

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

-Bloco #3-

Redes: métricas  
2º semestre de 2008

Virgílio A. F. Almeida  
Agosto de 2008



Departamento de Ciência da Computação  
Universidade Federal de Minas Gerais

## Redes de Informação?

- Rede: uma coleção de **entidades** que estão interconectadas
  - Uma aresta entre dois nós denota uma interação entre as duas entidades.
  - Vemos essa interação como **information exchange**, daí a ideia de Redes de Informação
  - O termo é bem geral

# Por que interessa?

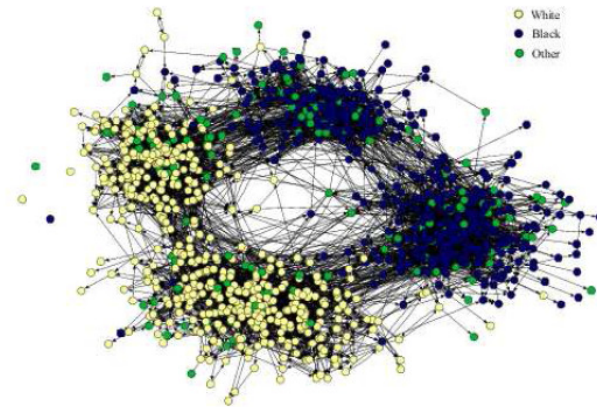
- Redes estão em todas áreas
  - Cada vez mais um número maior de sistemas pode ser modelado como redes e mais dados são coletados
  - Modelos de grafos tradicionais não se aplicam aos casos reais
- Redes em larga escala necessitam de novos algoritmos de análise
- Descobrir algoritmos capazes de fazer uso do “conhecimento” existente nas redes.
- Uma nova área: science of networks
  - Envolve múltiplas disciplinas: computação, física, biologia, economia, sociologia, matemática, etc.

# Tipos de redes

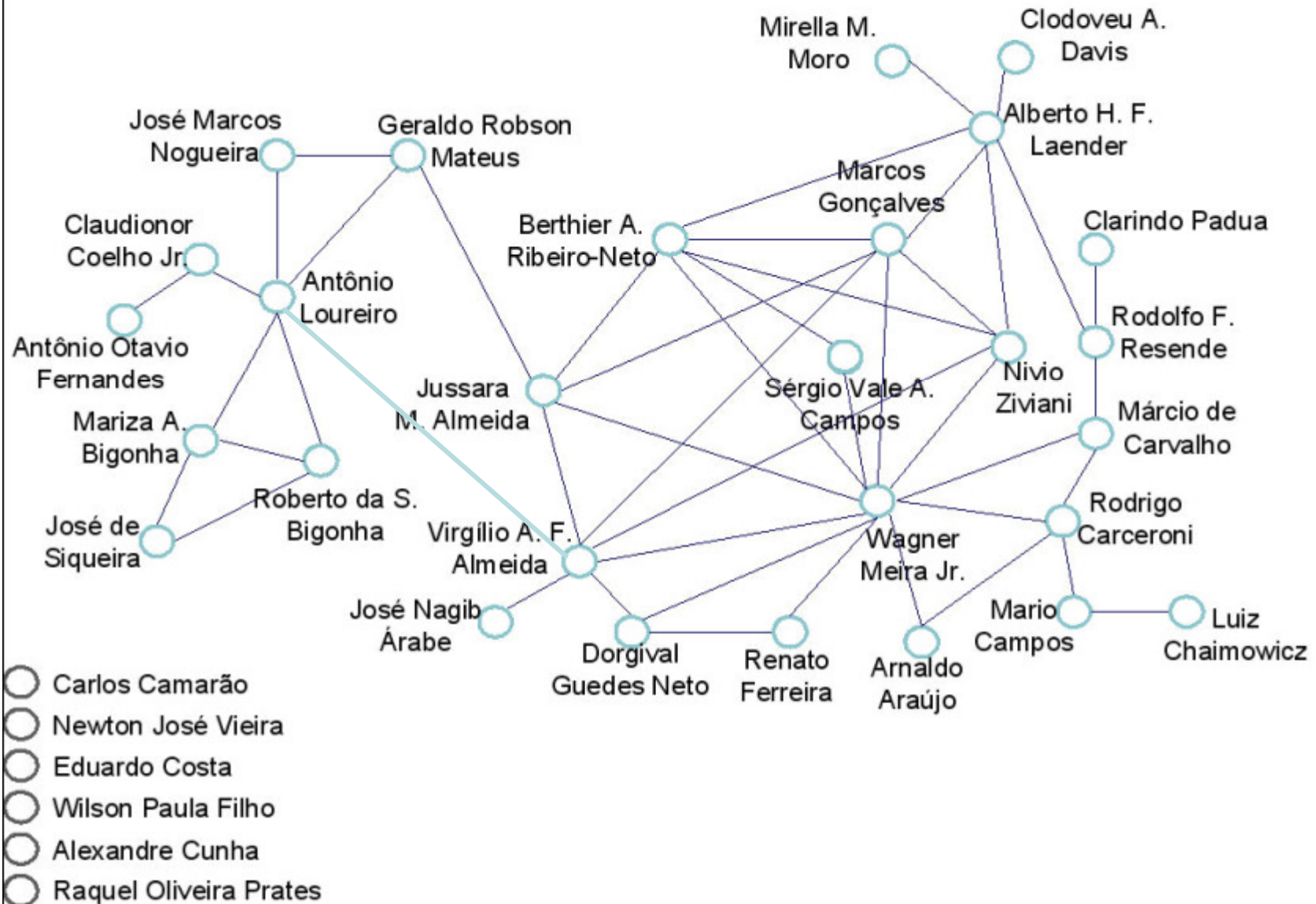
- Redes sociais
- Redes de conhecimento (informação)
- Redes tecnológicas
- Redes biológicas

# Redes Sociais

- Aresta denota/representa uma relação social
  - Redes de "conhecidos"
  - Rede de atores
  - Rede de co-autores
  - Rede de diretores
  - Rede de chamadas telefônicas
  - Rede de email
  - Rede de Instant Messaging
  - Redes de contatos sexuais
  - Redes de amigos em social media software



# Rede Co-Autoria, PPGC/UFMG

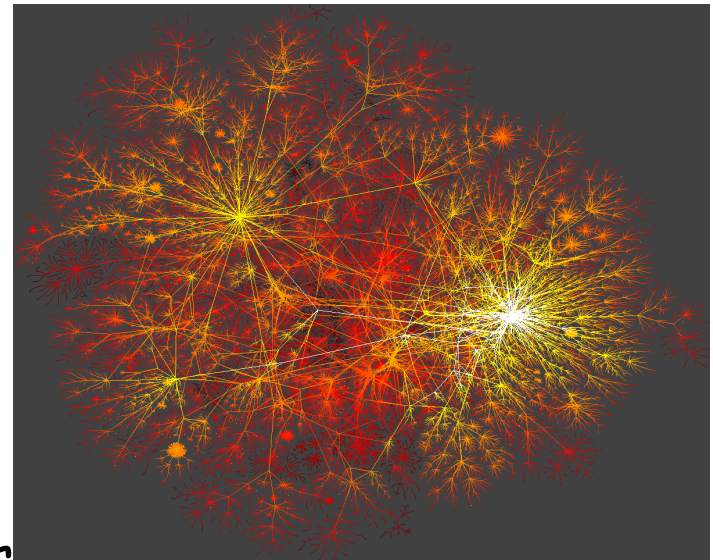


# Redes de conhecimento (informação)

- Nós armazenam informação, arestas associam informação
  - Rede de citação (orientada)
  - A Web (orientada)
  - Redes Peer-to-Peer
  - Redes de palavras
  - Rede de confiança "trust"
  - Rede de...

# Redes tecnológicas

- Redes construídas para distribuição de "commodity"
  - A Internet
    - router level, AS level
  - Rede de energia elétrica
  - Rede de linhas aéreas
  - Rede telefônica
  - Rede de transporte
    - Estrada, ruas, estrada de ferro
  - Grafos de Software





# Redes biológicas

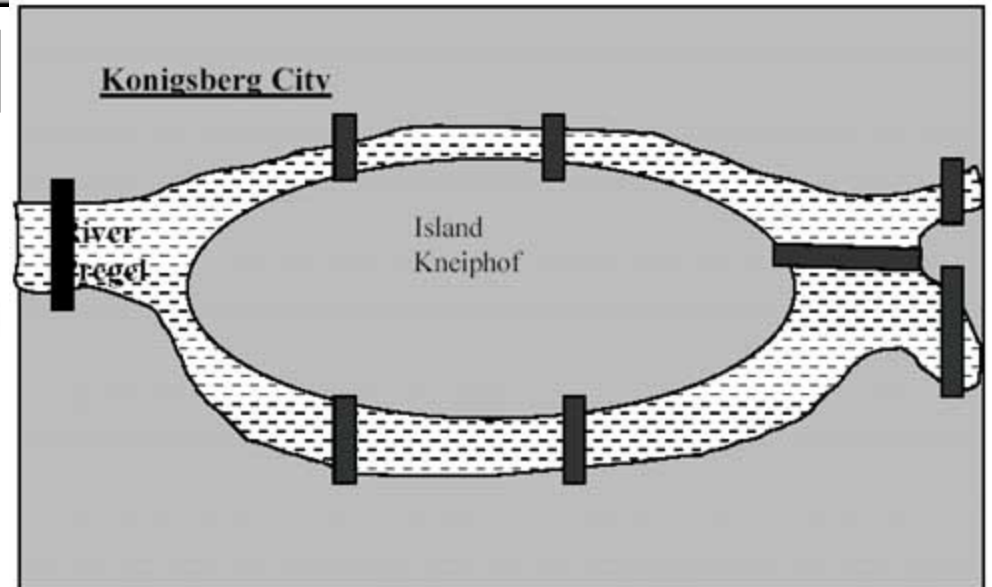
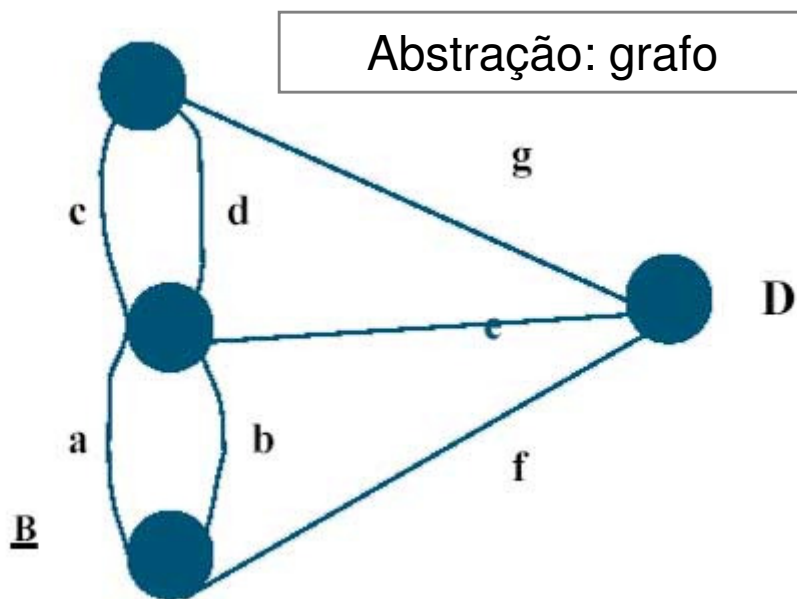
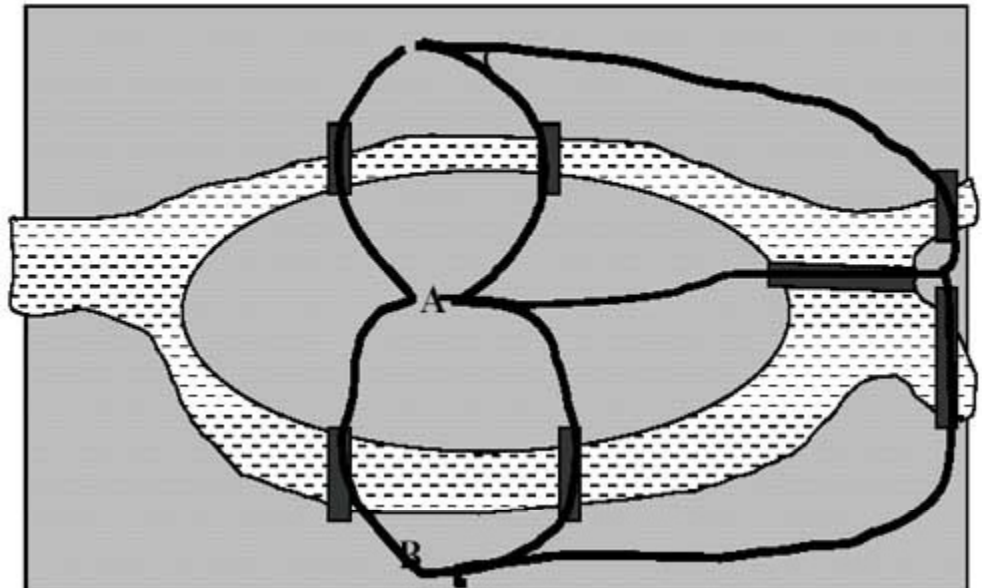
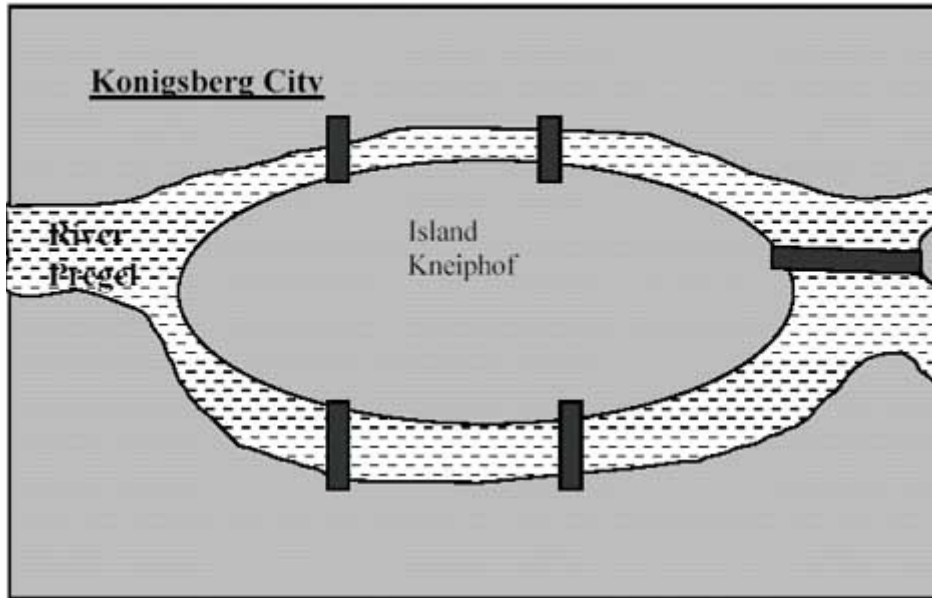
- Sistemas biológicos representados como redes
  - Redes de interação Proteína-Proteína
  - Redes de genes
  - Redes do tipo "Food Web"
  - Redes Neurais



# E agora?

- O "mundo" esta cheio de redes! E agora????
  - Entender suas topologias e medir suas propriedades
  - Estudar sua dinâmica de evolução
  - Criar modelos realistas
  - Criar algoritmos que aproveitem a estrutura e relações das redes

Konigsberg Bridge Problem: Does there exist any single path that crosses all seven bridges exactly once each?



# Erdős-Renyi Random graphs

- O modelo  $G_{n,p}$ 
  - $n$  : número de vértices
  - $0 \leq p \leq 1$
  - Para gerar uma amostra/sample, aleatoriamente, seleciona-se  $n$  inicialmente desconexos vértices e para cada par  $(i,j)$ , gere uma aresta  $(i,j)$  independentemente com probabilidade  $p$ .
- O modelo  $G_{n,m}$  é o conjunto de todos os grafos consistindo de  $n$  nodos e  $m$  arestas.
  - Para gerar uma amostra de grafo uniformemente, de maneira aleatória, simplesmente atribui-se  $m$  arestas entre pares escolhidos aleatoriamente de  $n$  nodos/vértice inicialmente não conectados.

## Distribuição de graus em um grafo aleatório

- A distribuição segue uma **binomial**

$$p(k) = B(n; k; p) = \binom{n}{k} p^k (1-p)^{n-k}$$

- Assumindo  $z=np$  fixo, qdo  $n \rightarrow \infty$   $B(n, k, p)$  pode ser aproximada por uma distribuição de **Poisson**

$$p(k) = P(k; z) = \frac{z^k}{k!} e^{-z}$$

- Altamente concentrada em torno da média, com uma cauda que cai exponencialmente.
- Exercício: demonstrar que a binomial converge para Poisson quando  $n \rightarrow \infty$

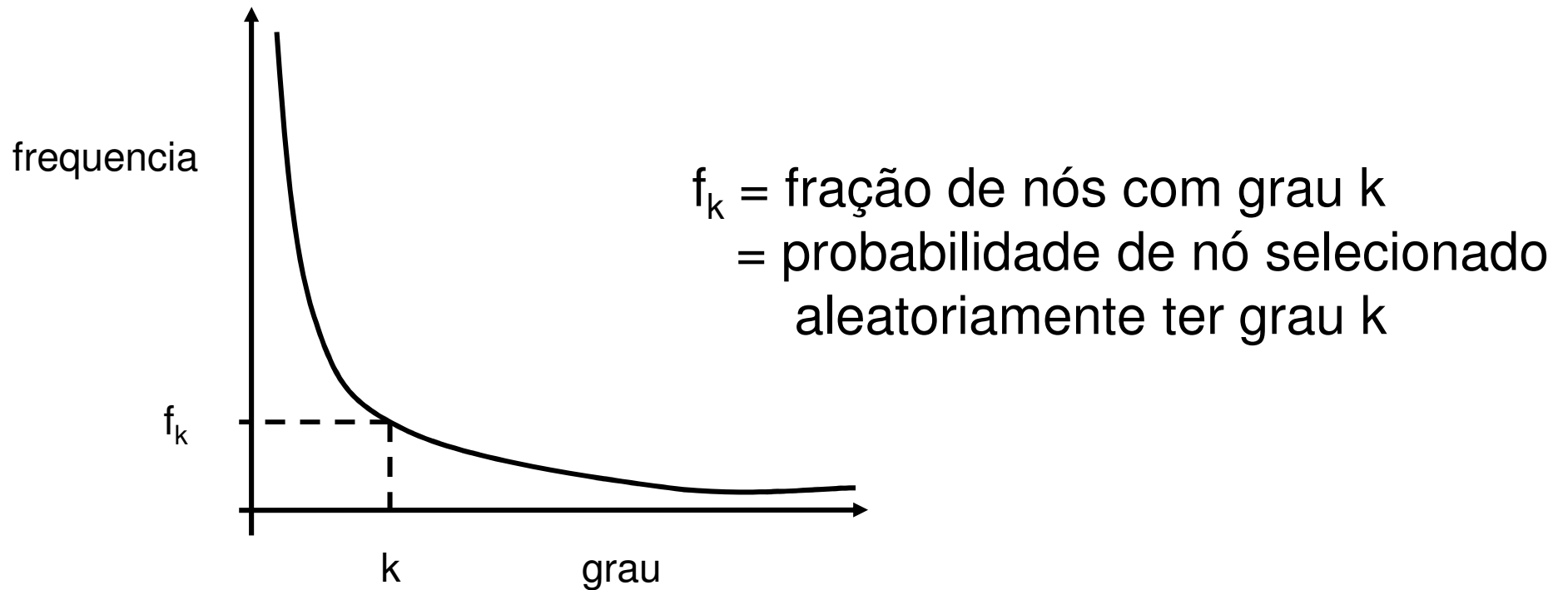
# Grafos randômicos e a realidade

- Uma teoria elegante estudada exaustivamente.
- Grafos Randomicos tem sido usados como modelos geradores de outros grafos.
- No entanto, não capturam a realidade...

# Propriedades das Redes

- Distribuição dos graus
- Fenômeno do "Small world"
- *Clustering Coefficient*
- *Mixing patterns (assortative/dissortative)*
- *Betweenness*
- Comunidades e clusters

# Distribuição dos graus



- Problema: determine a distribuição de probabilidade que melhor ajuste ("best-fit") a um conjunto de dados experimentais de uma rede. Proponha uma coleta de dados e faça o fitting!



# Distribuições Power-law

- O grau de distribuição de redes reais segue na maioria das vezes uma função do tipo **power law**:

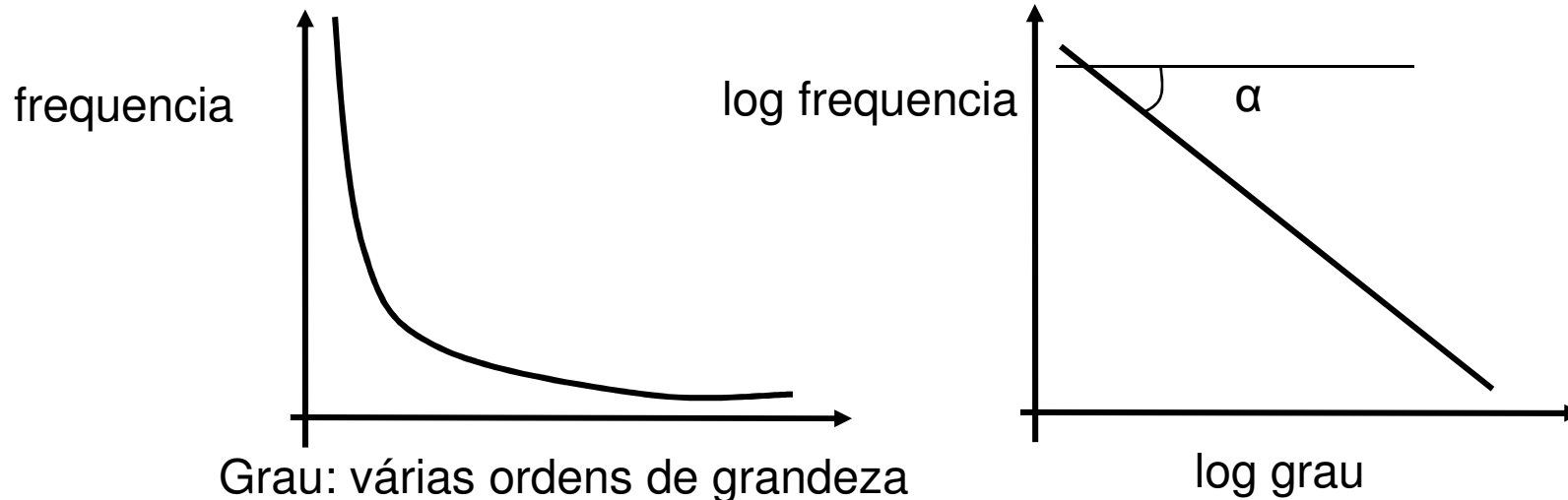
$$p(k) = Ck^{-\alpha}$$

- Distribuição de cauda pesada (Heavy-tail distribution)
  - Existe uma fração de nós não desprezível que tem graus muito altos (hubs)
  - **scale-free**: média não informativa e não tem características de escala.
- Num contraste gritante com grafos randomicos!
  - Altamente concentrado em torno da média
  - A probabilidade de nós com graus muito altos é exponencialmente pequena.

## Assinatura Power-law

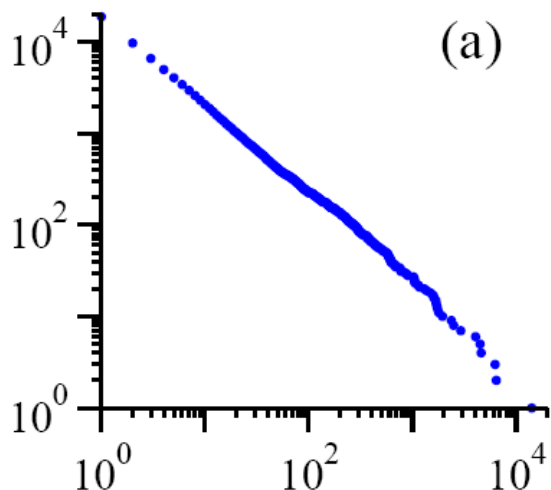
- Distribuição Power-law leva a uma linha reta num gráfico log-log.

$$\underbrace{\log p(k)}_y = -\alpha \underbrace{\log k}_x + \underbrace{\log C}_c, \quad \text{para } x > 0$$

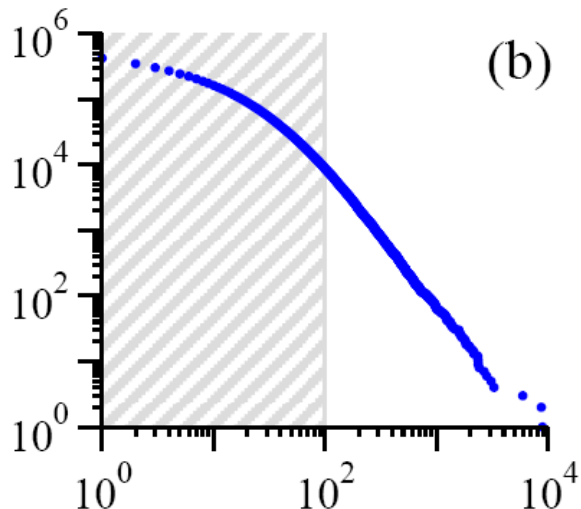


- $\alpha$  : expoente power-law (tipicamente  $2 \leq \alpha \leq 3$ )

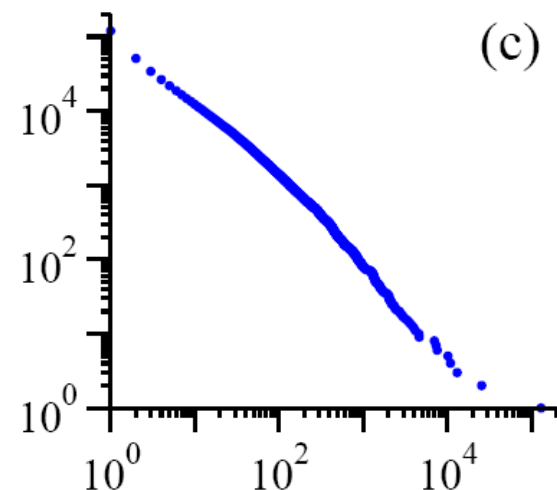
# Exemplos I



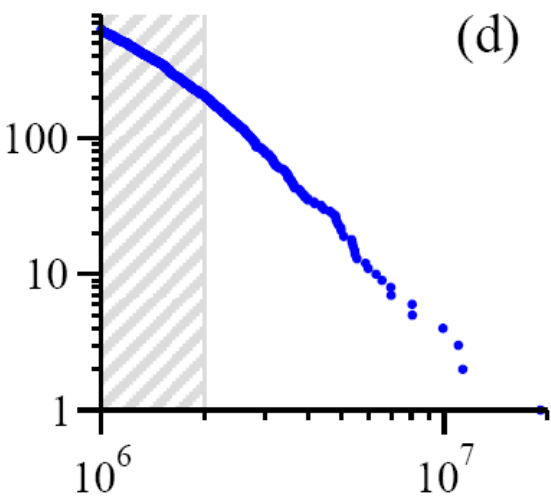
word frequency  
Moby Dick



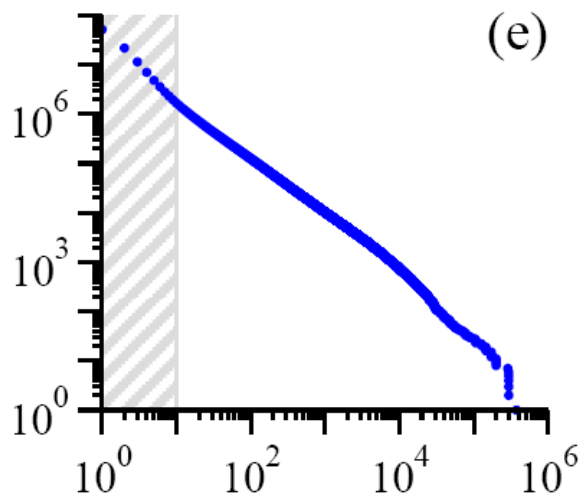
citations  
scientific papers 1981-1997



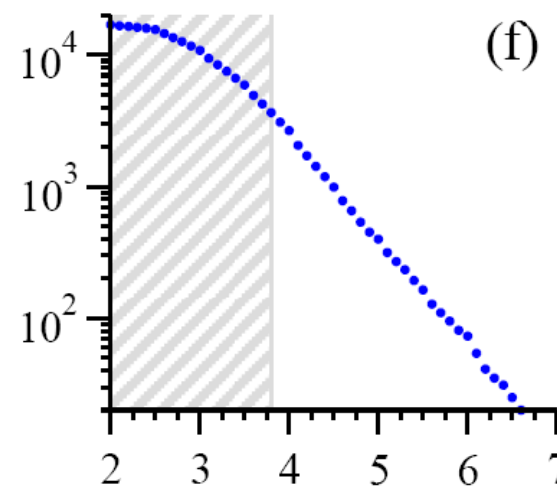
web hits  
AOL users visiting sites '97



books sold  
bestsellers 1895-1965

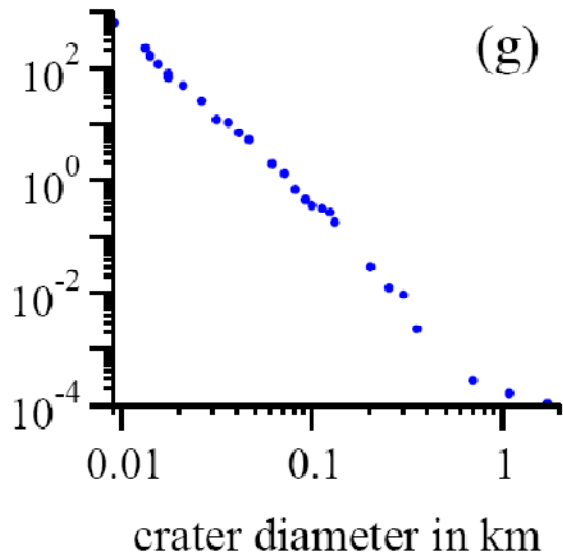


telephone calls received  
AT&T customers on 1 day

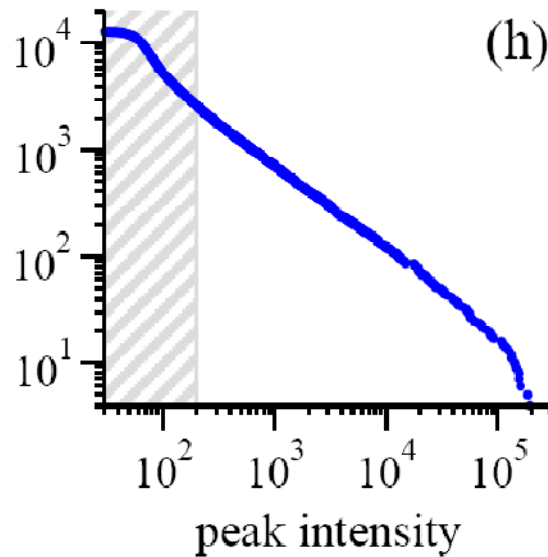


earthquake magnitude  
California 1910-1992

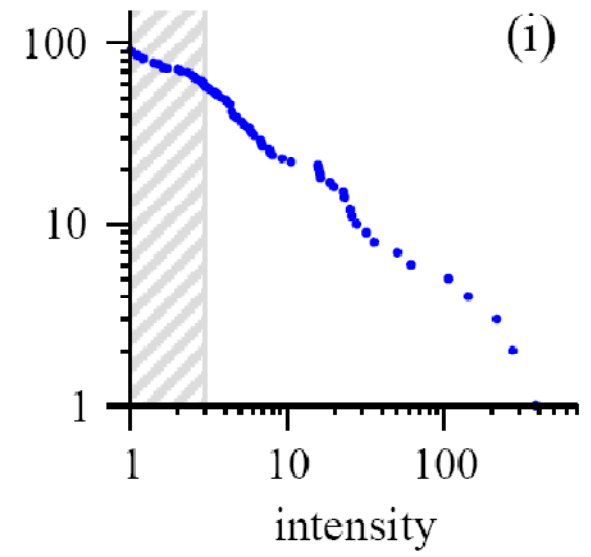
# Exemplos II



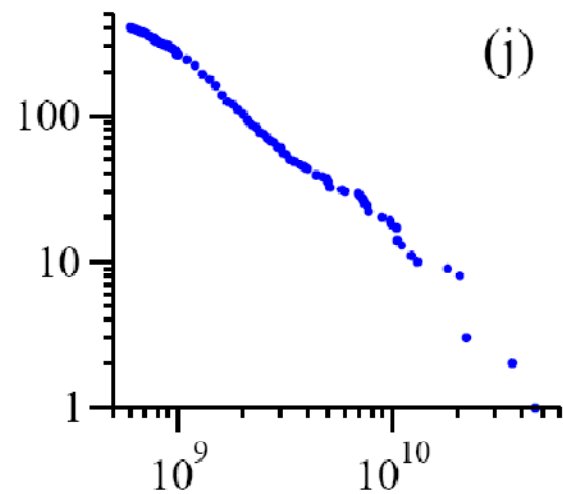
Moon



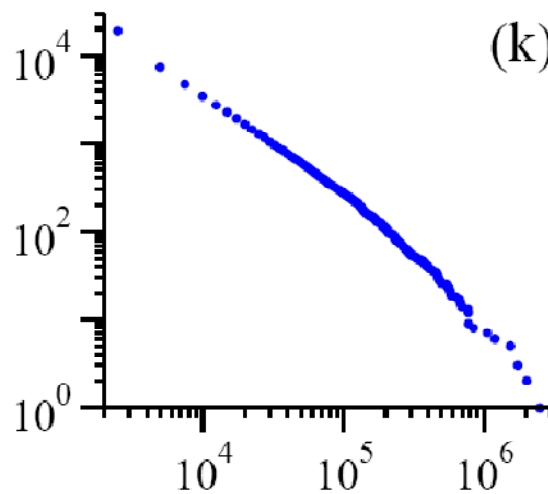
Solar flares



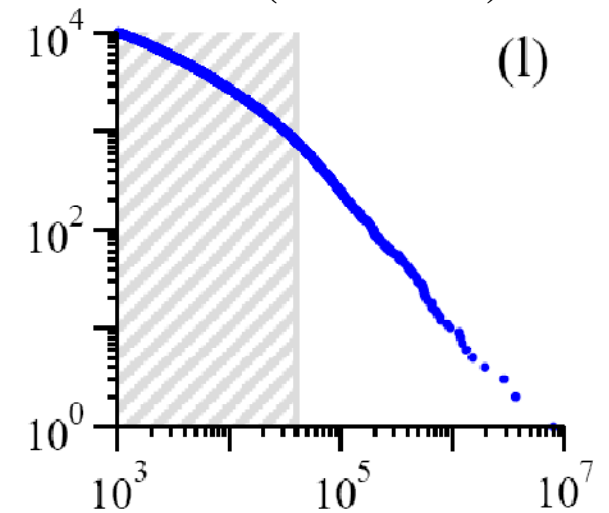
wars (1816-1980)



richest individuals 2003



US family names 1990



US cities 2003

# Power laws - leis de potência

No gráfico log-log:

$$\ln(p(x)) = c - \alpha \ln(x)$$

Recolocando o expoente:

$$p(x) = Cx^{-\alpha}$$

## ` ` Fitting power laws"

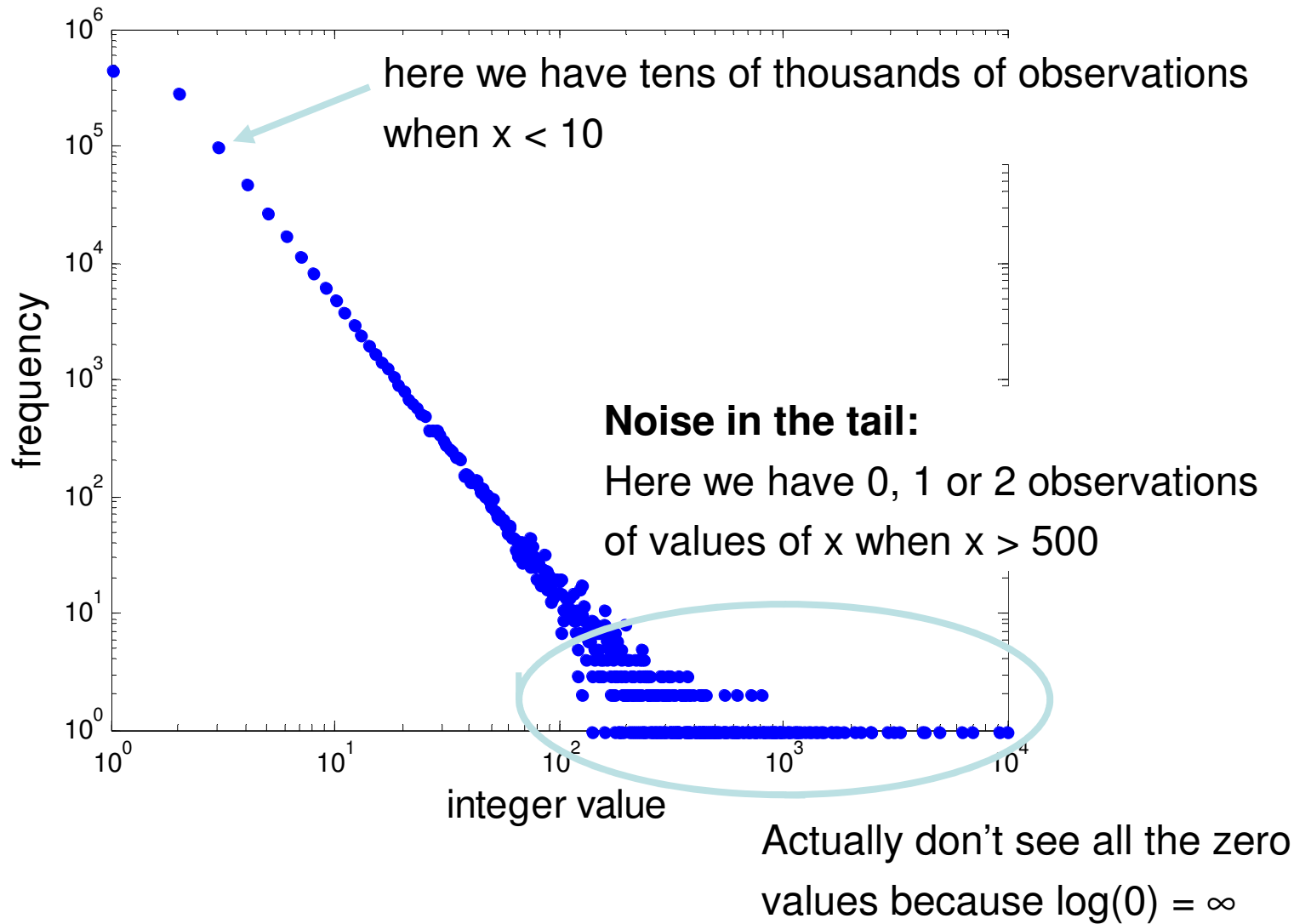
- Método mais simples : realizar uma regressão linear no espaço log-log para determinar a inclinação da reta.
- Este método pode não funcionar bem na prática. Por que?

Boa pergunta!!!

# Exemplo

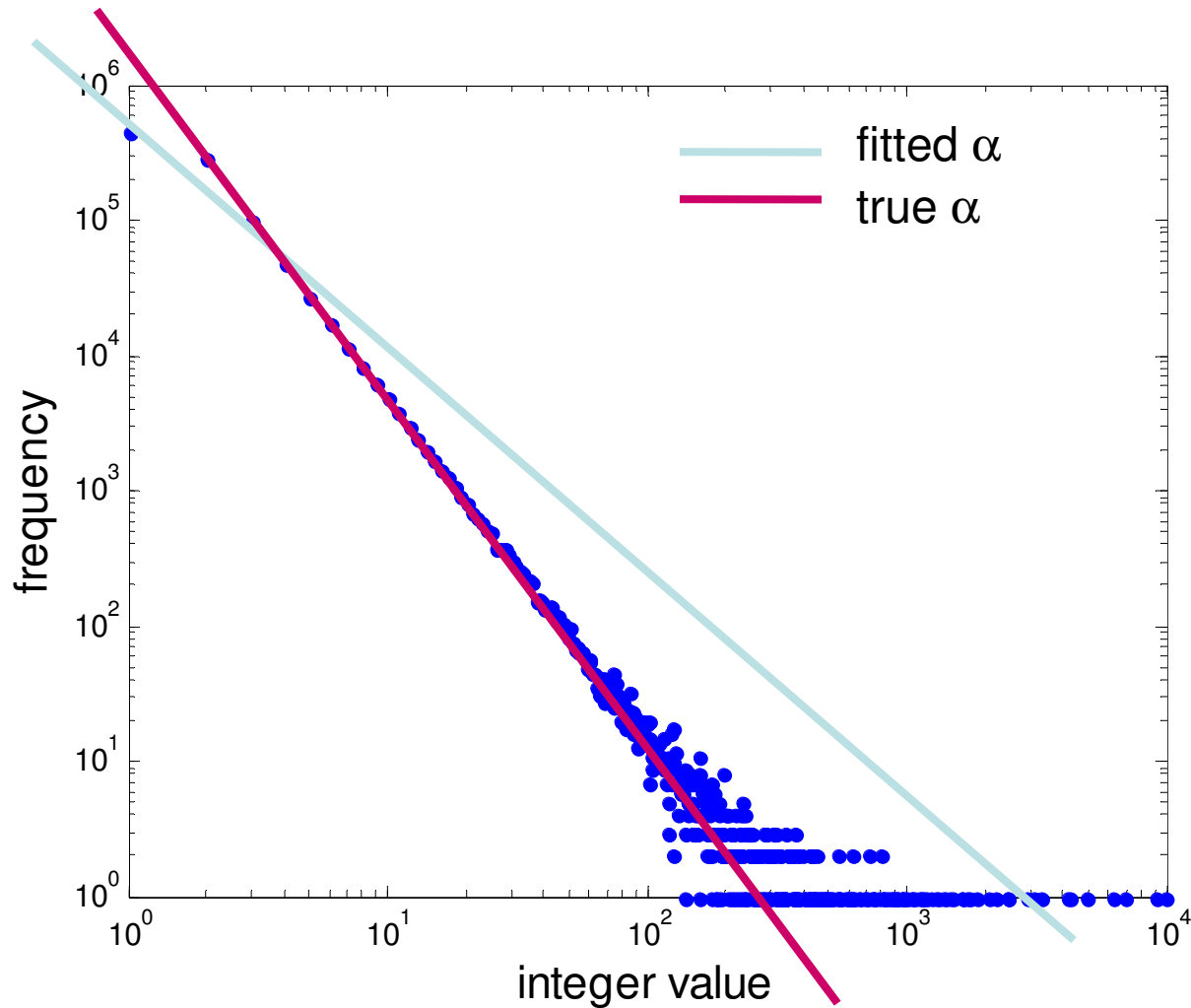
- Considere 1 milhão de números aleatórios de uma distribuição com  $a = 2.5$
- Pode ser gerado usando o método da transformação'
- Gere números aleatórios  $r$  no intervalo da unidade  $0 \leq r < 1$
- então  $x = (1-r)^{-1/(a-1)}$  é um número aleatório distribuído com lei de potência no intervalo  $1 \leq x < \infty$

- Alguns bins, mas com gráfico em escala log-log

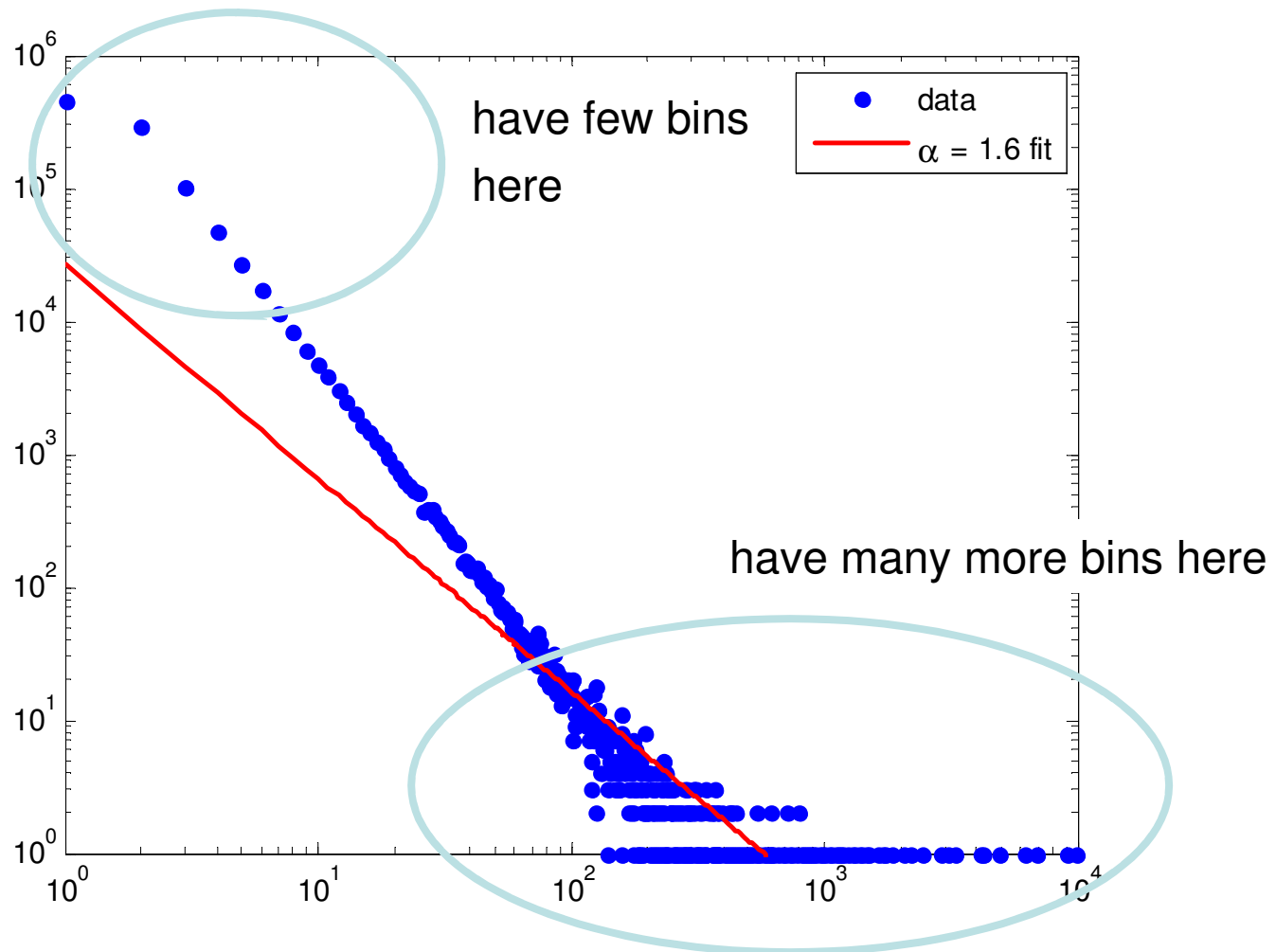




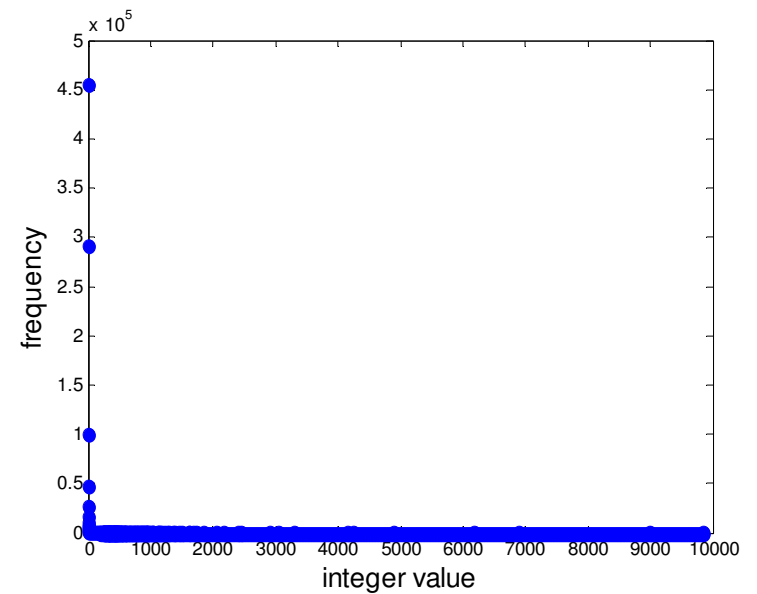
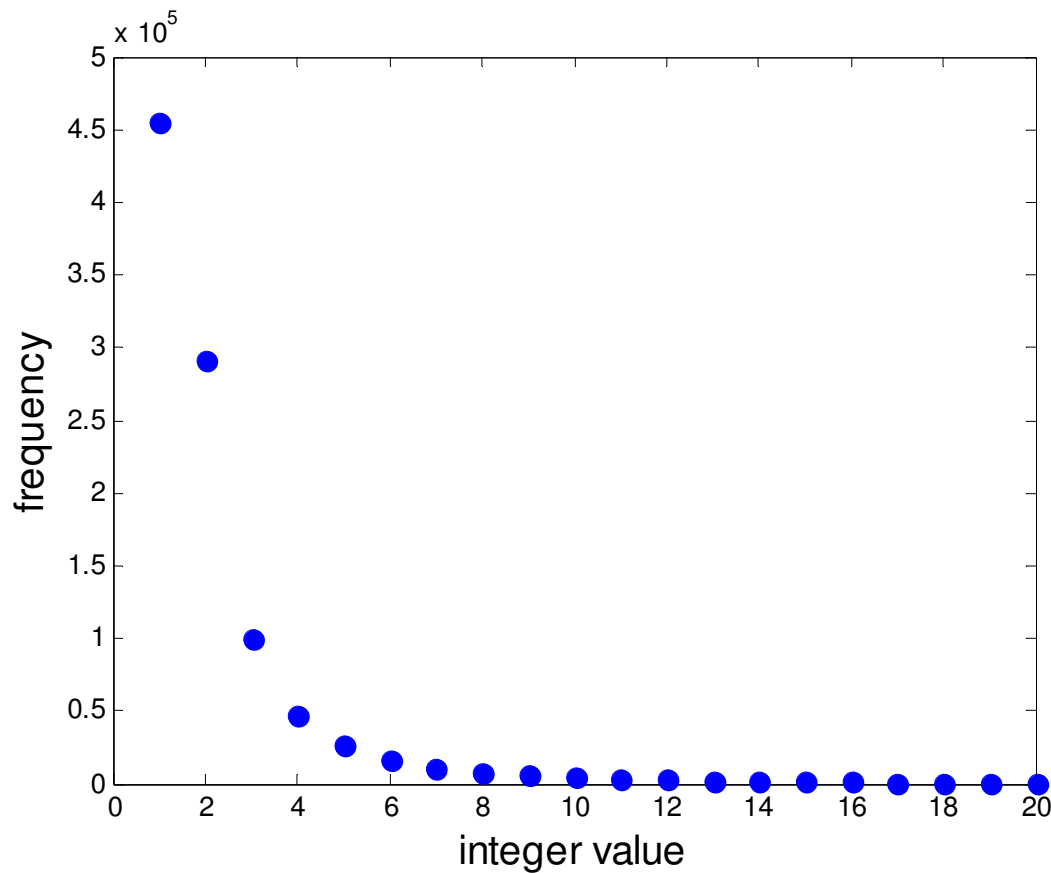
- *Fitting* uma linha reta através de uma regressão dos mínimos quadrados irá obter valores para o expoente que são muito baixos! Por que?



- Ruído na cauda torna o resultado da muito “inclinado”

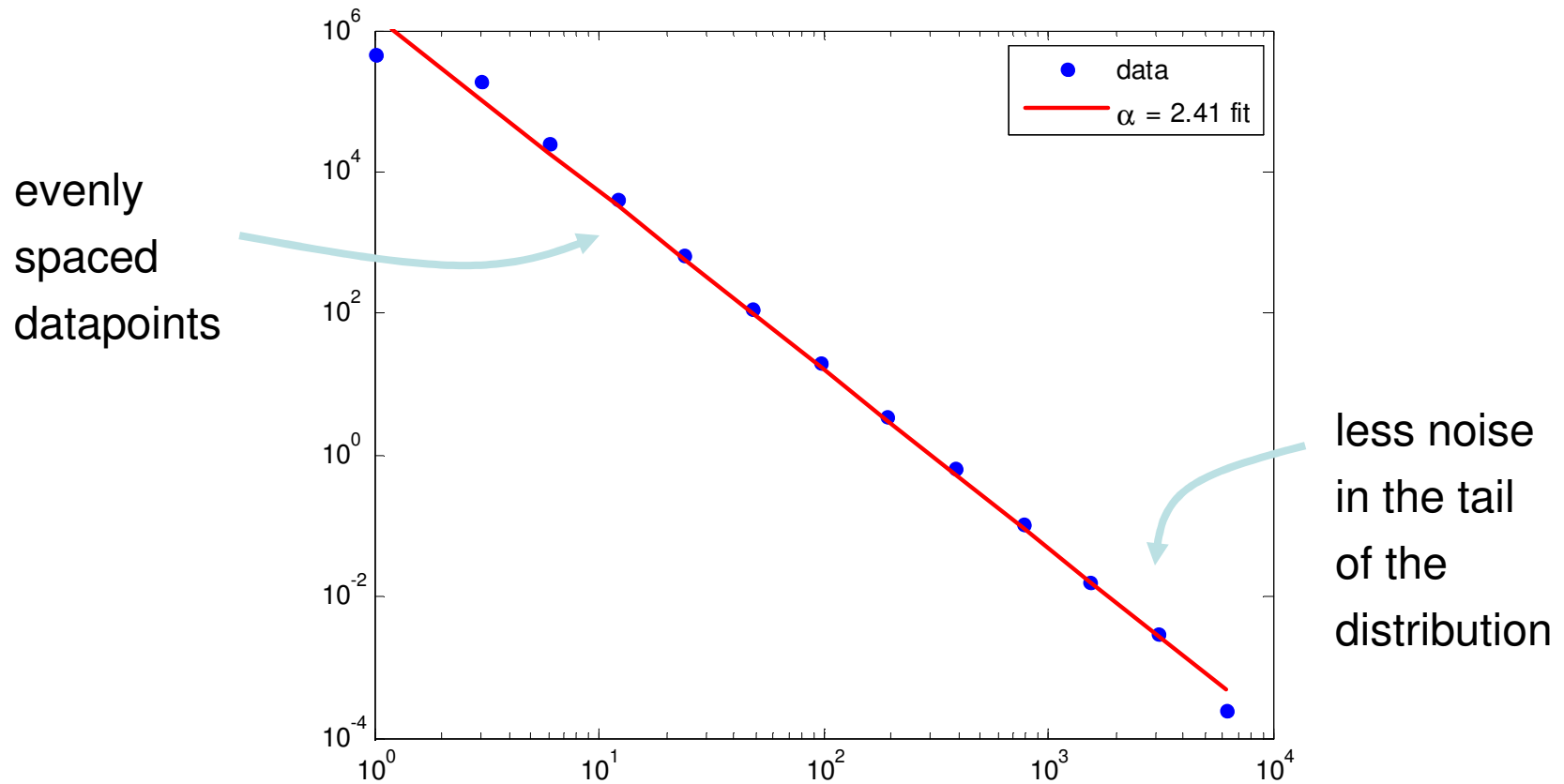


- Quantas vezes os números 1 ou 3843 ou 99723 ocorrem
- Lei de potência não é aparente
- Só faz sentido olhar nos menores bins



# Primeira solução

- Agrupe (“bin”) os dados em “bins” exponencialmente mais largos
  - 1, 2, 4, 8, 16, 32, ...
- Normalize pela largura do “bin”



- desvantagem: binning suaviza o dado mas perde informação

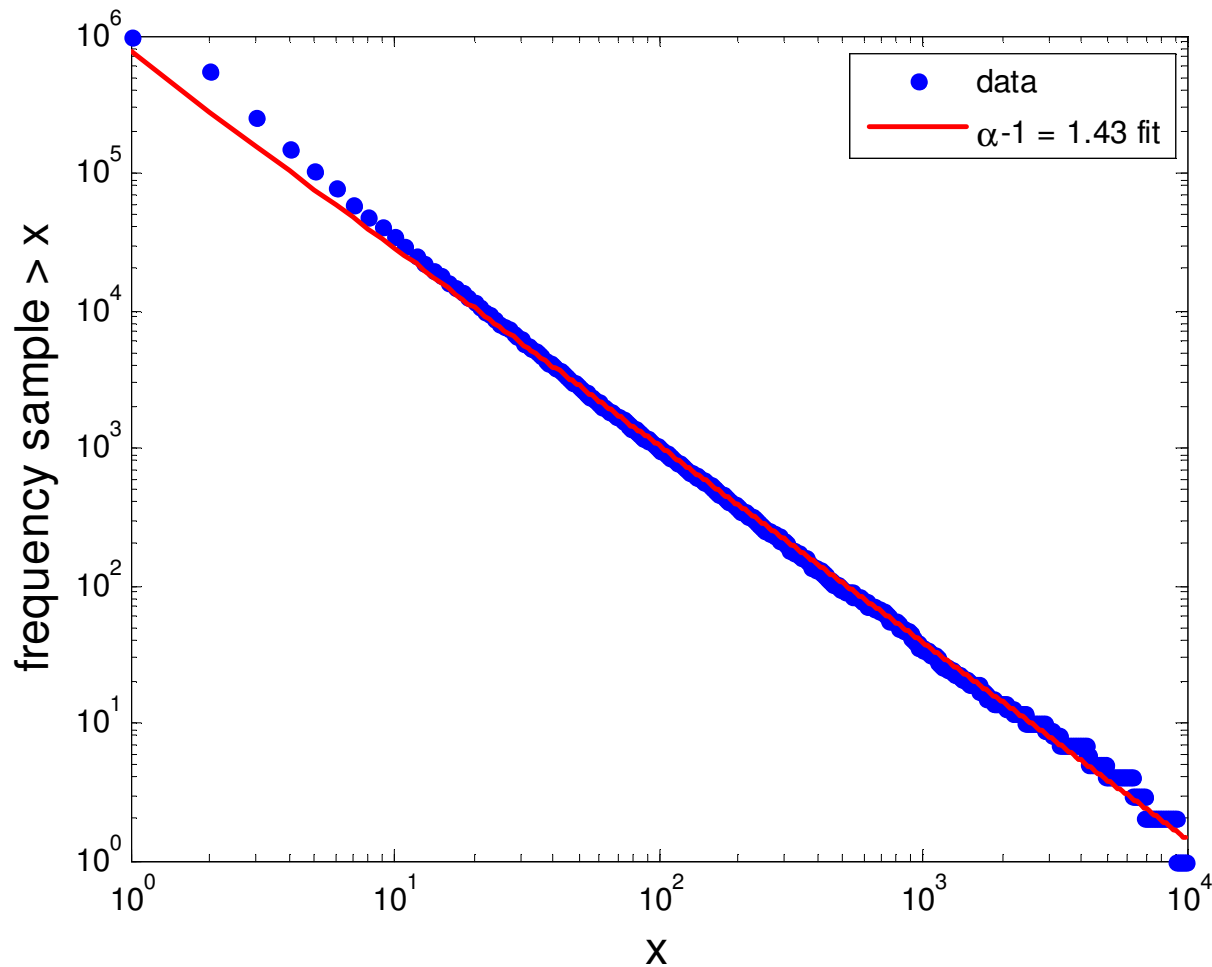
# Segunda solução

- Sem perda de informação
  - No need to bin, has value at each observed value of  $x$
- Usar distribuição acumulada Ou desacumulada? hehehe
  - i.e. quantos valores de  $x$  são pelo menos iguais a  $X$
  - A probabilidade acumulada de uma lei de potência é também uma lei de potência, mas com expoente  $\alpha - 1$

$$\int cx^{-\alpha} = \frac{c}{1-\alpha} x^{-(\alpha-1)}$$

# Agora a regressão funciona

- Expoente *fitted* (2.43) mais próximo do real (2.5)

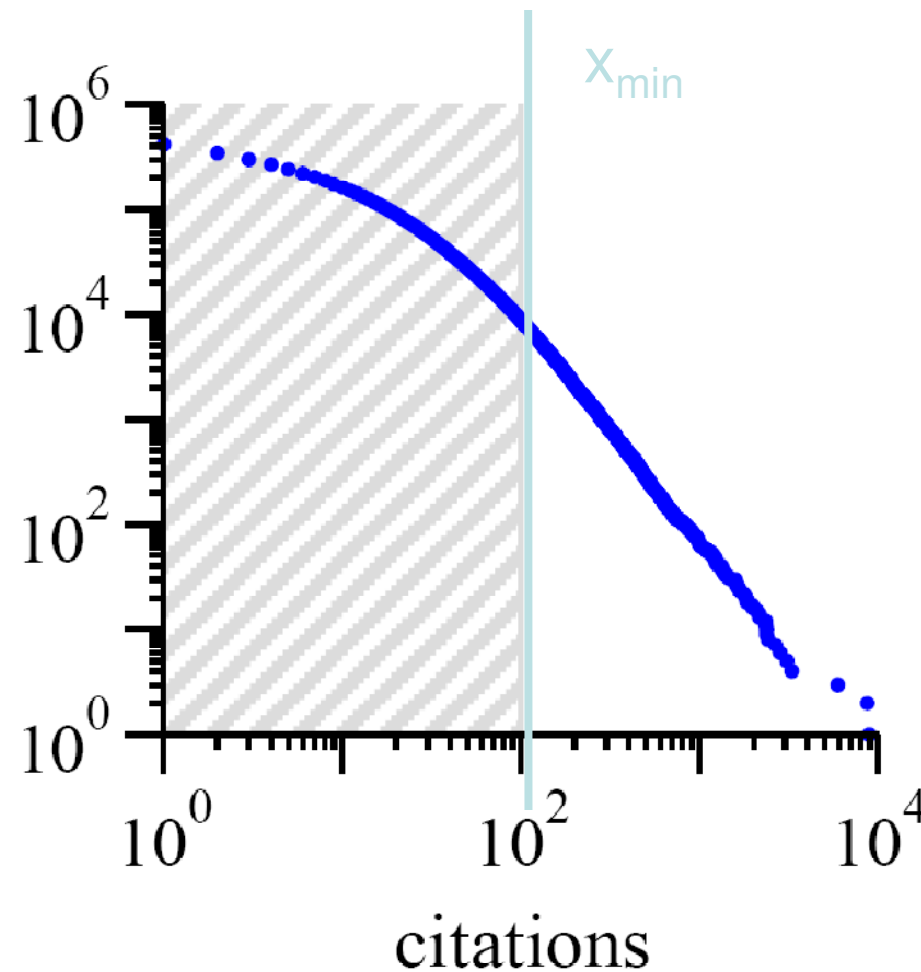


# Onde começar o fitting?

- Alguns conjuntos exibem a lei de potência apenas na cauda, o mais comum...
- Depois de fazer o "binning" ou usar uma distribuição acumulada, pode se iniciar na cauda o fitting.
- Então precisa-se de um  $x_{\min}$  o valor de  $x$  onde voce pensa que a "power-law" começa
- Certamente  $x_{\min}$  precisa ser maior que 0, pois  $x^{-a}$  é infinito em  $x = 0$

# Exemplo

- Distribuição de citações a artigos
- power law é evidente apenas na cauda ( $x_{\min} > 100$  citações)





# Bom método: MLE

Máxima Verossimilhança

- Você deve certificar se que tem uma distribuição com lei de potência (este método calcula o expoente mas não “goodness of fit”)

$$\alpha = 1 \square n \left[ \sum_{i=1}^n \ln \frac{x_i}{x_{\min}} \right]^{-1}$$

- $x_i$  são todos pontos de dados e você tem  $n$  deles
- Para o exemplo temos:  $\alpha = 2.503$  – bem perto!

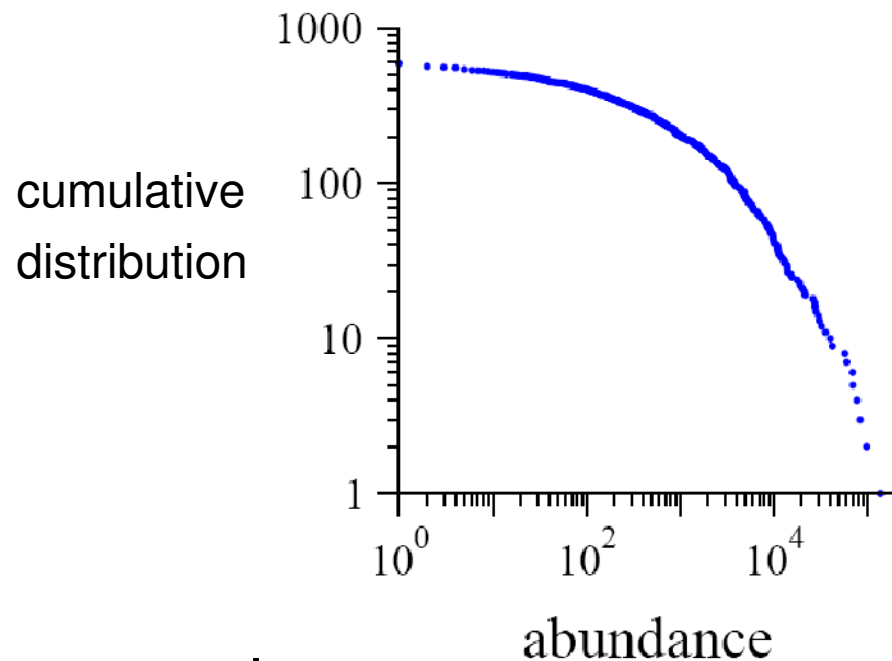
	$x_{\min}$	expoente $\alpha$
frequency of use of words	1	2.20
number of citations to papers	100	3.04
number of hits on web sites	1	2.40
copies of books sold in the US	2 000 000	3.51
telephone calls received	10	2.22
magnitude of earthquakes	3.8	3.04
diameter of moon craters	0.01	3.14
intensity of solar flares	200	1.83
intensity of wars	3	1.80
net worth of Americans	\$600m	2.09
frequency of family names	10 000	1.94
population of US cities	40 000	2.30

# Redes reais com lei de potência

	exponent $\alpha$ (in/out degree)
film actors	2.3
telephone call graph	2.1
email networks	1.5/2.0
sexual contacts	3.2
WWW	2.3/2.7
internet	2.5
peer-to-peer	2.1
metabolic network	2.2
protein interactions	2.4

# Nem tudo é uma "power law"

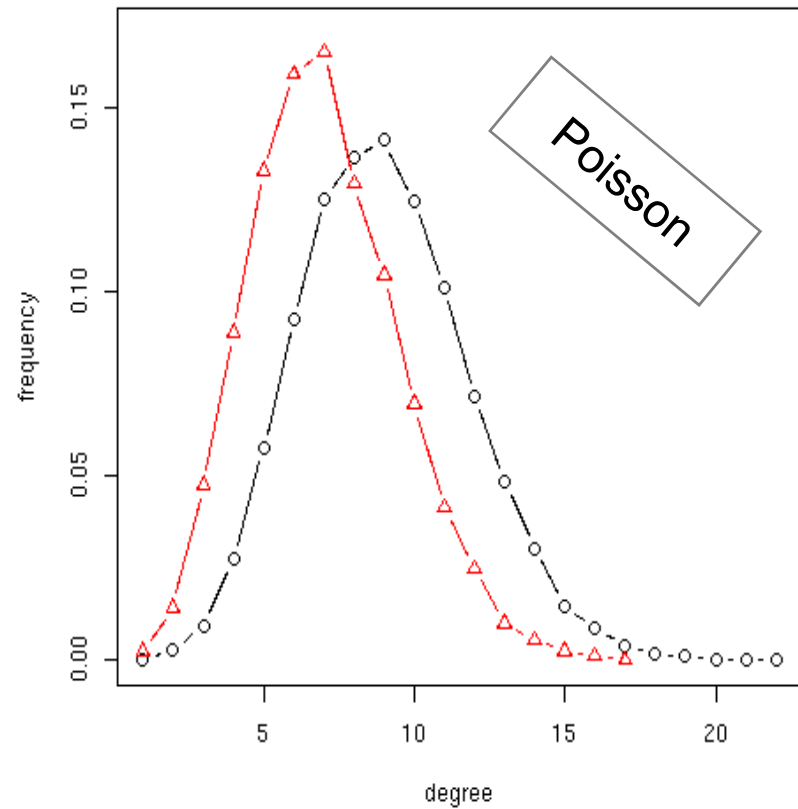
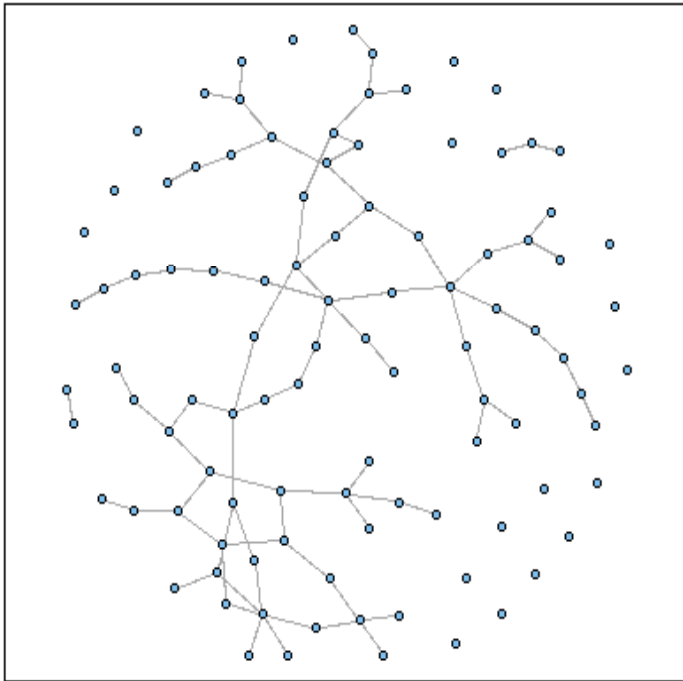
- Número de pássaros das 591 espécies vistos na América do Norte em 2003.



- Outro exemplo:
  - Tamanho dos incêndios florestais (em acres)

# Um exemplo de grafo randomico

Rede Aleatória ----- Distribuição de Gráus



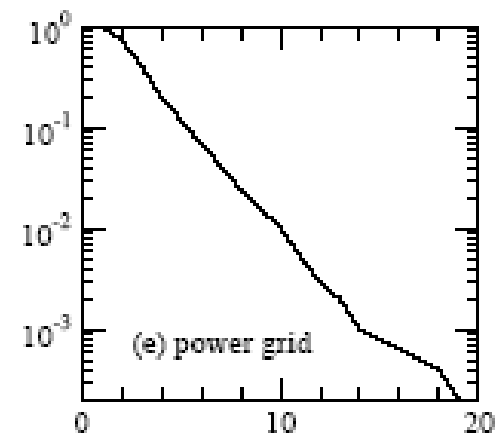
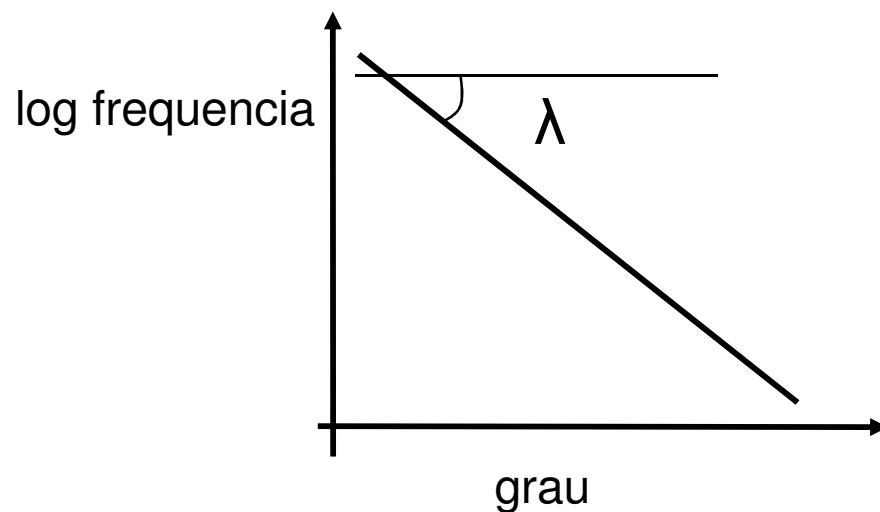
# Distribuição Exponencial

- Observada em algumas redes tecnológicas ou de colaboração

$$p(k) = \lambda e^{-\lambda k}$$

- Identificada por uma reta no gráfico log-linear.

$$\log p(k) = -\lambda k + \log \lambda$$



# Estatísticas Coletivas (M. Newman 2003)

	network	type	$n$	$m$	$z$	$\ell$	$\alpha$
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3
	company directors	undirected	7 673	55 392	14.44	4.60	-
	math coauthorship	undirected	253 339	496 489	3.92	7.57	-
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	-
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	-
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0
	email address books	directed	16 881	57 029	3.38	5.22	-
	student relationships	undirected	573	477	1.66	16.01	-
sexual contacts	undirected	2 810				3.2	
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7
	citation network	directed	783 339	6 716 198	8.57		3.0/-
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	-
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5
	power grid	undirected	4 941	6 594	2.67	18.99	-
	train routes	undirected	587	19 603	66.79	2.16	-
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4
	software classes	directed	1 377	2 213	1.61	1.51	-
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4
	marine food web	directed	135	598	4.43	2.05	-
	freshwater food web	directed	92	997	10.84	1.90	-
	neural network	directed	307	2 359	7.68	3.97	-

# Clustering coefficient (Transitividade)

número de arestas  
na vizinhança de  $v$

- Mede a densidade de triângulos (local clusters) no grafo
- Duas formas de medir:

$$C_v = \frac{|E_v|}{\binom{k_v}{2}} = \frac{2e_v}{k_v * (k_v - 1)}$$

grau do  
vértice

- A razão das médias

$$C^{(2)} = \frac{1}{n} C_v$$

$$C^{(1)} = \frac{\sum_i \text{triangulos centrados no nó } i}{\sum_i \text{triplas centradas no nó } i}$$



# clustering coefficient local

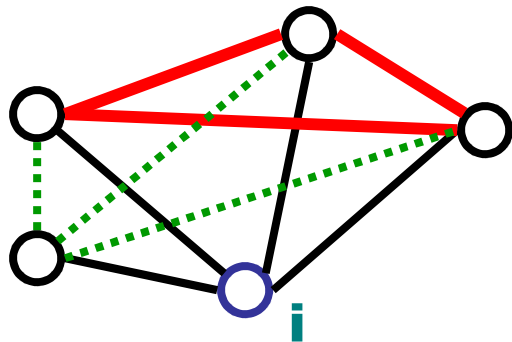
$$n_i = 4$$

Número max de conexões

$$4 * 3 / 2 = 6$$

**3** conexões presentes

$$C_i = 3 / 6 = 0.5$$



- link presente
- - - link ausente

## Clustering coefficient para grafos randomicos

- A probabilidade de dois de seus vizinhos serem tambem vizinhos é  $p$ , independente da estrutura local
  - clustering coefficient  $C = p$

Table 1: Clustering coefficients,  $C$ , for a number of different networks;  $n$  is the number of node,  $z$  is the mean degree. Taken from [146].

Network	$n$	$z$	$C$ measured	$C$ for random graph
Internet [153]	6,374	3.8	0.24	0.00060
World Wide Web (sites) [2]	153,127	35.2	0.11	0.00023
power grid [192]	4,941	2.7	0.080	0.00054
biology collaborations [140]	1,520,251	15.5	0.081	0.000010
mathematics collaborations [141]	253,339	3.9	0.15	0.000015
film actor collaborations [149]	449,913	113.4	0.20	0.00025
company directors [149]	7,673	14.4	0.59	0.0019
word co-occurrence [90]	460,902	70.1	0.44	0.00015
neural network [192]	282	14.0	0.28	0.049
metabolic network [69]	315	28.3	0.59	0.090
food web [138]	134	8.7	0.22	0.065

## Estatísticas Coletivas (M. Newman 2003)

	network	type	n	m	$C^{(1)}$	$C^{(2)}$
social	film actors	undirected	449 913	25 516 482	0.20	0.78
	company directors	undirected	7 673	55 392	0.59	0.88
	math coauthorship	undirected	253 339	496 489	0.15	0.34
	physics coauthorship	undirected	52 909	245 300	0.45	0.56
	biology coauthorship	undirected	1 520 251	11 803 064	0.088	0.60
	telephone call graph	undirected	47 000 000	80 000 000		
	email messages	directed	59 912	86 300		0.16
	email address books	directed	16 881	57 029	0.17	0.13
	student relationships	undirected	573	477	0.005	0.001
sexual contacts	undirected	2 810				
information	WWW nd.edu	directed	269 504	1 497 135	0.11	0.29
	WWW Altavista	directed	203 549 046	2 130 000 000		
	citation network	directed	783 339	6 716 198		
	Roget's Thesaurus	directed	1 022	5 103	0.13	0.15
	word co-occurrence	undirected	460 902	17 000 000		0.44
technological	Internet	undirected	10 697	31 992	0.035	0.39
	power grid	undirected	4 941	6 594	0.10	0.080
	train routes	undirected	587	19 603		0.69
	software packages	directed	1 439	1 723	0.070	0.082
	software classes	directed	1 377	2 213	0.033	0.012
	electronic circuits	undirected	24 097	53 248	0.010	0.030
	peer-to-peer network	undirected	880	1 296	0.012	0.011
biological	metabolic network	undirected	765	3 686	0.090	0.67
	protein interactions	undirected	2 115	2 240	0.072	0.071
	marine food web	directed	135	598	0.16	0.23
	freshwater food web	directed	92	997	0.20	0.087
	neural network	directed	307	2 359	0.18	0.28

# Métricas de rede

- $d_{ij}$  = menor caminho entre  $i$  e  $j$

- Diâmetro:

$$d = d_G = \max_{i,j} d_{ij}$$

- Comprimento médio do caminho:

$$\ell = \frac{1}{n(n-1)/2} \sum_{i>j} d_{ij}$$

- Média Harmônica

$$\ell^{-1} = \frac{1}{n(n-1)/2} \sum_{i>j} d_{ij}^{-1}$$

# Estatísticas Coletivas (M. Newman 2003)

	network	type	$n$	$m$	$z$	$\ell$
social	film actors	undirected	449 913	25 516 482	113.43	3.48
	company directors	undirected	7 673	55 392	14.44	4.60
	math coauthorship	undirected	253 339	496 489	3.92	7.57
	physics coauthorship	undirected	52 909	245 300	9.27	6.19
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92
	telephone call graph	undirected	47 000 000	80 000 000	3.16	
	email messages	directed	59 912	86 300	1.44	4.95
	email address books	directed	16 881	57 029	3.38	5.22
	student relationships	undirected	573	477	1.66	16.01
sexual contacts	undirected	2 810				
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18
	citation network	directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87
	word co-occurrence	undirected	460 902	17 000 000	70.13	
technological	Internet	undirected	10 697	31 992	5.98	3.31
	power grid	undirected	4 941	6 594	2.67	18.99
	train routes	undirected	587	19 603	66.79	2.16
	software packages	directed	1 439	1 723	1.20	2.42
	software classes	directed	1 377	2 213	1.61	1.51
	electronic circuits	undirected	24 097	53 248	4.34	11.05
	peer-to-peer network	undirected	880	1 296	1.47	4.28
biological	metabolic network	undirected	765	3 686	9.64	2.56
	protein interactions	undirected	2 115	2 240	2.12	6.80
	marine food web	directed	135	598	4.43	2.05
	freshwater food web	directed	92	997	10.84	1.90
	neural network	directed	307	2 359	7.68	3.97

# Mixing patterns

Paramos aqui!!!  
01/04/2009

- Assuma-se que se tem vários tipos de nós. Qual a probabilidade de dois nós de diferente tipos serem conectados? → assortative mixing

E : mixing matrix

$p(i,j)$  = probabilidade de mixing

$$p(i,j) = \frac{E(i,j)}{\sum_{i,j} E(i,j)}$$

$p(j | i)$  = probabilidade condicional de mixing

$$p(j|i) = \frac{E(i,j)}{\sum_j E(i,j)}$$

		women			
		black	hispanic	white	other
men	black	506	32	69	26
	hispanic	23	308	114	38
	white	26	46	599	68
	other	10	14	47	32

TABLE III Couples in the study of Catania *et al.* [85] tabulated by race of either partner. After Morris [302].

# Propriedades das redes

- Mixing patterns
  - a.k.a. assortative mixing
  - 1958 casais heteros em San Francisco:

		women			
		black	hispanic	white	other
men	black	506	32	69	26
	hispanic	23	308	114	38
	white	26	46	599	68
	other	10	14	47	32

# Mixing Coefficients

- Mixing Patterns

- Como quantificar?

- Seja  $E_{ij}$  = # de arestas conectando vertices tipos  $i$  e  $j$
- $\mathbf{E}$  = matriz com elements  $E_{ij}$
- Assim, a matriz normalizada

$$\mathbf{e} = \frac{\mathbf{E}}{\|\mathbf{E}\|}$$

- $\|\mathbf{E}\|$  = soma de todos elementos  $\mathbf{E}$
- Elemento  $e_{ij}$  de  $\mathbf{e}$  = fração de arestas entre  $i$  and  $j$



# Mixing coefficient

- Newman 2003

$$a_i = \sum_j p(i,j)$$

$$b_i = \sum_j p(j,i)$$

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}$$

- Características:

- $r = 1$  matriz diagonal  $\rightarrow \sum_i e_{ii} = 1$

- $r = 0$  matriz uniforme  $\rightarrow \sum_i e_{ii} = \sum_i a_i b_i$

		women				$a_i$
		black	hispanic	white	other	
men	black	0.258	0.016	0.035	0.013	0.323
	hispanic	0.012	0.157	0.058	0.019	0.247
	white	0.013	0.023	0.306	0.035	0.377
	other	0.005	0.007	0.024	0.016	0.053
$b_i$		0.289	0.204	0.423	0.084	

TABLE I: The mixing matrix  $e_{ij}$  and the values of  $a_i$  and  $b_i$  for sexual partnerships in the study of Catania *et al.* [23]. After Morris [24].

$r=0.621$

Altamente Assortativa, Por que? Quais fatores?

# Correlação de graus

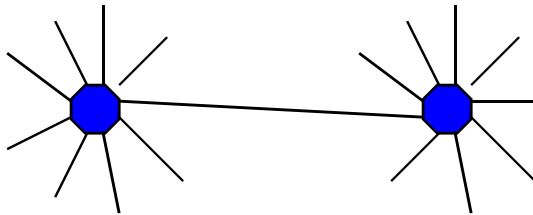
- Nós de graus altos tendem a se conectar a nós altos?
- Pastor Satoras et al.
  - Exercício: Para os dados que voce coletou, faça um gráfico do grau médio dos vizinhos em função do grau.
- Newman
  - Compute o coeficiente de correlação dos graus de duas extremidades de uma aresta
  - assortative/disassortative

# Assortatividade

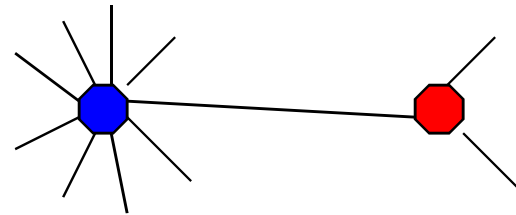
Primeira aproximação

Propriedade do nó = seu **grau**

**Assortative networks**



**Disassortative networks**



- Redes reais sempre exibem uma das duas tendências,
- redes “similares” exibem comportamentos “similares”.



Social networks



Techological networks

# Medidas de Correlação Grau-a-Grau

- Grau medio dos vizinhos proximos de nós de grau  $k$
- Probabilidade que um vertice  $k$  esteja conectado a  $k'$   $P(k'|k)$
- Probabilidade  $P(k) = N_k / N$
- Para efeitos de cálculo em redes reais, é mais fácil trabalhar com o grau medio dos vizinhos mais proximos de  $k$ , definido por:

$$Knn(k) = \sum_{k'} k' P(k'|k)$$

- Para se analisar *assortative*, faz em função de  $k$ , ou seja

Se  $knn$  é uma função crescente de  $k$ : →  
Assortative

Se  $knn$  é uma função decrescente de  $k$ : →  
Disassortative

Se nenhum dos casos acima: → Non assortative

# Conectividade da Internet

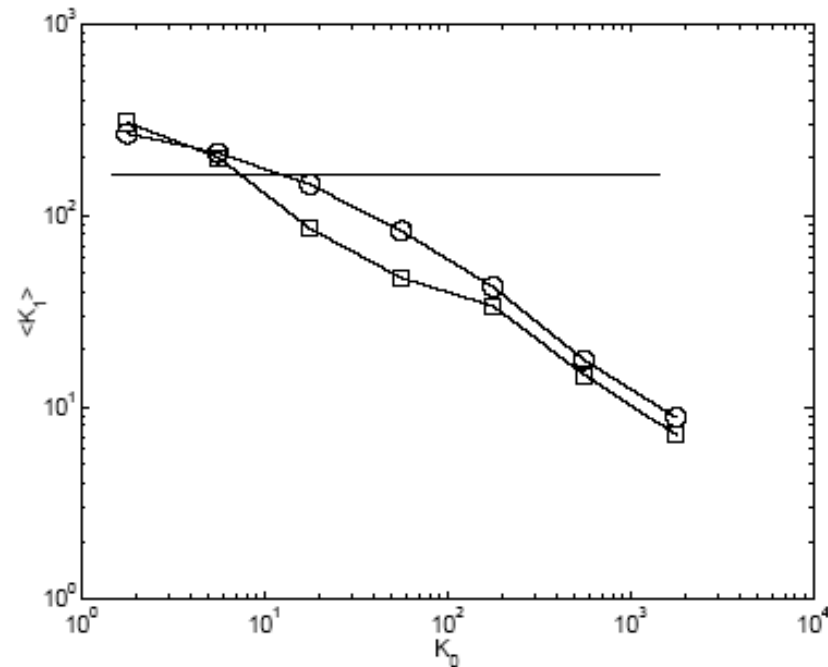
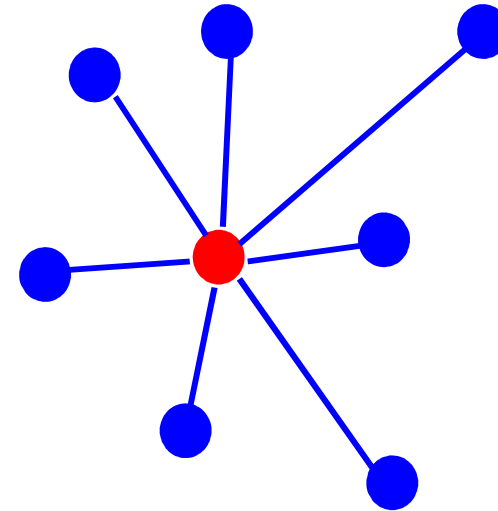


FIG. 2. The average connectivity  $\langle K_1 \rangle_{K_0}$  of neighbors of nodes with connectivity  $K_0$  in the Internet (squares) and its typical randomized counterpart (circles). Error bars in multiple realizations of the randomized network are smaller than symbol sizes. The horizontal line is the analytical result  $\langle K_1 \rangle_{K_0} = \text{const} = \langle K^2 \rangle / \langle K \rangle \simeq 165$  valid for a random network in which multiple edges between pairs of nodes are allowed [2].

# Relações de conectividade

Pastor Satorras, Vazquez & Vespignani,  
PRL 87, 258701 (2001)

Media do grau dos vizinhos mais próximos



$$\langle k_{nn}(k) \rangle = \sum_{k'} k' p(k'|k)$$

← Função de correlação de graus

# Hierarquia de conectividade

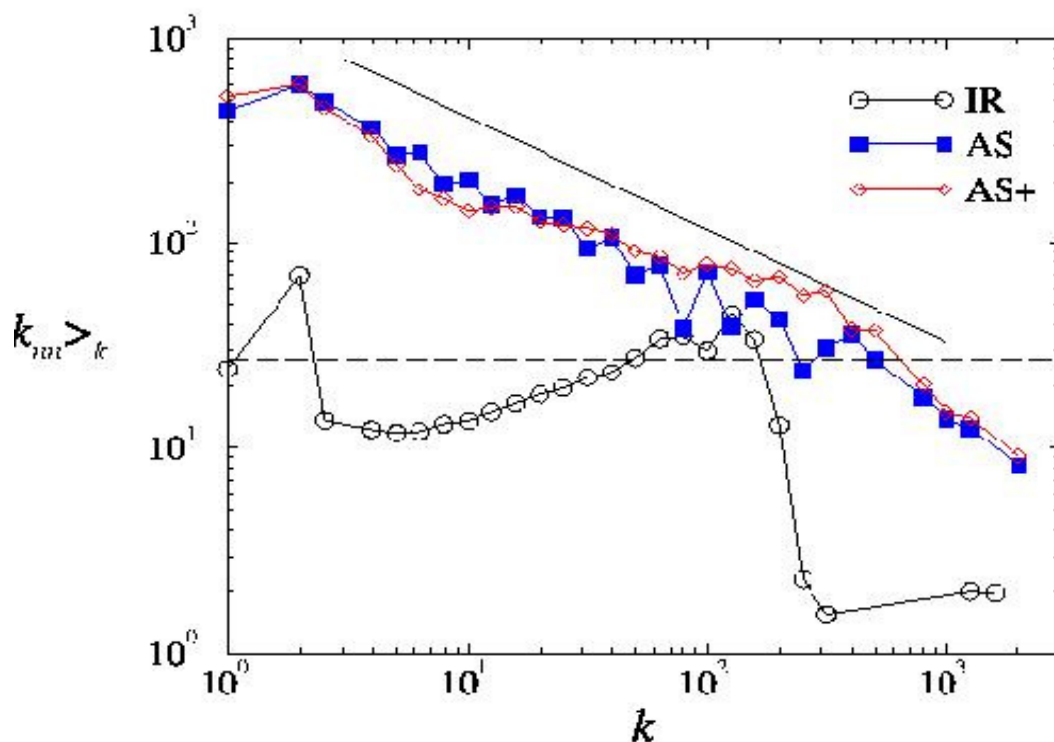
Média do grau dos vizinhos proximos

$$\langle k_{nn}(k) \rangle = \sum_{k'} k' p(k'|k)$$

← Degree correlation function

Alto grau de ASs conecta  
a um baixo grau de ASs

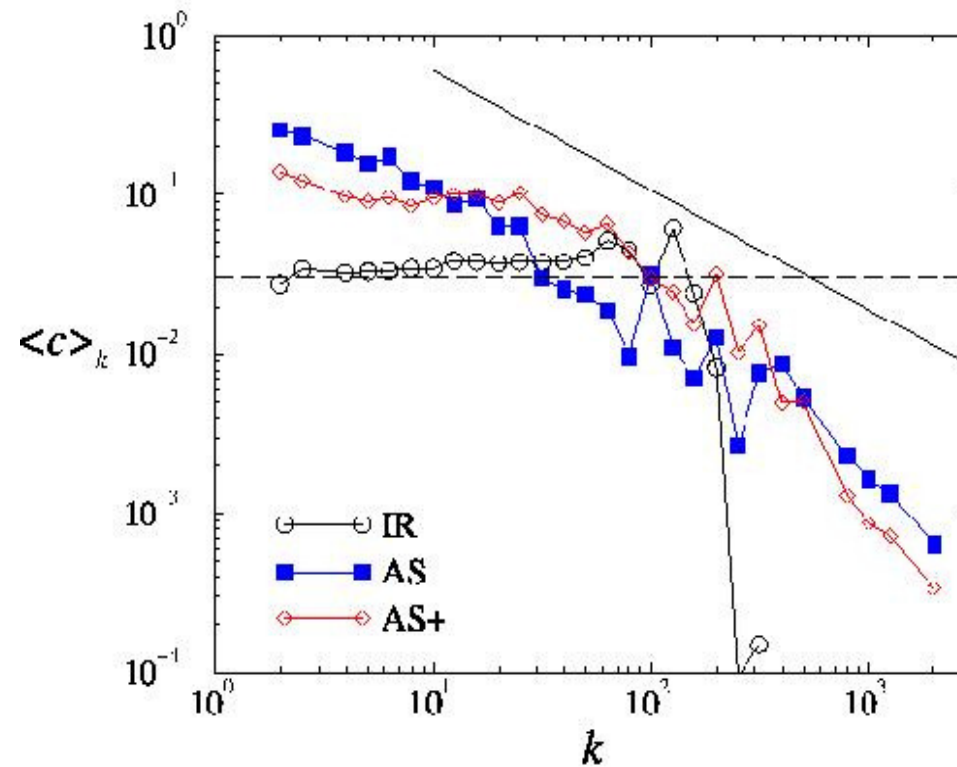
Baixo grau de ASs  
conecta a ASs com  
alto grau



Não há hierarquia para mapeamento de roteadores

# Hierarquia de "Clustering"

CC em função do grau do vértice



**ASs de baixo grau tem links regiões da Internet altamente conectadas! Por que?**

Nível de roteador, não há hierarquia! Por que?



	network	type	size $n$	assortativity $r$	error $\sigma_r$
social	physics coauthorship	undirected	52 909	0.363	0.002
	biology coauthorship	undirected	1 520 251	0.127	0.0004
	mathematics coauthorship	undirected	253 339	0.120	0.002
	film actor collaborations	undirected	449 913	0.208	0.0002
	company directors	undirected	7 673	0.276	0.004
	student relationships	undirected	573	-0.029	0.037
	email address books	directed	16 881	0.092	0.004
technological	power grid	undirected	4 941	-0.003	0.013
	Internet	undirected	10 697	-0.189	0.002
	World-Wide Web	directed	269 504	-0.067	0.0002
	software dependencies	directed	3 162	-0.016	0.020
biological	protein interactions	undirected	2 115	-0.156	0.010
	metabolic network	undirected	765	-0.240	0.007
	neural network	directed	307	-0.226	0.016
	marine food web	directed	134	-0.263	0.037
	freshwater food web	directed	92	-0.326	0.031

**Consequencias da assortatividade:** - Disseminação de Epidemias, Comunidades Isoladas....

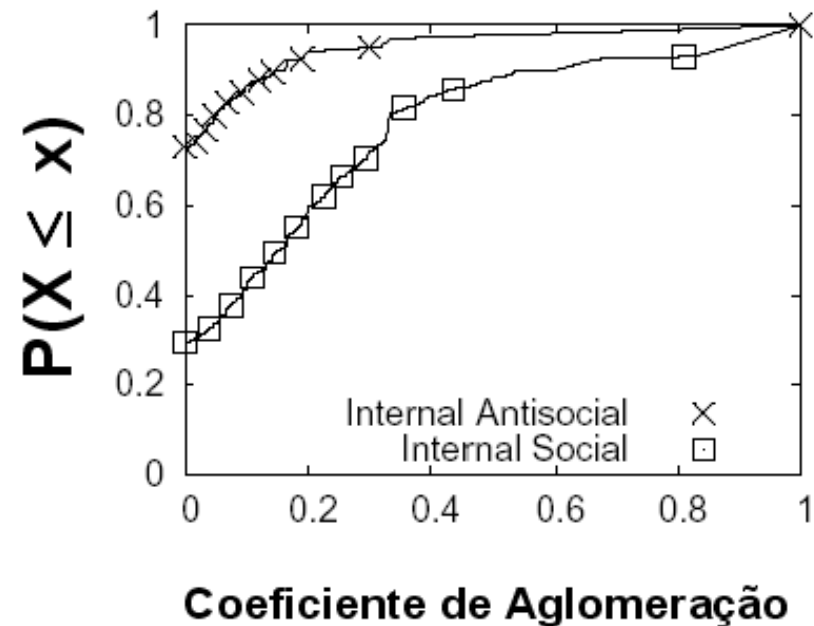
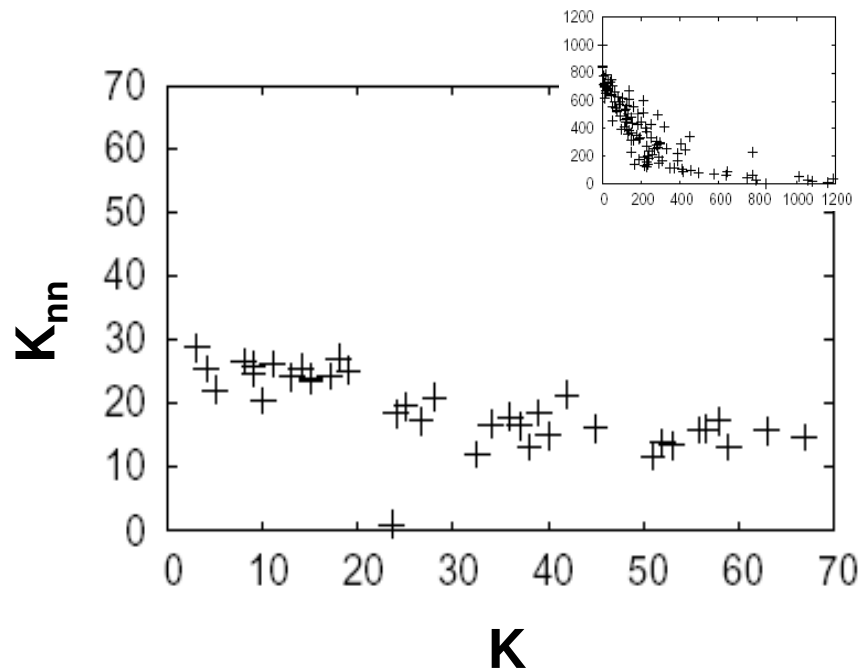
➤ Newman, *PRE*, bf 67 : 026126 , (2003).

# Quantifying social vs. antisocial behavior in email networks

Gomes, Luiz H.; Bettencourt, Luis M. A.; Almeida, Virgilio A. F.;

Almeida, Jussara M.; Castro, Fernando D. O.

<http://adsabs.harvard.edu/abs/2006physics...1141G>

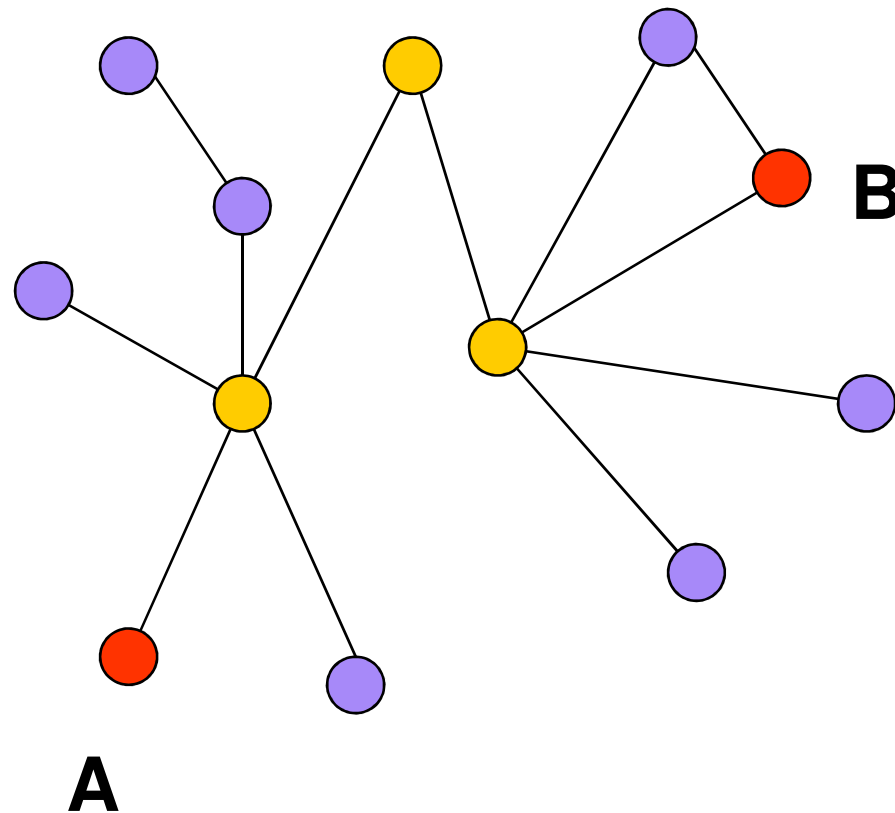


# Estatísticas Coletivas (M. Newman 2003)

	network	type	$n$	$m$	$z$	$\ell$	$\alpha$	$C^{(1)}$	$C^{(2)}$	$r$	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	–0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	–0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	–0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	–0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	416, 421

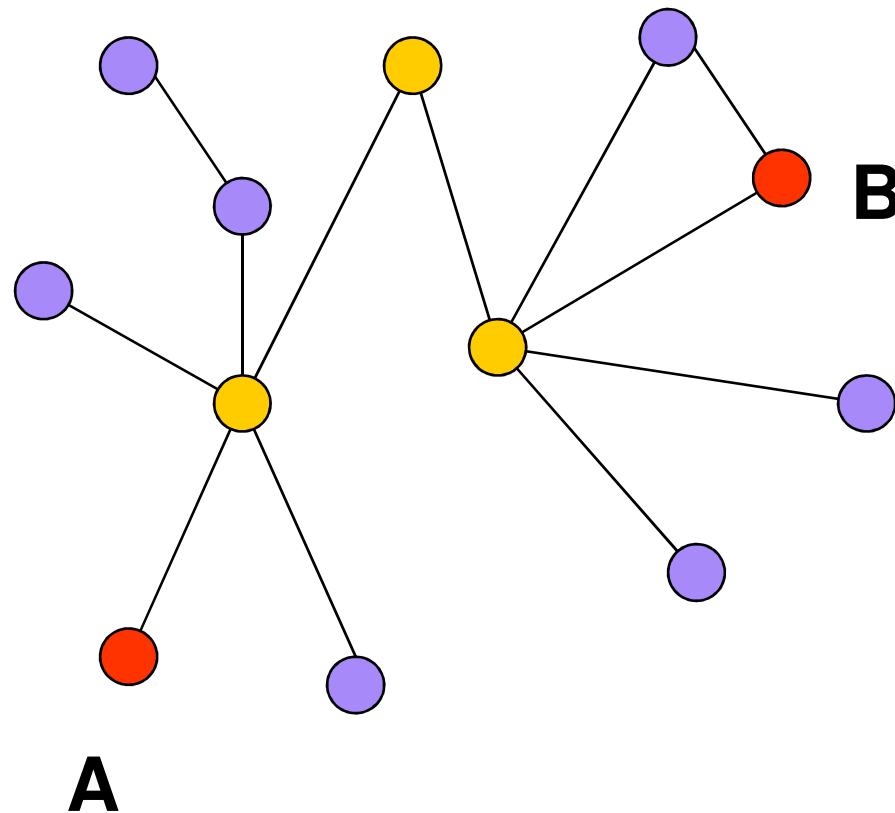
TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices  $n$ ; total number of edges  $m$ ; mean degree  $z$ ; mean vertex–vertex distance  $\ell$ ; exponent  $\alpha$  of degree distribution if the distribution follows a power law (or  $\alpha_{in}$ / $\alpha_{out}$  if not; in/out-degree exponents are given for directed graphs); clustering coefficient  $C^{(1)}$  from Eq. (3); clustering coefficient  $C^{(2)}$  from Eq. (6); and degree correlation coefficient  $r$ , Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

## *Betweenness* mede a centralidade de nós



O **betweenness**  $b_i$  do nó  $i$  é o número de caminhos mínimos entre pares de nós que passa pelo nó  $i$ .

## Betweenness mede a centralidade de nós

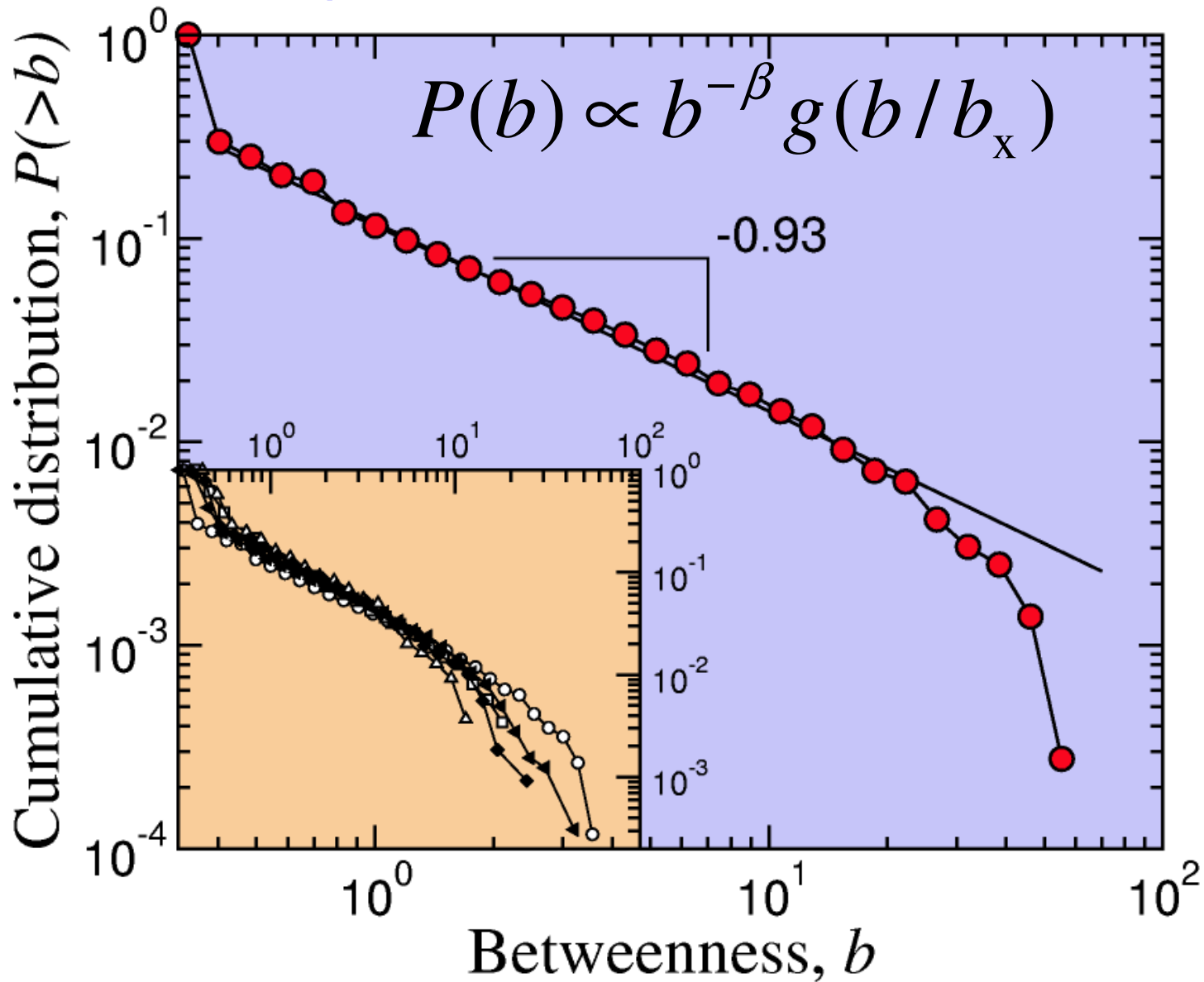


*betweenness = load = Betweenness centrality*

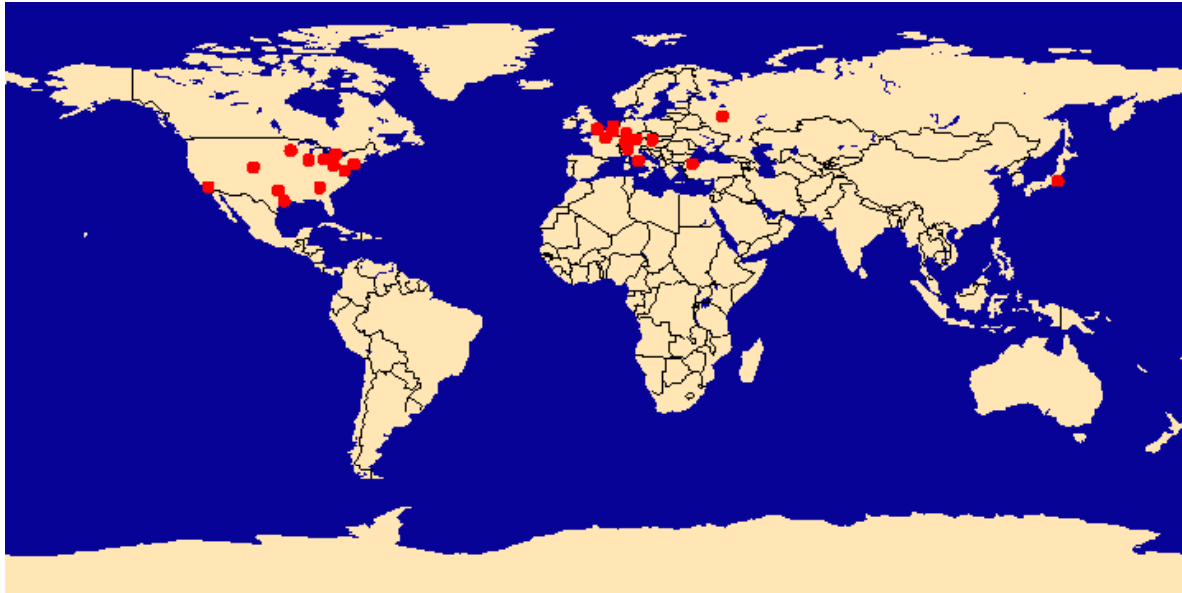
$L_{h,j}$  no. Total de shortest paths de  $h$  a  $j$  e  $L_{h,i,j}$  é o no. total de shortest paths que passa pelo vértice  $i$ . Assim, ,

$$b_i = \sum L_{h,i,j} / L_{h,j}$$

The betweenness distribution also follows a truncated power law

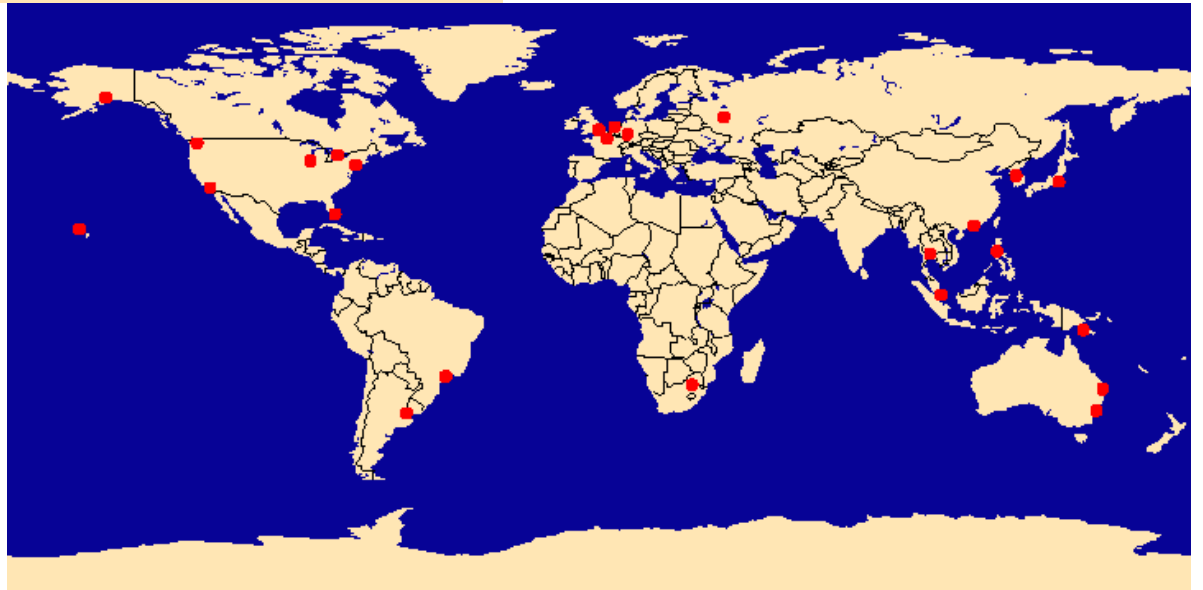


Aeroportos: cidades mais conectadas não são as mais centrais

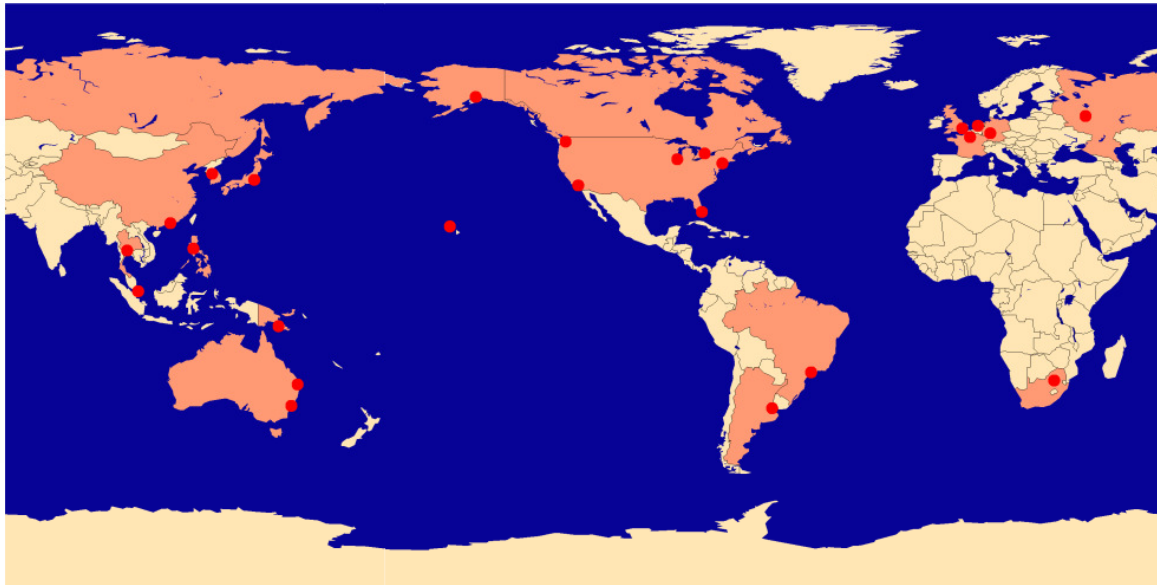


Cidades Mais  
**conectadas**

Mais  
**centrais**



Cidades mais centrais são importantes!



Cidades  
mais  
**centrais**



# Referências

- M. E. J. Newman, *The structure and function of complex networks*, SIAM Reviews, 45(2): 167-256, 2003
- M. E. J. Newman, *Random graphs as models of networks* in *Handbook of Graphs and Networks*, S. Bornholdt and H. G. Schuster (eds.), Wiley-VCH, Berlin (2003).
- N. Alon J. Spencer, *The Probabilistic Method*

1. Pode ser 8,14 (se vc assumir que convidados não são parte do grafo ) ou 2,5,14 (o contrario)

2. 0 0 1 2 2 2 2 1 1 1 1 2 2 2

3. Joana-Bruno-Raquel-Tiago (ou inverso)

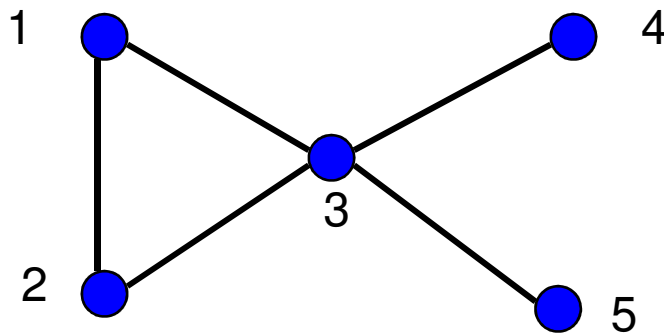
# Is the path length enough?

- Random graphs have diameter

$$d = \frac{\log n}{\log z}$$

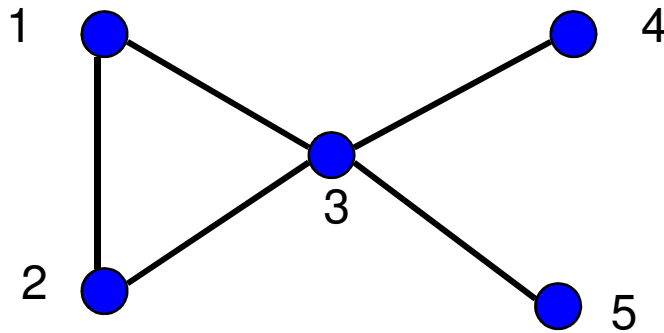
- $d = \log n / \log \log n$  when  $z = \omega(\log n)$
- Short paths should be combined with other properties
  - ease of navigation
  - high clustering coefficient

# Exemplo



$$C^{(1)} = \frac{3}{1+1+6} = \frac{3}{8}$$

## Exemplo



$$C^{(2)} = \frac{1}{5} (1 + 1 + 1/6) = \frac{13}{30}$$

$$C^{(1)} = \frac{3}{8}$$

- Os dois clustering coefficients levam a medidas diferentes

# Exercício - entrega 1/09/2008

- Use os dados de atores, referentes aos primeiros 100.000 filmes da seguinte rede:
- <http://www.nd.edu/~networks/resources.htm>
- Crie o grafo da rede
  - Calcule as principais métricas da rede:
  - Total de vértice e arestas
  - Grau médio da rede
  - Distribuição de graus
  - Caminho (não orientado) mínimo médio
  - Coeficiente de agrupamento (clustering coefficient) médio da rede
  - Tamanho do maior componente fortemente conectado
- Apresente uma visualização da rede
- Os resultados devem ser apresentados no formato de um ``paper'' no formato definido em:

<http://www.ieee.org/web/publications/pubservices/confpub/infocom08#templates>