

Introdução a Redes Complexas

Jornadas de Atualização em Informática (JAI)
CSBC 2011

Encontro 1/3

Daniel R. Figueiredo
LAND – PESCCOPPE/UFRJ



Do que trata Redes Complexas?



- ❑ Entender como as coisas se conectam e as consequências desta conectividade

“Coisas que se conectam” → **Redes**

“Como, por que, e consequências” → **Complexo**

- ❑ **Estrutura** assume papel central
 - necessária para compreender fenômenos que ainda não explicamos
- ❑ A rede não é complexa!

Exemplo com Twitter



- ☐ Dezenas de milhões de pessoas, centenas de milhões de seguidos
- ☐ Rede de quem segue quem
- ☐ Qual é sua *estrutura*? Como ela vem evoluindo? Por que?
- ☐ Qual é o impacto da estrutura na funcionalidade (disseminar informação)? Como ter *voz* no Twitter?



Veremos muitos exemplos!

Objetivo do Mini-curso

- ❑ Primeira exposição a esta maneira de abordar problemas
 - estrutura no centro
- ❑ Apresentar descobrimentos empíricos
 - caracterização de diferentes redes, propriedades peculiares de muitas redes
- ❑ Modelos matemáticos que representam estruturas
 - intuições e propriedades
- ❑ Apresentar alguns fenômenos importantes
 - funcionalidade e processos em redes

Despertar sua curiosidade!

Organização do Mini-curso

- ❑ Três encontros (Qui 17h, Sex 11h, Sex 17h)

- duas horas cada, brinde na Sex 17h

- ❑ **Encontro 1**

- Redes por todos os lados, formalismo e propriedades, descobrimentos empíricos

- ❑ **Encontro 2**

- Lei de potência, modelos de redes: $G(n,p)$, Barabási-Albert, Watts-Strogatz, propriedade dos modelos

- ❑ **Encontro 3**

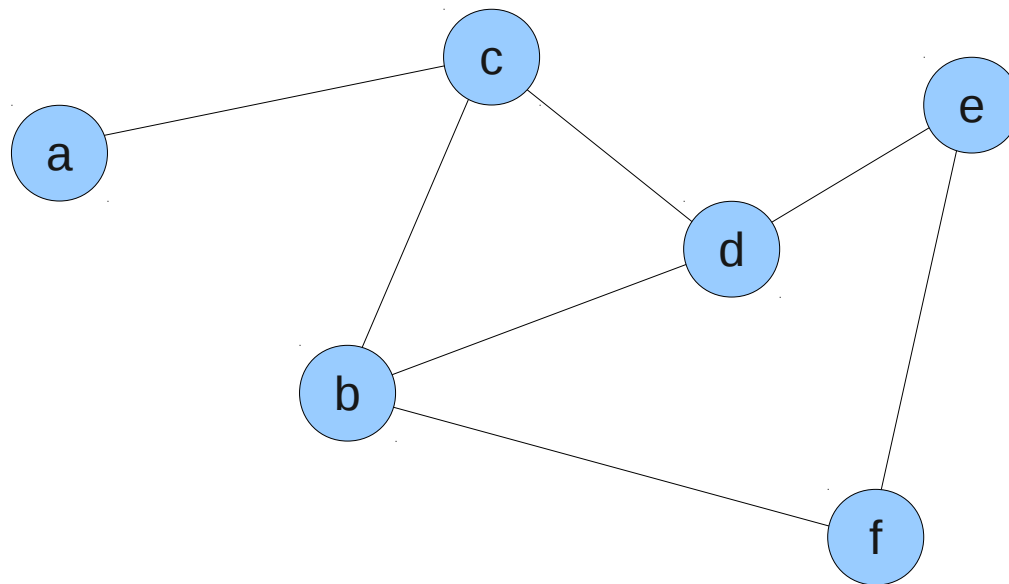
- Funcionalidade em redes: robustez, busca e navegabilidade

Interatividade com participação do público!

Redes

O que é uma rede?

- ❑ Conjunto de pontos interligados por segmentos de retas

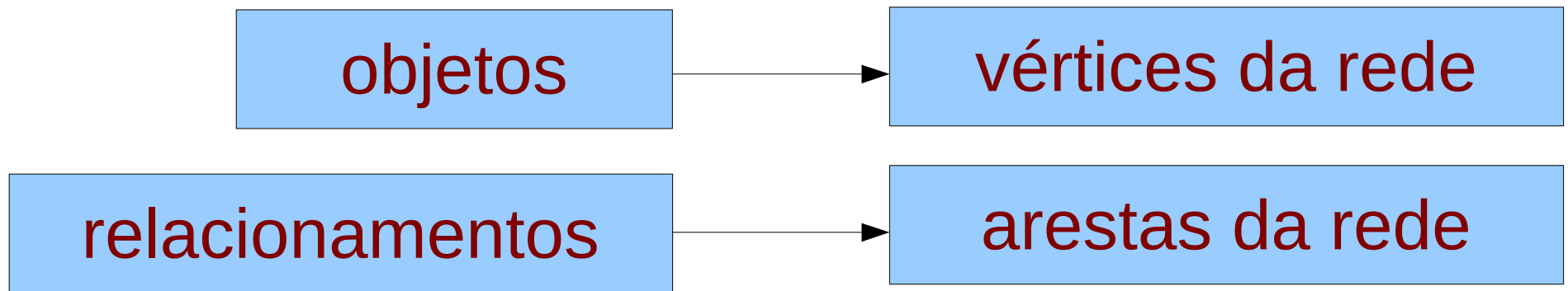


← **Definição
burocrática!**



Redes, outra definição

- ❑ Abstração que permite codificar relacionamento entre **pares** de objetos

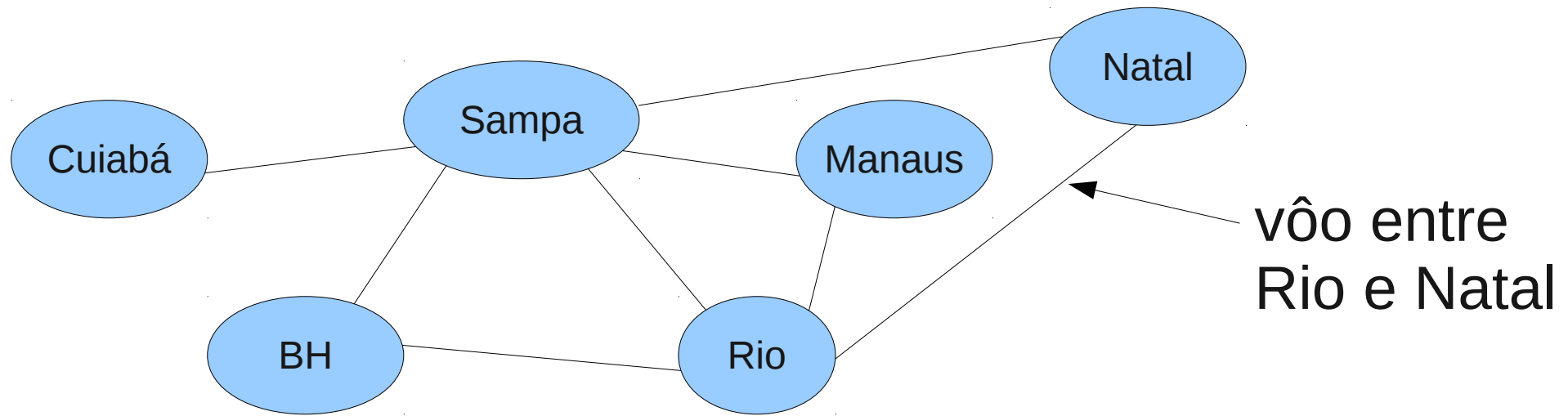


- ❑ Redes como ferramenta de **abstração matemática**
- ❑ Na computação, redes é chamada de grafos



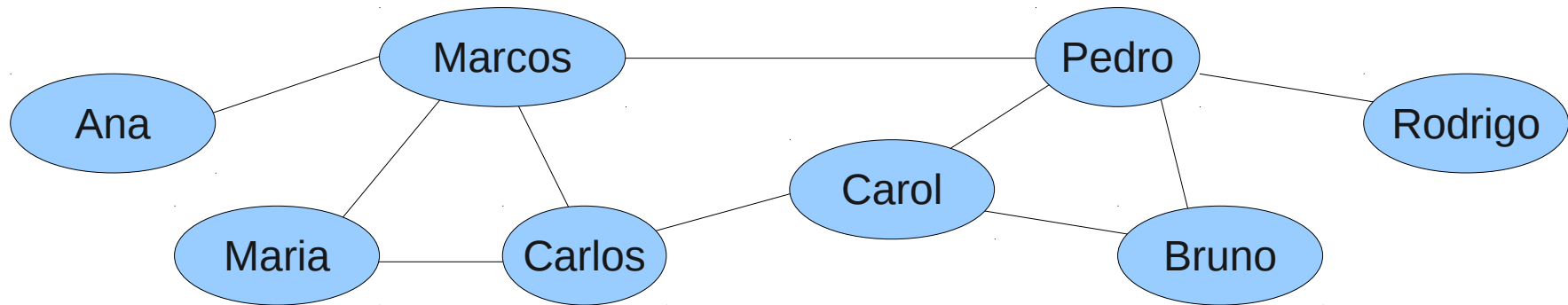
Rede de Transporte Aéreo

- ❑ Objeto: cidades
- ❑ Relacionamento: voo comercial entre cidades

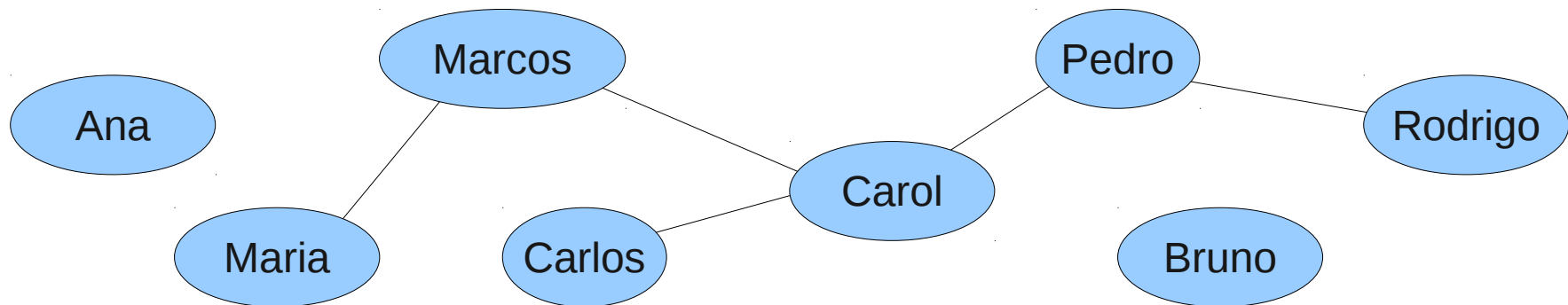


Redes Sociais

- ❑ Objeto: pessoas
- ❑ Relacionamento: ser amigo no Facebook



- ❑ Outro relacionamento: ter se beijado



**Diferentes relacionamentos sobre
mesmo conjunto de objetos**

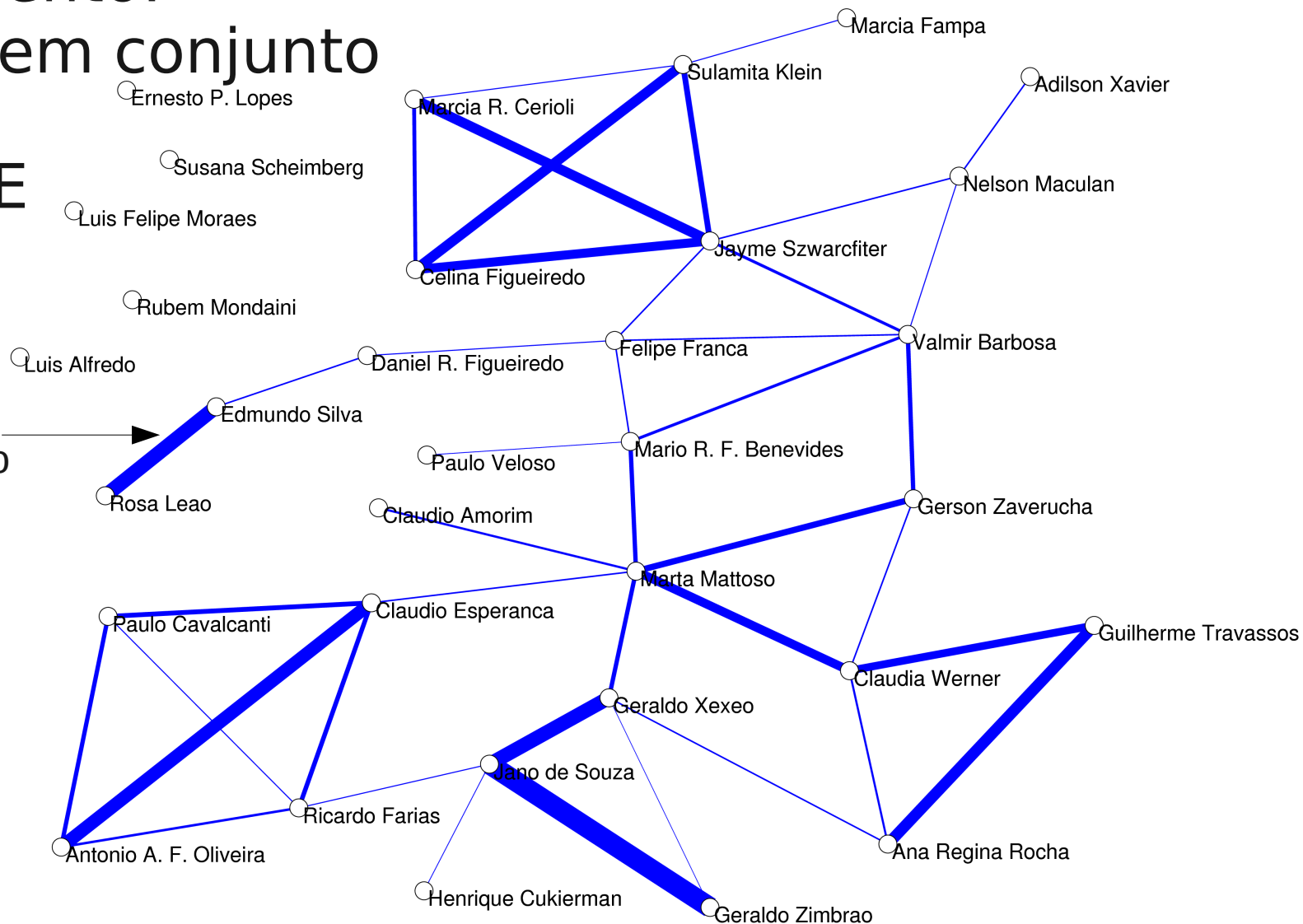
Rede de Colaboração

- Objeto: pesquisadores
- Relacionamento: publicação em conjunto

PESC/COPPE (DBLP)

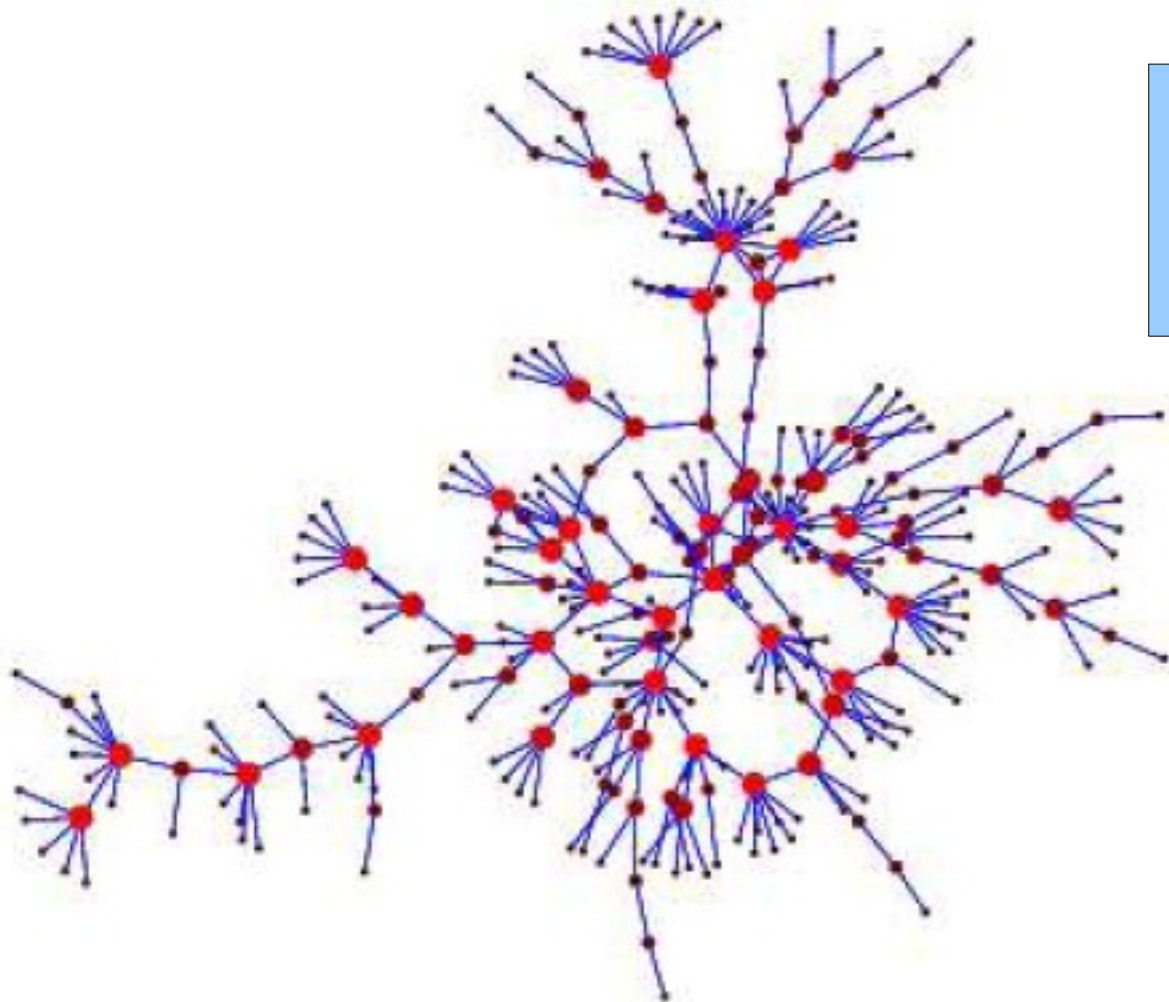
Intensidade do relacionamento →

O que estrutura nos diz?



Rede de Contato Sexual

- ❑ Objeto: pessoas
- ❑ Relacionamento: tiveram uma relação sexual



Importância da estrutura?

Rede da Web

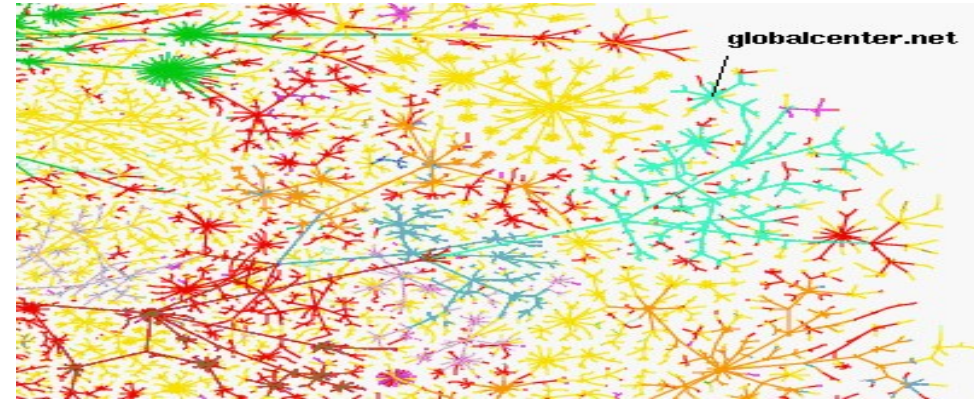
- ❑ Páginas web e hiperlinks entre elas
- ❑ Relacionamento assimétrico (*hiperlinks*)



- ❑ 20+ bilhões de páginas
(worldwidewebsite.com)

Ranqueamento de Páginas

- ❑ Importante função na Web
 - ranquear páginas por relevância

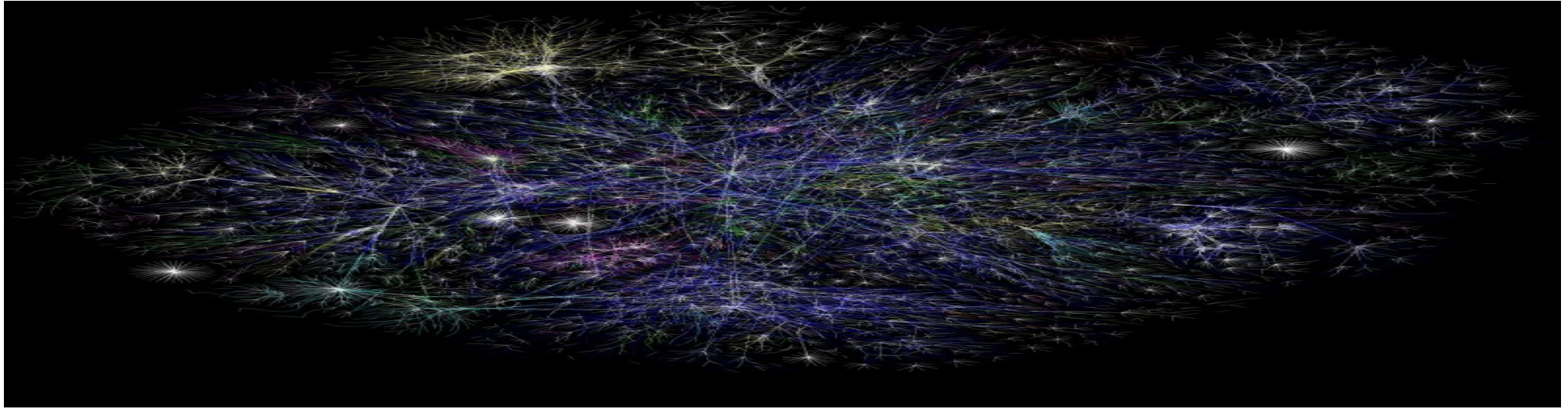


Como explorar a estrutura para inferir relevância de uma página?

- ❑ Páginas com mais links apontando para elas são mais relevantes
 - ❑ Peso do hiperlink é proporcional a relevância da página
- ➔
- ❑ Ideia do PageRank (início da Google)
 - ❑ Transformou as máquinas de busca

Internet

- ❑ Uma das maiores redes físicas construídas



- ❑ 30+ mil redes (sistemas autônomos)
- ❑ 1.2+ milhões de roteadores, 13+ bilhões de enlaces
- ❑ Construída de forma distribuída

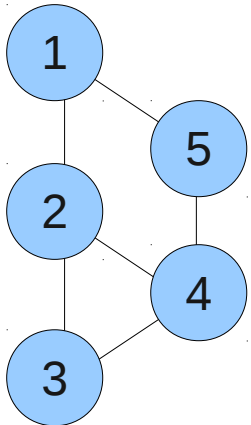
**Como descobrir a estrutura
da Internet (roteadores)?**

Descobrendo a Internet



Como descobrir a estrutura da Internet (roteadores)?

- ❑ Não há um oráculo que conhece a rede
- ❑ *traceroute*: descobre caminhos entre dois pontos na rede



- ❑ Supor operador tem acesso a 1
- ❑ $\text{tr}(1,3) = 1,2,3$
- ❑ $\text{tr}(1,4) = 1,5,4$
- ❑ Não descobre enlace (3,4) e (2,4)
- ❑ Problema de pesquisa, fundamental
- ❑ Como determinar o *bias* na rede descoberta?

Rede Neural

- ❑ *C. elegans* (roundworm)
- ❑ Possui sistema nervoso simples



- ❑ Rede Neural
 - 302 neurônios
 - completamente mapeada
- ❑ Importância da estrutura?

Classificando as Redes

- ❑ **Redes sociais:** relacionamento entre pessoas ou grupo de pessoas
- ❑ **Redes de informação:** relacionamento entre algum tipo de informação ou conhecimento
- ❑ **Redes tecnológicas:** redes físicas construídas pelo homem, em geral para distribuir *commodities*
- ❑ **Redes biológicas:** redes construídas pela natureza

Classificação apenas de referência

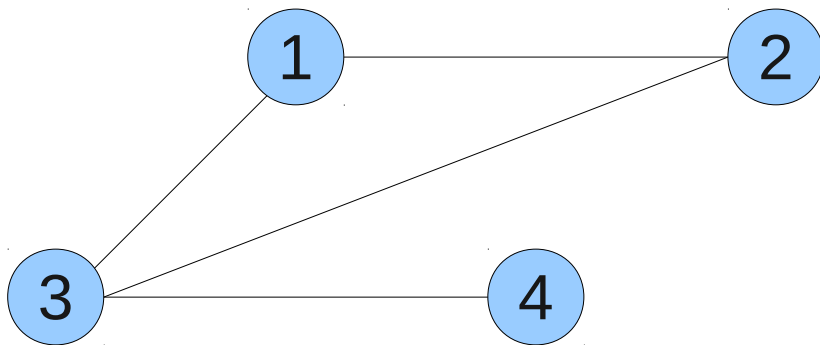
- ❑ Twitter: rede social ou rede de informação?

Rede (ou Grafo)

- Definindo formalmente uma rede $R = (V, E)$
- V = conjunto de objetos
 - chamaremos de vértices ou nós
- E = conjunto de pares relacionados (par não ordenado)
 - chamaremos de arestas
- Exemplo: $R = (V, E)$
 - $V = \{1, 2, 3, 4\}$
 - $E = \{(1,2), (1,3), (2,3), (3,4)\}$

Representação Gráfica

- Desenho de R (o que vimos até agora)
 - representação gráfica dos conjuntos
- Exemplo: $R = (V, E)$
 - $V = \{1, 2, 3, 4\}$
 - $E = \{(1,2), (1,3), (2,3), (3,4)\}$



- Estaremos interessados em redes grandes
- Não será possível desenhá-las!

Adjacência e Incidência

- Vértices adjacentes são vértices “vizinhos”
 - mais precisamente...
- Dois vértices a e b são **adjacentes** se existe $e = (a, b)$ no conjunto E
- Aresta e é **incidente** aos vértices a e b
- Exemplo: $R = (V, E)$
 - $V = \{1, 2, 3, 4\}$
 - $E = \{(1,2), (1,3), (2,3), (3,4)\}$

Vértices e Arestas

- Número de vértices de uma rede

- $n = |V|$ ← cardinalidade do conjunto

- Número de arestas de uma rede

- $m = |E|$

- Dado $R = (V, E)$, qual é o maior número de arestas de R ?

- número de pares não ordenados em um conjunto de $n = |V|$ objetos → $\binom{n}{2} = \frac{n(n-1)}{2} \leq n^2$

Grau

- Grau de um vértice v
 - número de vértices adjacentes a v
 - função $g(v)$
- Exemplo: $R = (V, E)$
 - $V = \{1, 2, 3, 4\}$ ■ $g(1) = ?$
 - $E = \{(1,2), (1,3), (2,3), (3,4)\}$ ■ $g(4) = ?$
- Dado $R = (V, E)$ qualquer
 - Grau mínimo de um vértice? —————▶ **zero**
 - Grau máximo de um vértice? —————▶ **$n - 1$**

Grau Médio

- Média do grau dos vértices

$$\bar{g} = 1/n \sum_{u \in V} g(u) \quad \longleftarrow \text{Grau do vértice } u$$

- Também pode ser calculado diretamente

$$\bar{g} = 2m/n$$

- Logo temos $\sum_{u \in V} g(u) = 2m$

- Cada aresta tem duas pontas!

Distribuição do Grau

□ Distribuição (empírica) do grau dos vértices

$$p_k = P[D = k] = \frac{\text{Número de vértices com grau } k}{\text{Número total de vértices}}$$

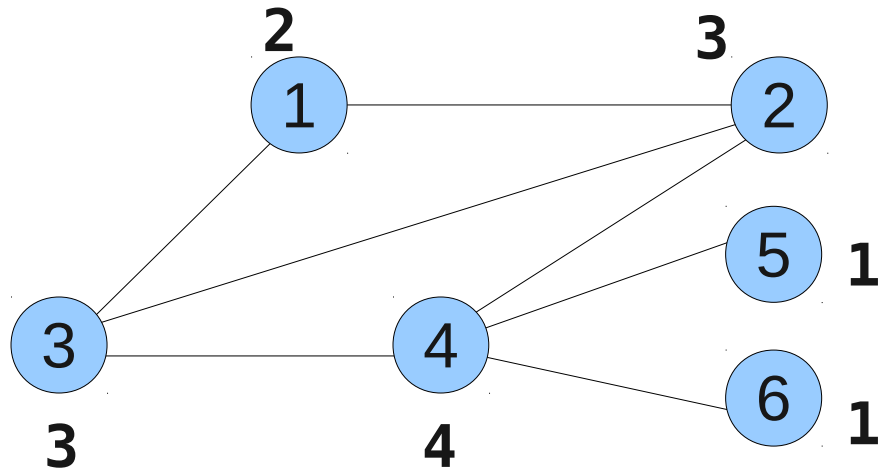
↖
Variável aleatória que
denota grau de um vértice

□ CCDF – Complementary Cumulative Distribution Function

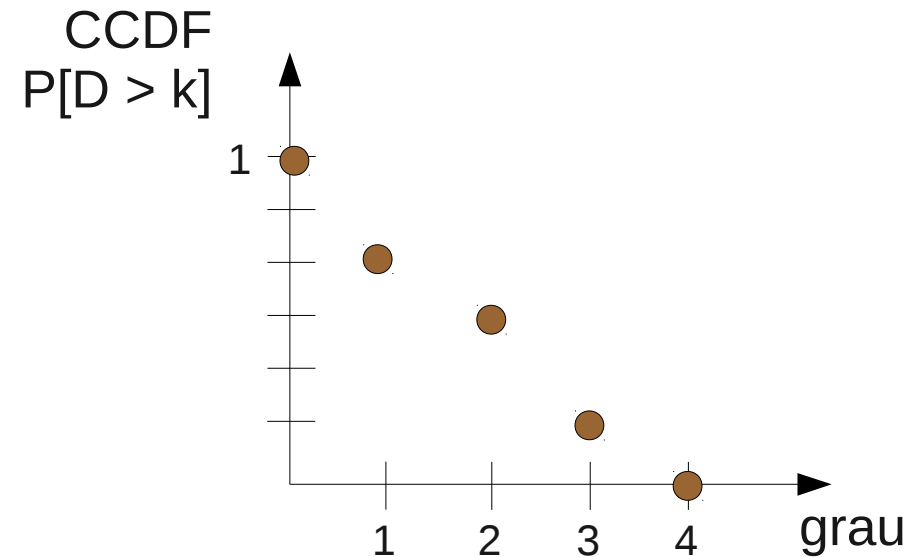
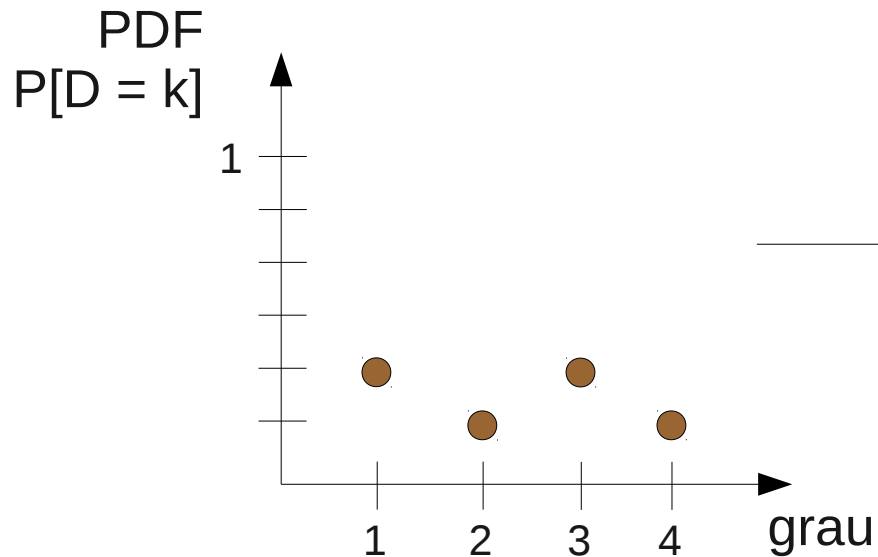
○ Fração de vértice com grau $> k$

$$P[D > k] = 1 - \sum_{i=0}^k p_i$$

Exemplo

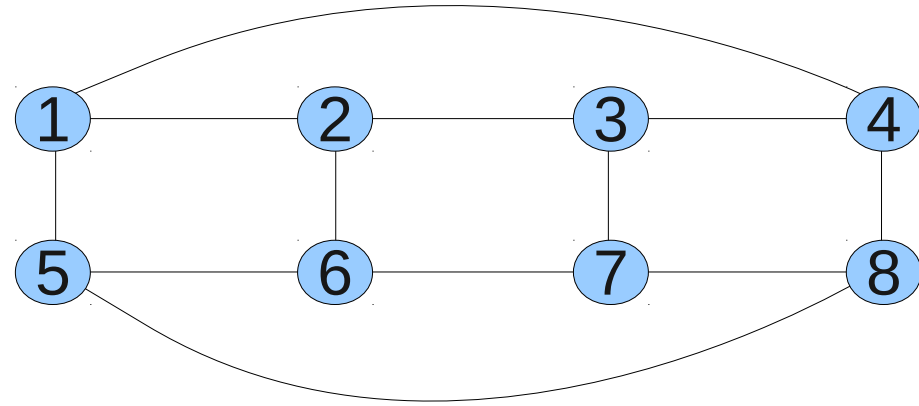


□ Distribuição (empírica) do grau dos vértices?



Caminho em Redes

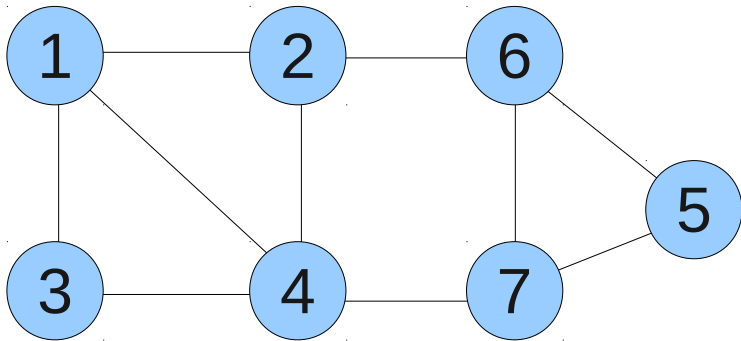
- Caminho simples entre dois vértices
 - sequência de vértices conectados por arestas, sem repetição de vértices



- Ex. caminho entre 1 e 7?
 - 1, 2, 3, 7 $\longrightarrow$ (1,2), (2,3), (3,7)
 - 1, 5, 6, 3, 7 é caminho?
- Comprimento do caminho
 - número de *arestas* que o forma

Distância

- Comprimento do **menor** caminho entre dois vértices
- Função $d(u,v)$, onde u e v são vértices
 - infinito quando não há caminho



- Exemplo

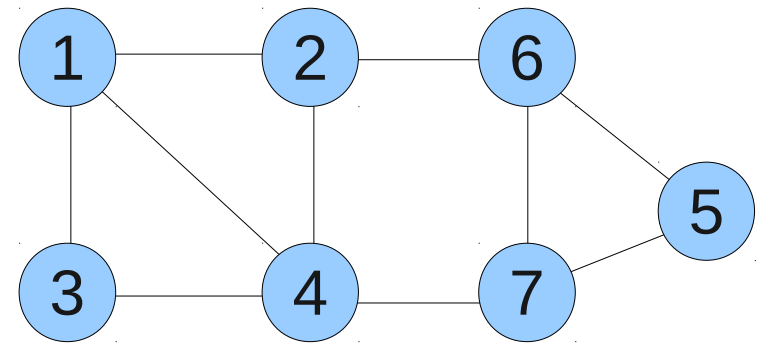
- $d(6, 3) = ?$

- $d(7, 1) = ?$

Distância Média

- Distância média da rede
 - média entre todos os pares (não ordenados) de vértices

$$\bar{d} = \frac{\sum_{u,v \in V} d(u,v)}{\binom{n}{2}}$$



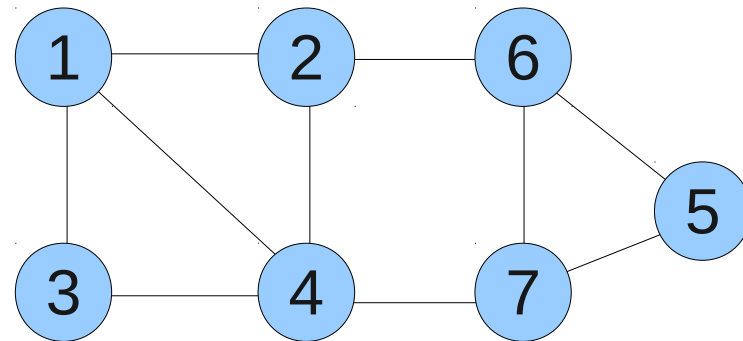
- Se $d(u,v)$ não definida, ignorar par (u,v)
 - também do denominador

Diâmetro

- Maior distância entre dois vértices da rede
- Maior caminho *mínimo* entre dois vértices quaisquer
 - maior distância

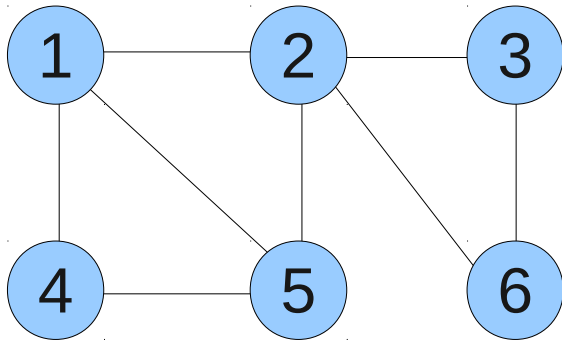
$$r = \max_{u, v \in V} d(u, v)$$

■ Exemplo



Clusterização

- Medida a respeito de triângulos na rede
- Comparação entre triângulos existentes e “quase-triângulos”
 - Quase triângulos: caminho com duas arestas



- Alta clusterização dá uma ideia de redundância
 - muitos caminhos alternativos

Medida de Clusterização

- Fração de aresta entre vizinhos
- Definida para cada vértice da rede

$$C_i = \frac{E_i}{\binom{d_i}{2}}$$

← E_i : # de arestas entre os vizinhos de i

← d_i : grau do vértice i

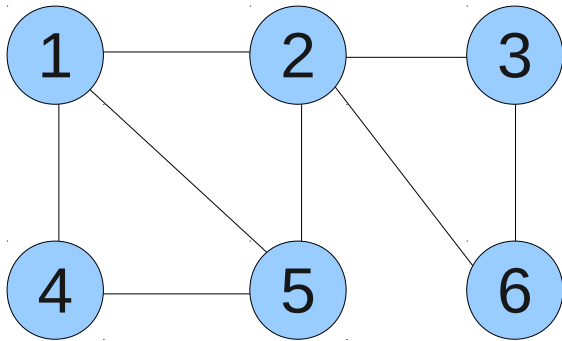
$$C_i = \frac{2 E_i}{d_i(d_i - 1)}$$

- Clusterização do grafo
 - média da clusterização dos vértices

$$C = 1/n \sum_{v \in V} C_i$$

Calculando Clusterização

■ Exemplo



■ Quanto vale C ?

Exemplos

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	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

Três Características

- Fundamentais e observadas em diversas redes reais
 - início do estudo empírico em Redes Complexas



- Efeito “mundo pequeno”



- Efeito “meus amigos são amigos”



- Efeito “abaixo a democracia”

Mundo Pequeno

- “It's a small world, after all”
- Distância média entre vértices é muito pequena
- Mesmo para redes muito grandes
 - Web Graph – 2×10^8 vértices ← 7.5
 - Rede de atores – 10^6 vértices ← 3
- “Six degrees of separation”
 - Separação (média) entre qualquer duas pessoas no mundo!
- Experimento de Milgram na década de 60



Meus Amigos também são Amigos

- Se A está relacionado com B e C
- Então B e C têm mais chance de estarem relacionados
- Triângulos se fecham nas redes
- Coeficiente de clusterização



■ Resultados empíricos

■ AS Graph – 10^4 vértices ← 0.39

■ Rede de metabolismo – 765 vértices ← 0.67

Abaixo a Democracia



- Graus dos vértices é muito desigual
- Muitos com grau pequeno, poucos muito grande

Igual a distribuição de renda no Brasil!

- Distribuição tem causa pesada (lei de potência)
- AS Graph – 10^4 vértices
 - Grau médio: **5.9**
 - Grau máximo: **~2100**
- Citações – 7.8×10^5 vértices
 - Grau médio: **8.6**
 - Grau máximo: **~9000**

A Curiosidade

- **Fato I:** Muitas redes possuem propriedades estruturais peculiares (não esperadas)
 - Graus, distâncias, etc.
- **Fato II:** Muitas redes distintas possuem propriedades estruturais similares
 - Web, Facebook, Rede de neurônio



- “Million dollar question”

Por que?

- Ainda estamos tentando responder a esta pergunta!