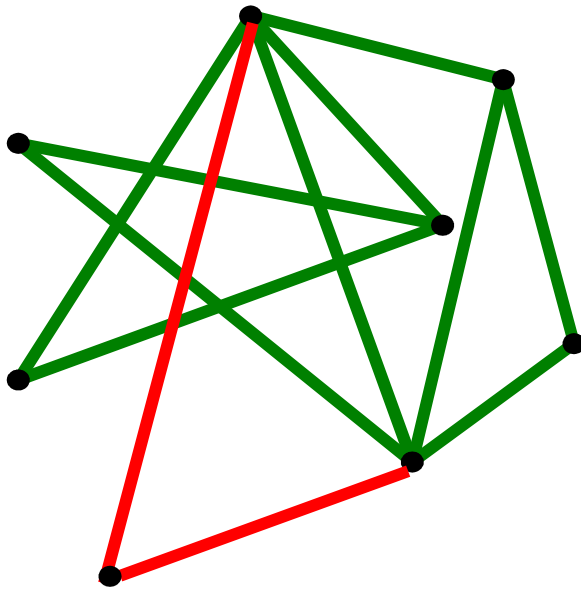
**Citação** : papers muito citados são mais prováveis de serem citados novamente.

# Modelo BA: Scale-free

(1) **Crescimento:** A cada passo de tempo adicionamos um novo nodo com  $m$  arestas (conectado a nodos já presentes no sistema).

(2) **Preferential Attachment:** A probabilidade  $\Pi$  que um novo nodo será conectado ao nodo  $i$  depende da conectividade  $k_i$  daquele nodo

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



# Modelo Bianconi & Barabasi

## Competition in Evolving Networks

(1) **Crescimento:** A cada passo de tempo adicionamos um novo nodo com  $m$  arestas (conectado a nodos já presentes no sistema).

(2) **Preferential Attachment and Competition** A probabilidade  $\Pi$  que um novo nodo será conectado ao nodo  $i$  depende da conectividade  $k_i$  daquele nodo e de um fator de “fitness”

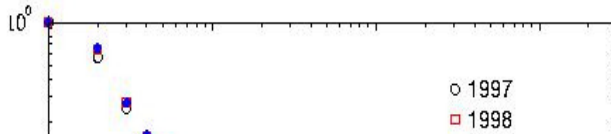
$$\Pi(k_i) = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$
$$k_{\eta_i}(t, t_0) = m \left( \frac{t}{t_0} \right)^{\beta(\eta_i)}$$

Propriedades principais:

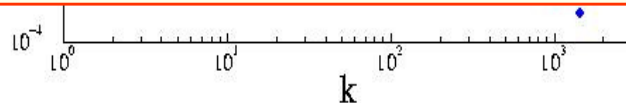
- redes complexas
- preferential attachment
- clustering

**Modeling of scale-free networks  
by Barabasi et al. (1999)**

**Distribuição de conectividade**



- **A Internet e a World-Wide-Web**
- **Redes de Proteínas**
- **Redes Sociais**
- **Food-webs e redes ecologicas**



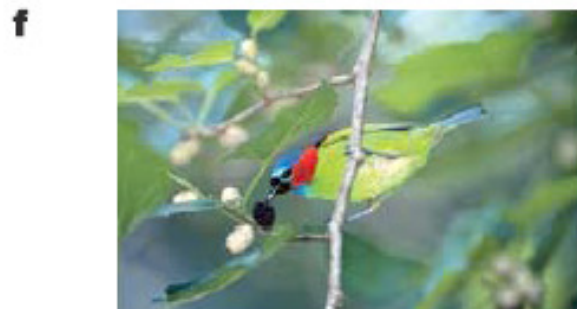
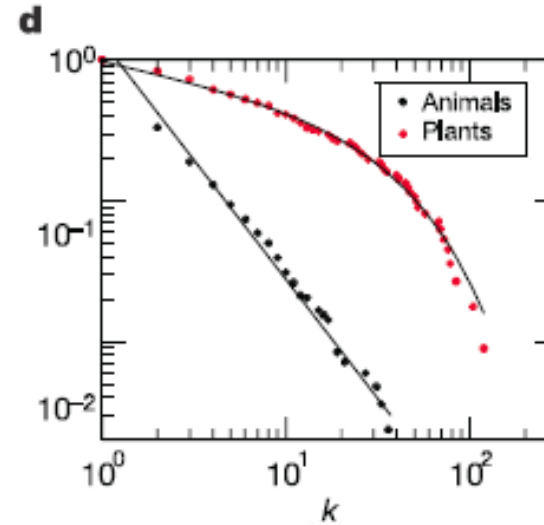
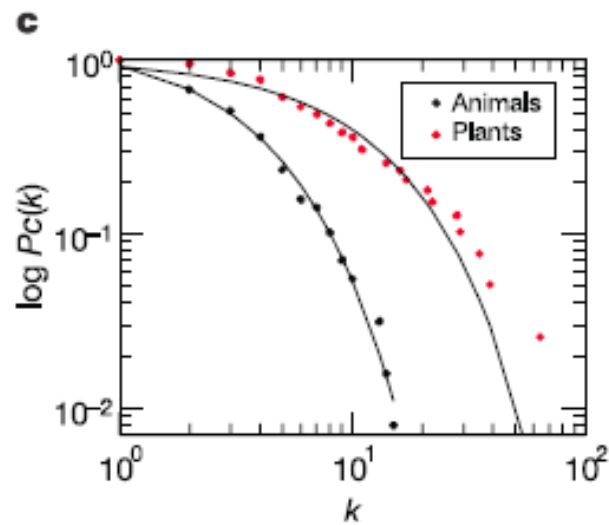
- $\langle k \rangle = \text{const}$
- $\langle k^2 \rangle \rightarrow \infty$

**Propriedades Scale-free**



**Variabilidade**

# Redes Ecológicas



**Distribution of linkage density in ecological networks.** **a–e**, The cumulative probabilities  $P_c(k)$ , for  $k \geq k$ , where  $P(k)$  is the probability a species is linked to other species, and is given by  $P(k) \sim k^{-\gamma} e^{-k/\gamma}$  where  $e^{-k/\gamma}$  is a cut-off at some characteristic scale  $\gamma$ . Panels **a** (log–log) and **b** (log–linear) show the distribution of linkage density in the animal web<sup>77</sup>; and panels **c** (log–log) and **d** (log–linear) show the distribution in the plant web<sup>77</sup>; and **e**, **f** show the distribution in the animal web<sup>77</sup>; and **e**, **g** show the distribution in the plant web<sup>77</sup>.

(filled circles) and black circles,  $k$  is the number of links to other species, and  $P_c(k)$  is the cumulative probability of a species having  $k$  or more links to other species.





# Redes de Linguagem

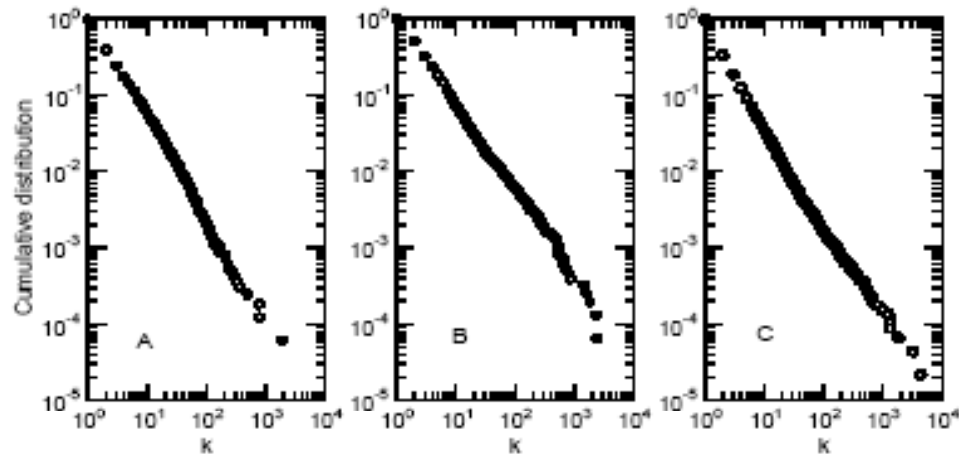


FIG. 3 Scaling laws in language webs. Here three different corpora have been used and the co-occurrence networks have been analysed for: (a) basque, (b) english and (c) russian. Specifically, we computed the degree distribution  $P(k)$  (box 1) and in order to smooth the fluctuations the cumulative distribution has been used, being defined as  $P_{>}(k) = \sum_{j>k} P(j)$ . Each corpus has  $10^4$  lines of text. Although some differences exist (44), their global patterns are rather similar, thus suggesting common principles of organization.

## Virus Naturais em computadores

- DNS-cache computer viruses
- Routing tables corruption

## Virus transportados por vírus

- ftp, file exchange, etc.

Topologia Internet

Rede de E-mail

Ebel et al. (2002)

## Computer worms

- Difusão de e-mail
- Auto-replicação

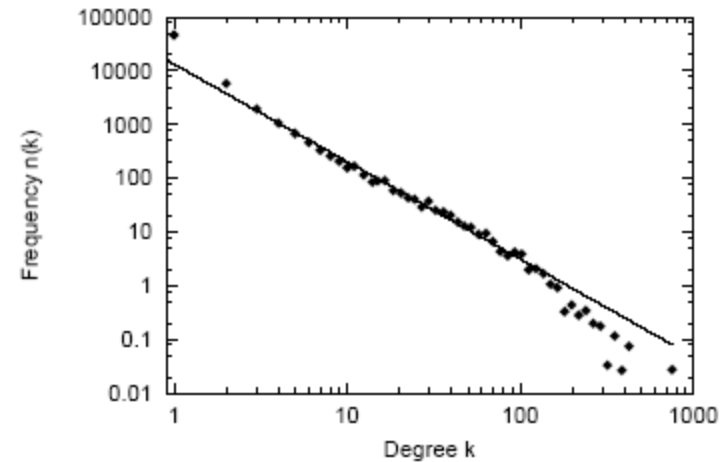


FIG. 1: Degree distribution of the e-mail network. The

## Leituras para Próxima Aula

- Tópicos: robustez e vulnerabilidade em SF networks
- Albert, H. Jeong, and A.-L. Barabási, Error and attack tolerance in complex networks, Nature 406
- Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999). On power-law relationships of the Internet topology. Computer Communication Review, vol. 29, pp. 251-262.