

Análise de Redes

Aula 02 – Introdução a Grafos

Prof. Patrick Terrematte



Teoria de Grafos

- Propriedades
 - Ordem e Tamanho
 - Caminhos e medidas
 - Grau e Distribuição de Grau
 - Coeficiente de Clusterização
 - Medidas de Centralidade
- Tipos de Redes
 - Redes Aleatórias
 - Redes 'Mundo Pequeno' (*Small Worlds*)
 - Redes Livre de Escala

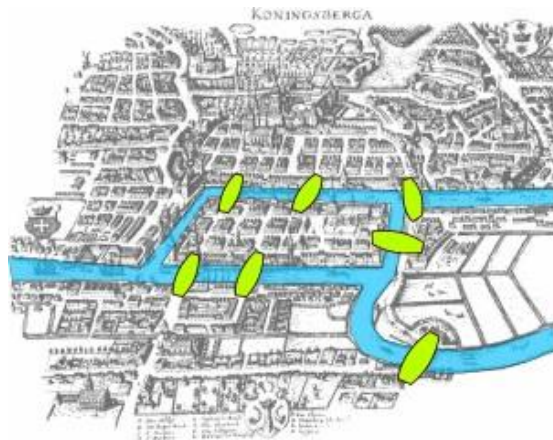
Teoria de Grafos



<http://www.visualcomplexity.com/vc/>

As Pontes de Königsberg

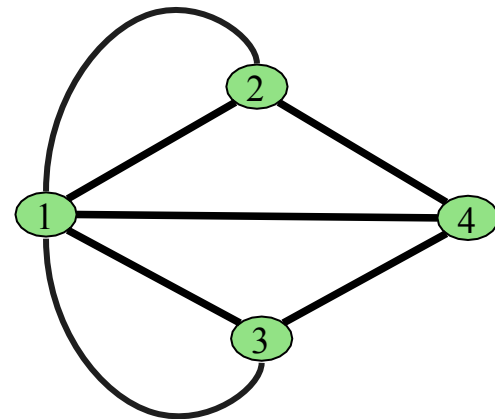
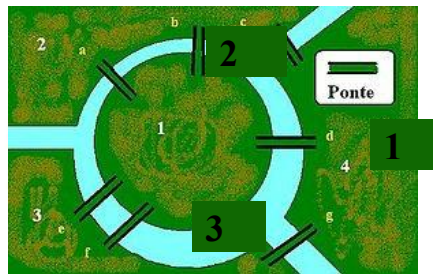
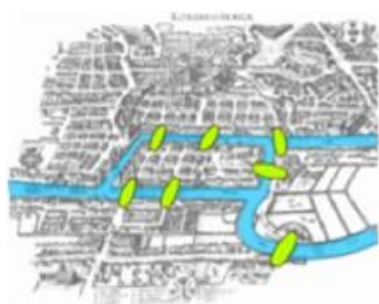
- Na cidade de Königsberg havia um conjunto de 7 pontes que cruzavam o rio Pregel e conectavam duas ilhas centrais entre si e com as margens.
- **Problema:**
*Há um **caminho** que passe por todas as pontes, visitando cada ponte uma única vez?*



Em 1735, o matemático suíço **Leonard Euler** mostrou que **não existe** uma solução para o problema.

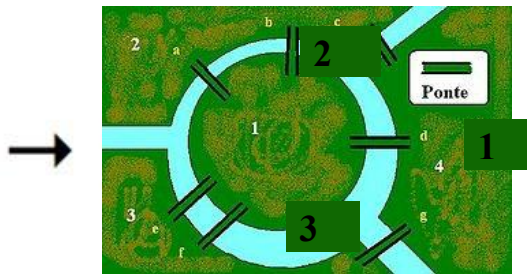
As Pontes de Königsberg

- **Caminho Euleriano** é um caminho (em um grafo) que visita **uma aresta** apenas **uma vez**.
- A demonstração foi baseada em **grafos**.
- Para que exista um caminho que percorra todos os vértices passando por cada aresta uma única vez, é **necessário** que **0** ou **2** dos vértices tenham um número **ímpar** de arestas.

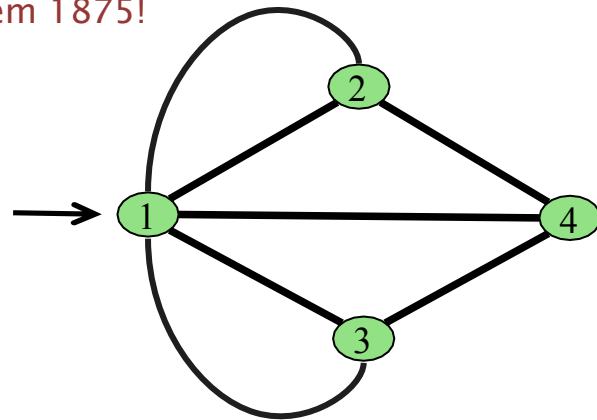


As Pontes de Königsberg

- **Teorema de Euler:** Se um grafo não-direcionado tiver dois, ou nenhum vértice ímpar, ele possui pelo menos um caminho Euleriano.
 - **Racional:** se houver um caminho cruzando todas as pontes, mas nunca a mesma ponte duas vezes, então os vértices com número ímpar de arestas devem ser o ponto inicial ou final deste caminho.
 - Como tornar este um caminho Euleriano? Problema resolvido em 1875!



4



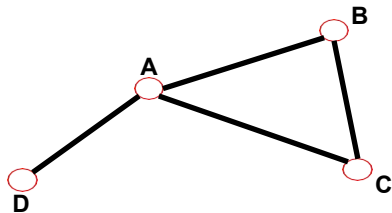
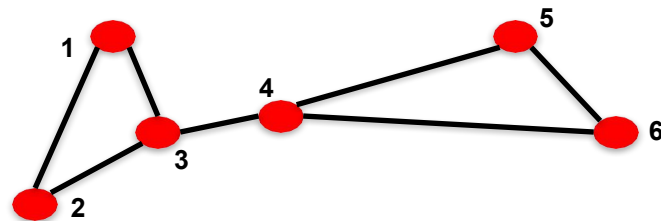
Grafos

Conjunto composto pelo **par ordenado** $G = (N, L)$

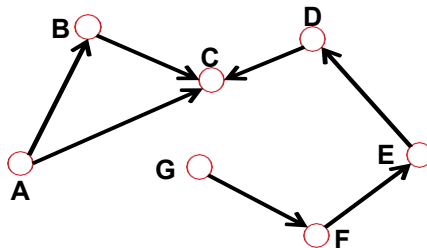
■ **Ordem:** # vértices $n(G) = 6$

■ **Tamanho:** # arestas $l(G) = 7$

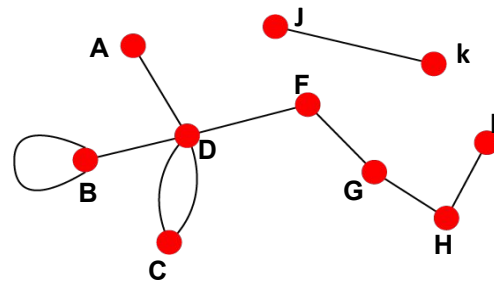
■ Dado $G = (N, L)$, o maior número de arestas de G = onde n é a ordem do grafo.



Não-orientados Links de
co-autoria Redes de
atores Interações
proteicas



Orientados
URLs na www Chamadas
telefônicas Reações
metabólicas



Não-conectados
Componentes gigantes
isolados

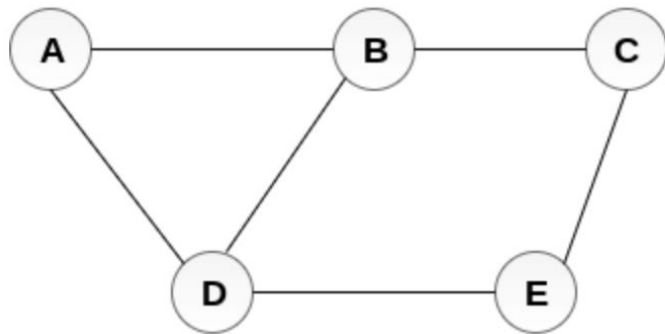
Grafos

Uma matriz de adjacência $A^{n \times n}$ representa elementos a_{ij} tais que cada e_{ij} representa uma aresta.

$$A_{ij} = \begin{matrix} & A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{matrix}$$

Grafos

Uma matriz de adjacência $A^{n \times n}$ representa elementos a_{ij} tais que cada e_{ij} representa uma aresta.

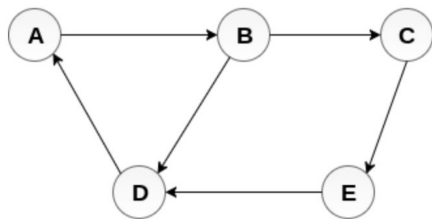


Undirected Graph

	A	B	C	D	E
A	0	1	0	1	0
B	1	0	1	1	0
C	0	1	0	0	1
D	1	1	0	0	1
E	0	0	1	1	0

Adjacency Matrix

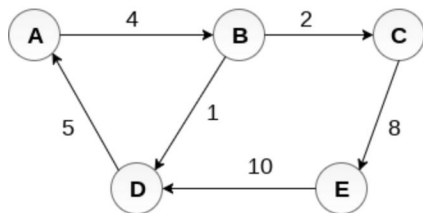
Grafos



Directed Graph

	A	B	C	D	E
A	0	1	0	0	0
B	0	0	1	1	0
C	0	0	0	0	1
D	1	0	0	0	0
E	0	0	0	1	0

Adjacency Matrix



Weighted Directed Graph

	A	B	C	D	E
A	0	4	0	0	0
B	0	0	2	1	0
C	0	0	0	0	8
D	5	0	0	0	0
E	0	0	0	10	0

Adjacency Matrix

Grafos ou Redes?

REDES - sistemas reais

World Wide Web

Rede metabólica

Rede social

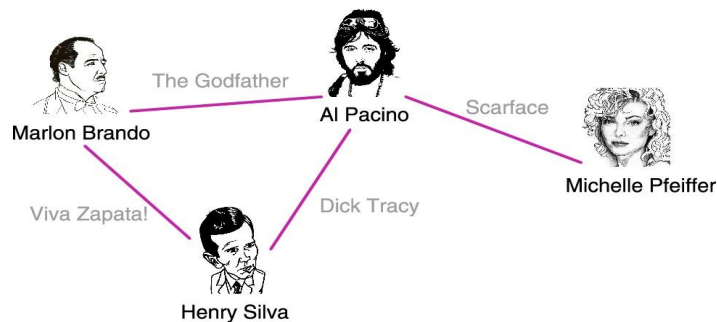
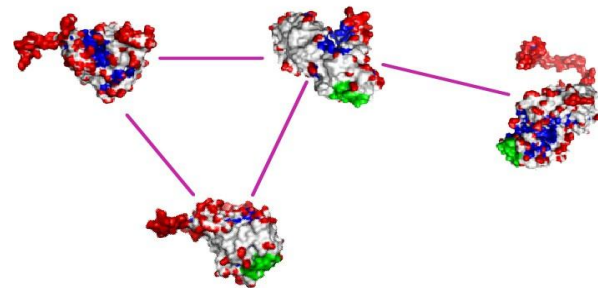
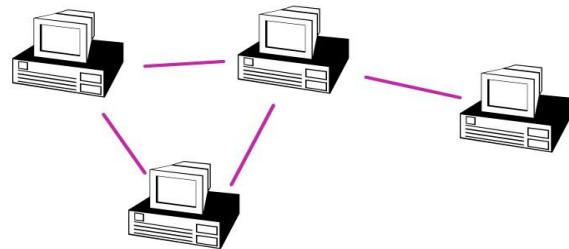
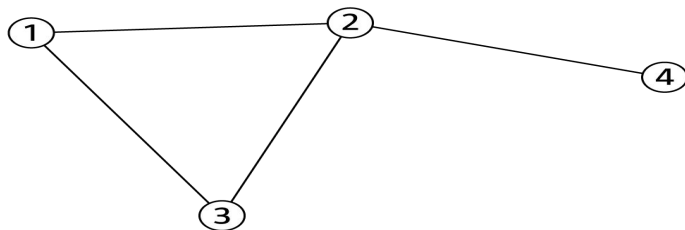
Nomenclatura: nó, ligação.

GRAFO - representação matemática

Grafo web

Grafo social

Nomenclatura: vértice, aresta.



Tipos de Redes

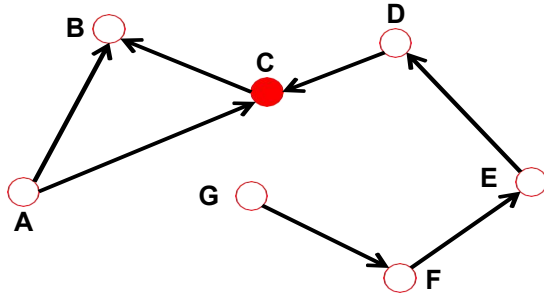
Network	Nodes	Links	Directed / Undirected	N	L	⟨K⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.34
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile-Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorships	Undirected	23,133	93,437	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Papers	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

GRAU, GRAU MÉDIO E DISTRIBUIÇÃO DE GRAU

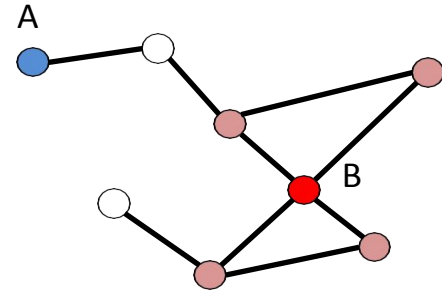
Grau e Grau Médio

■ **Grau (K):** número de arestas incidentes ao vértice.

- Em grafos orientados, k_{in} e k_{out} .
 - **Fonte (source):** vértice com $k^{in} = 0$
 - **Sumidouro (sink):** vértice com $k^{out} = 0$



$$k_C^{in} = 2 \quad k_C^{out} = 1 \quad k_C = 3$$



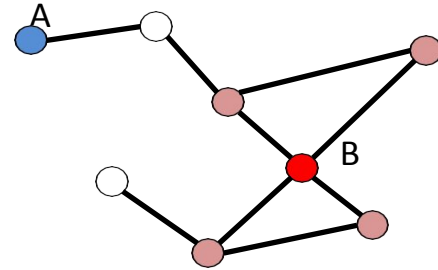
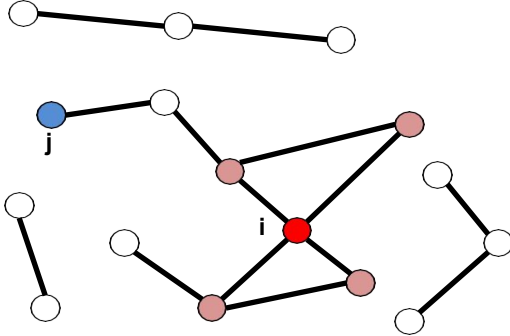
$$k_A = 1 \quad k_B = 4$$

Grau e Grau Médio

■ **Grau (K):** número de arestas incidentes ao vértice.

■ **Grau Médio $\langle K \rangle$:**

$$\langle k \rangle = \frac{1}{n} \sum_i k_i = \frac{2m}{n} = \frac{2|E|}{|V|}$$



$$k_A = 1 \quad k_B = 4$$

Grau e Grau Médio

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	k
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.9 0

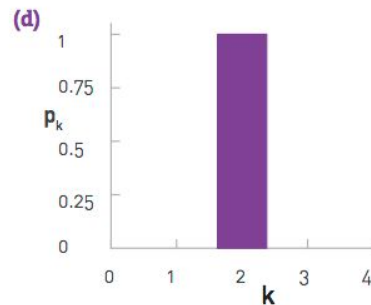
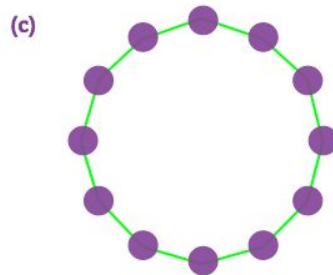
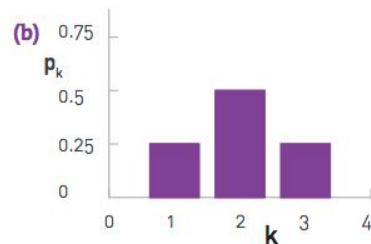
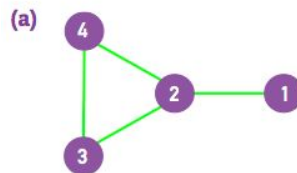
Distribuição de Grau

- **P(k):** probabilidade que um vértice escolhido aleatoriamente tenha grau **k**.
- **Distribuição empírica de grau:** frequência de vértices com grau **k**.

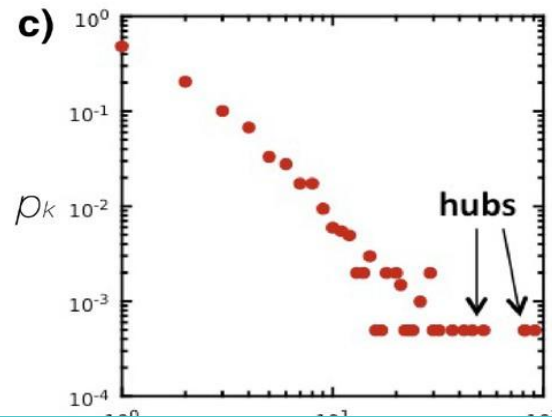
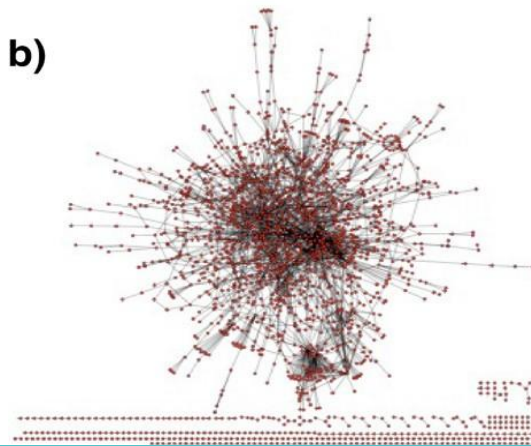
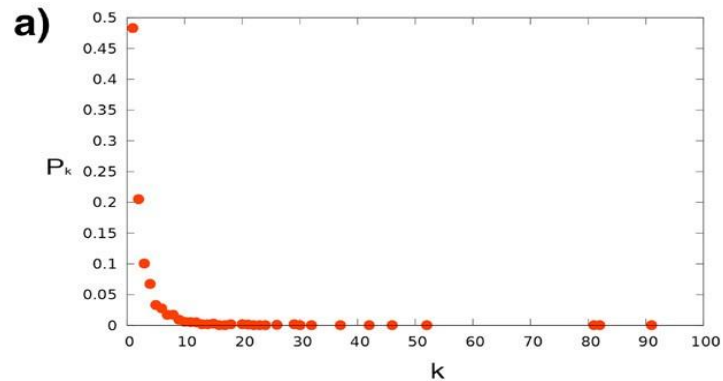
$$P(k_i = k) = P(k) = P_k = \frac{n_k}{\sum_k n_k} = \frac{n_k}{n}$$

k_i = grau de cada nó,

n_k = # vértices com grau k



Distribuição de Grau em redes PPI

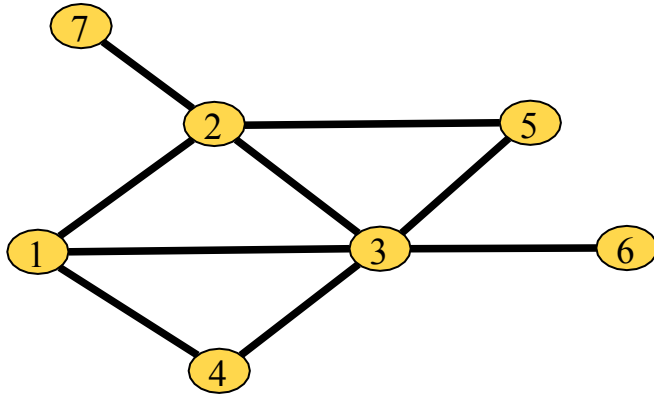


MATRIZ DE ADJACÊNCIA

Representação de Grafos

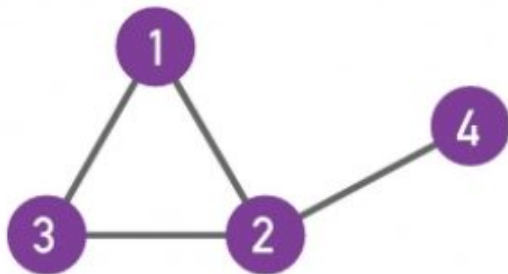
Matriz de adjacência $(n \times n)$

- $a_{ij}=1$, se existe aresta entre os vértices i e j
- $a_{ij}=0$, caso contrário



	1	2	3	4	5	6	7
1	0	1	1	1	0	0	0
2	1	0	1	0	1	0	1
3	1	1	0	1	1	1	0
4	1	0	1	0	0	0	0
5	0	1	1	0	0	0	0
6	0	0	1	0	0	0	0
7	0	1	0	0	0	0	0

MATRIZ DE ADJACÊNCIA (grafo não direcionado)



$$A_{ij} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{matrix}$$

$$k_2 = \sum_{j=1}^4 A_{2j} = \sum_{i=1}^4 A_{i2} = 3$$

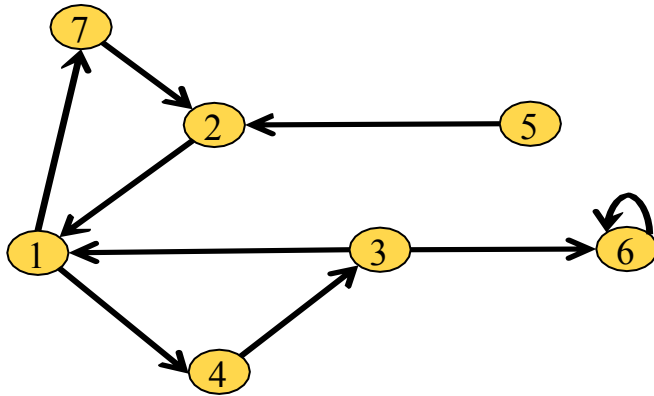
$$A_{ij} = A_{ji} \quad A_{ii} = 0$$

$$L = \frac{1}{2} \sum_{i=1}^N A_{ij}$$

$$\langle k \rangle = \frac{2L}{N}$$

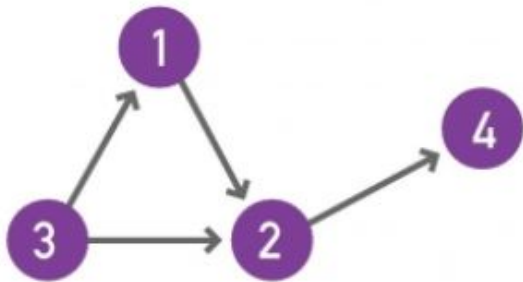
Representação de Grafos

- Qual a diferença desta matriz de adjacência para a anterior?
- E para a posterior?



	1	2	3	4	5	6	7
1	0	0	0	1	0	0	1
2	1	0	0	0	0	0	0
3	1	0	0	0	0	1	0
4	0	0	1	0	0	0	0
5	0	1	0	0	0	0	0
6	0	0	0	0	0	1	0
7	0	1	0	0	0	0	0

MATRIZ DE ADJACÊNCIA (grafo direcionado)



$$A_{ij} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{matrix}$$

$$k_2^{\text{in}} = \sum_{j=1}^4 A_{2j} = 2, \quad k_2^{\text{out}} = \sum_{i=1}^4 A_{i2} = 1$$

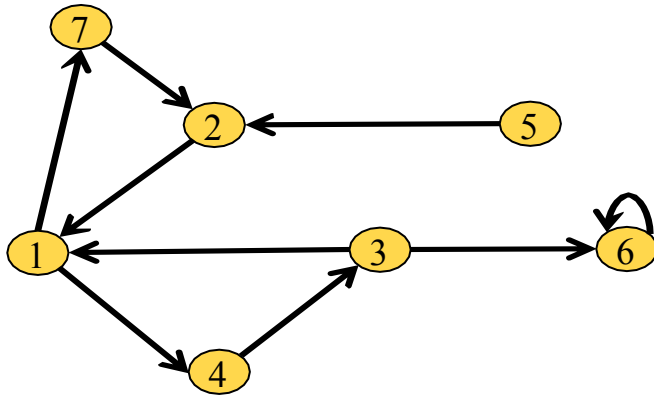
$$A_{ij} \neq A_{ji} \quad A_{ii} = 0$$

$$L = \sum_{i,j=1}^N A_{ij}$$

$$\langle k^{\text{in}} \rangle = \langle k^{\text{out}} \rangle = \frac{L}{N}$$

Representação de Grafos

- **Lista de adjacência:** lista de vértices com seus respectivos vértices adjacentes.
- Computacionalmente vantajosa com grafos esparsos ($N^2 \gg L$)



1: 4, 7

2: 1

3: 1, 6

4: 3

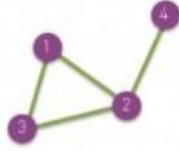
5: 2

6: 6

7: 2

Resumo: Tipos de Redes

a. Undirected

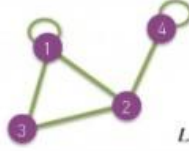


$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \quad A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^N A_{ij} \quad \langle k \rangle = \frac{2L}{N}$$

b. Self-loops

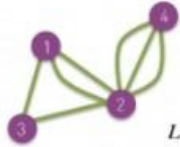


$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$$\exists i, A_{ii} \neq 0 \quad A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1, i \neq j}^N A_{ij} + \sum_{i=1}^N A_{ii} \quad ?$$

c. Multigraph
(undirected)

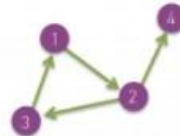


$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \quad A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^N A_{ij} \quad \langle k \rangle = \frac{2L}{N}$$

d. Directed

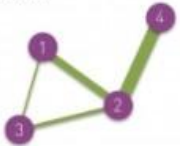


$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{ij} \neq A_{ji}$$

$$L = \sum_{i,j=1}^N A_{ij} \quad \langle k \rangle = \frac{L}{N}$$

e. Weighted
(undirected)

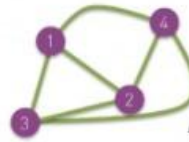


$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \quad A_{ij} = A_{ji}$$

$$\langle k \rangle = \frac{2L}{N}$$

f. Complete Graph
(undirected)



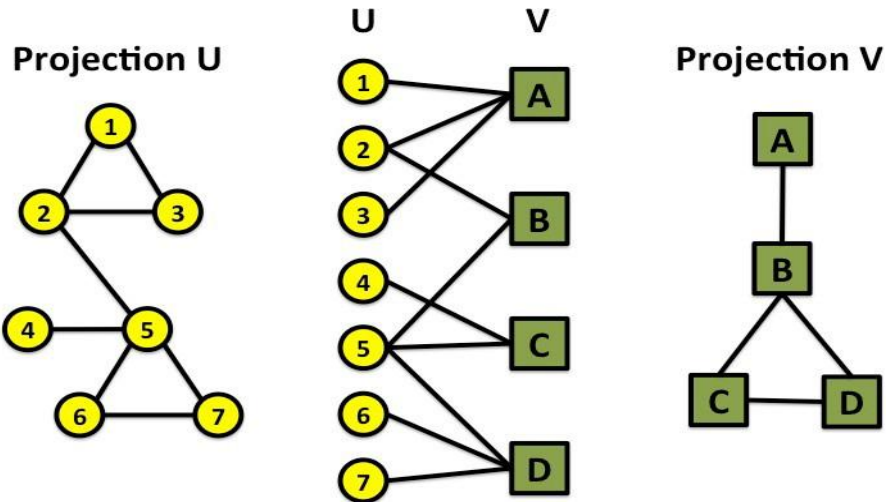
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \quad A_{ij} = 1$$

$$L = L_{\max} = \frac{N(N-1)}{2} \quad \langle k \rangle = N-1$$

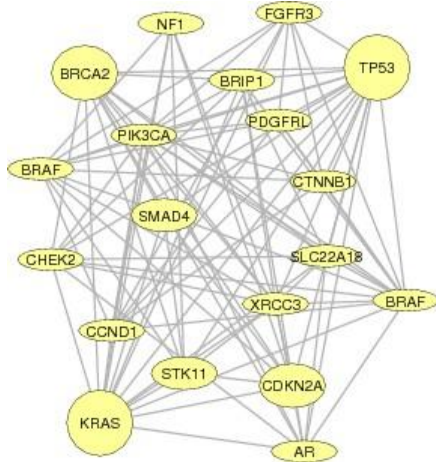
Grafos Bipartidos (Bígrafo)

- Grafo cujos nós podem ser divididos em dois conjuntos disjuntos U e V, de modo que cada link conecte um nó em U a um em V;
- Ou seja, U e V são conjuntos independentes.



Rede de atores de Hollywood
Redes de colaborações
Rede de doenças (diseasome)

Grafos Bipartidos (Bígrafo)

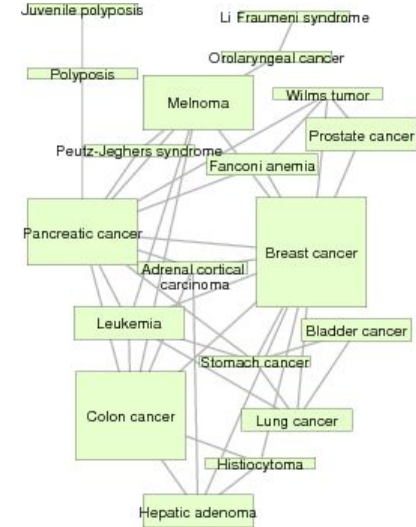
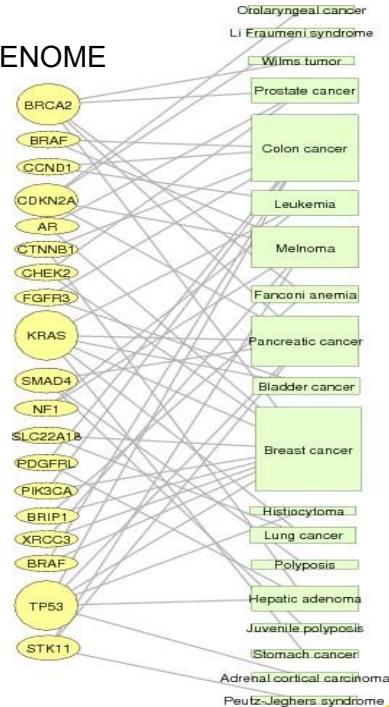


Gene network

DISEASOME

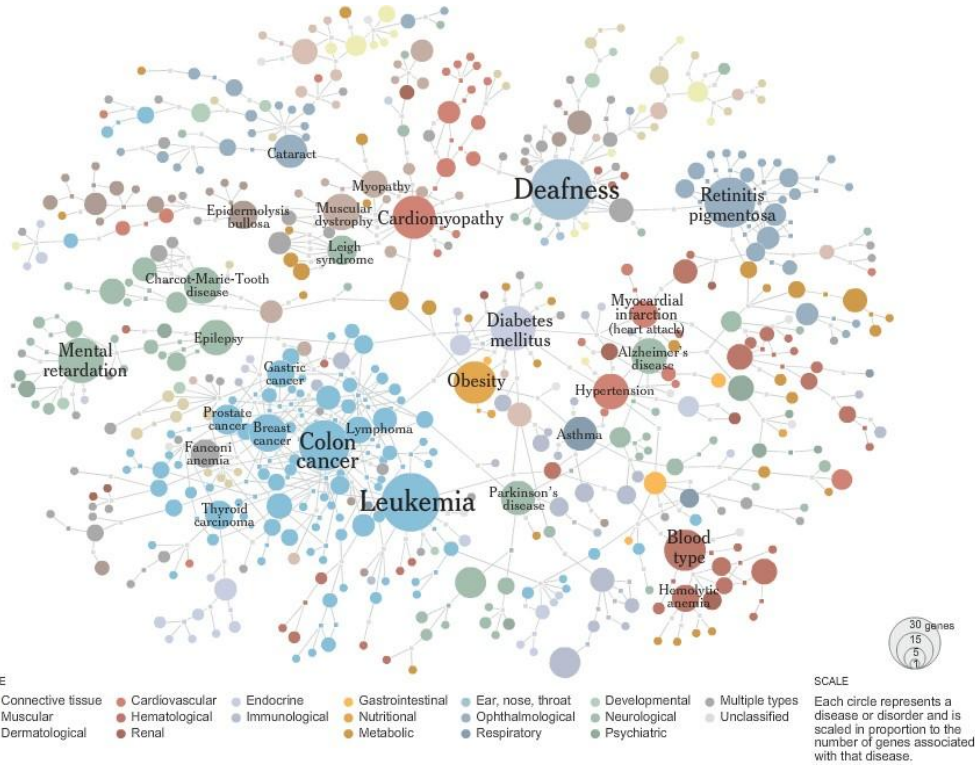
PHENOME

GENOME



Disease network

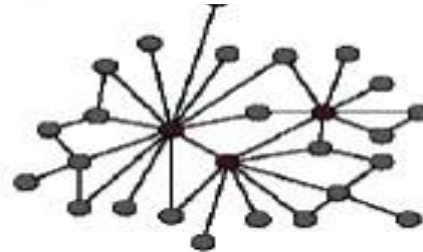
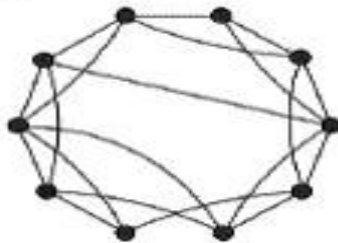
Human Diseasesome Network



Sources: Marc Vidal; Albert-Laszlo Barabasi; Michael Cusick;
Proceedings of the National Academy of Sciences

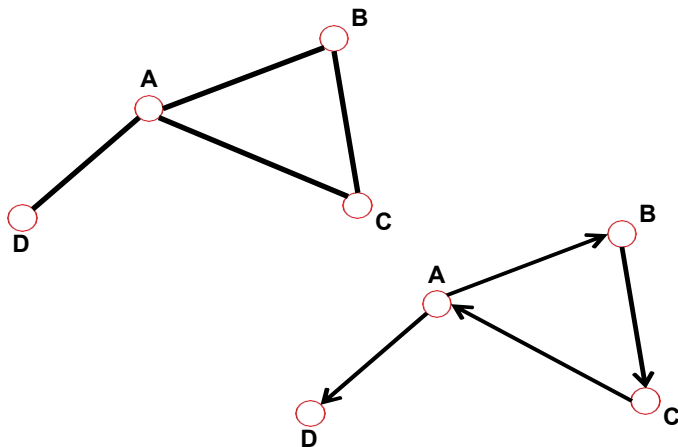
The New York Times

Propriedades de Grafos



Caminho

- **Caminho**: sequência de vértices consecutivos conectados por arestas $\langle s, u, v, \dots, t \rangle$.
- Em um **grafo direcionado**, o caminho segue o sentido da aresta. $AB \neq BA$.
- **Distância** (*caminho mínimo, caminho geodésico*): o menor caminho entre dois vértices.



Grafo não-direcionado

$\langle B, C, A, D \rangle$ é caminho de comprimento **3**.

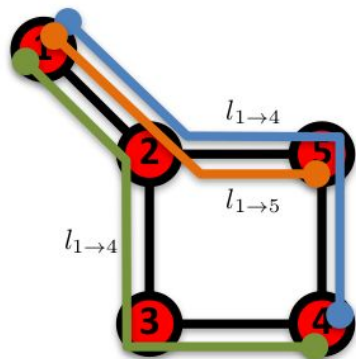
$\langle B, A, D, C \rangle$ não é caminho.

Grafo direcionado

$\langle A, B, C \rangle$ é caminho.

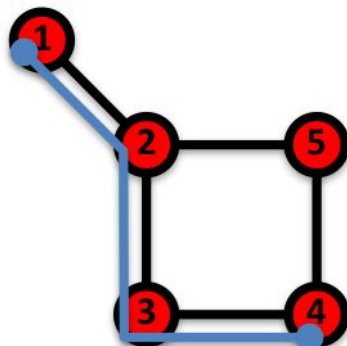
$\langle A, C, B \rangle$ não é caminho.

Caminhos



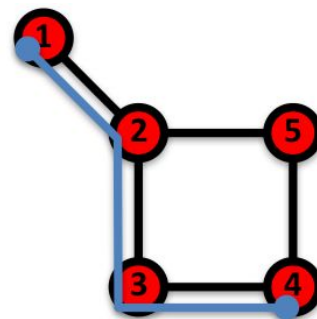
$$l_{1 \rightarrow 4} = 3 \quad l_{1 \rightarrow 5} = 2$$

Distância: menor comprimento entre 2 vértices (**caminho mínimo**).



$$l_{1 \rightarrow 4} = 3$$

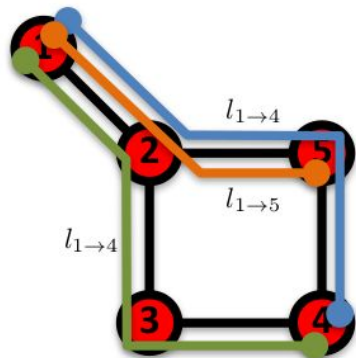
Diâmetro: maior distância entre quaisquer 2 vértices (**maior caminho mínimo**).



$$(l_{1 \rightarrow 2} + l_{1 \rightarrow 3} + l_{1 \rightarrow 4} + l_{1 \rightarrow 5} + l_{2 \rightarrow 3} + l_{2 \rightarrow 4} + l_{2 \rightarrow 5} + l_{3 \rightarrow 4} + l_{3 \rightarrow 5} + l_{4 \rightarrow 5}) / 10 = 1.6$$

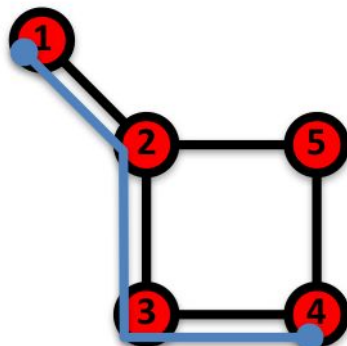
Caminho médio: média das distâncias entre todos os pares de vértices.

Caminhos



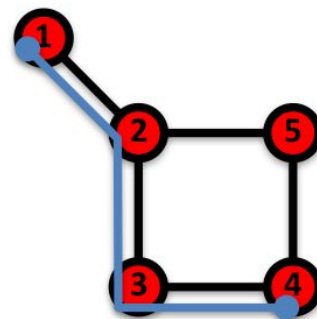
$$l_{1 \rightarrow 4} = 3 \quad l_{1 \rightarrow 5} = 2$$

Distância: menor comprimento entre 2 vértices (**caminho mínimo**).



$$l_{1 \rightarrow 4} = 3$$

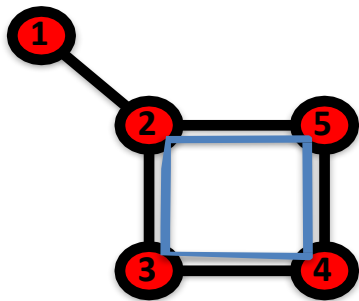
Diâmetro: maior distância entre quaisquer 2 vértices (**maior caminho mínimo**).



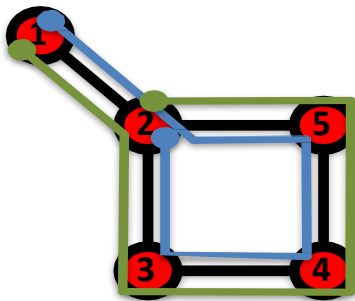
$$\langle d \rangle \equiv \frac{1}{L_{\max}} \sum_{i,j>i} d_{ij}$$

Caminho médio: média das distâncias entre todos os pares de vértices.

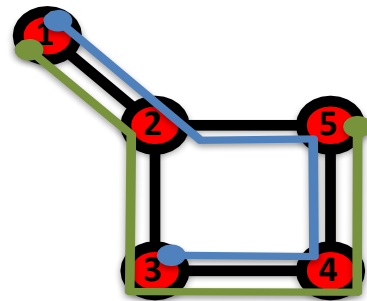
Caminhos



Ciclo: caminho que começa e termina no mesmo vértice.



Caminho Euleriano: caminho que passa por cada aresta uma única vez.



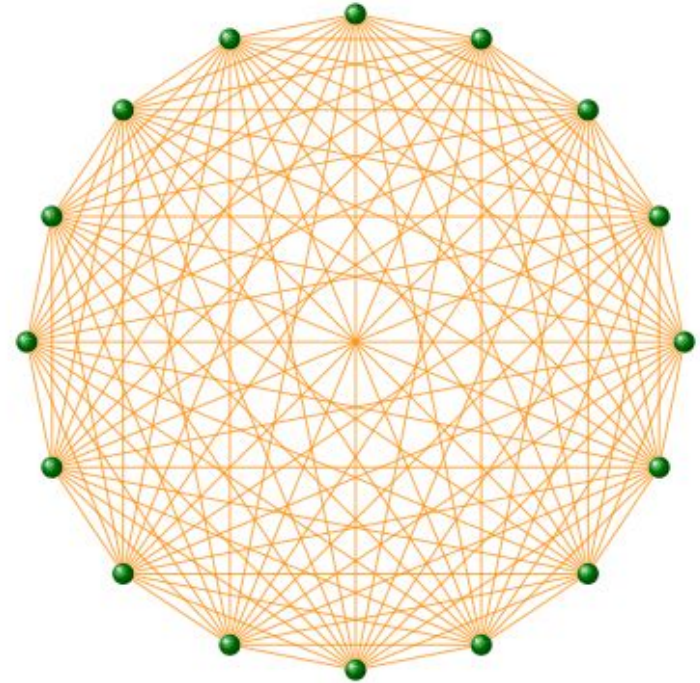
Caminho Hamiltoniano: caminho que passa por cada vértice uma única vez.

Grafos Completos

- Grafo com tamanho $L = L_{\max}$ e grau médio $\langle k \rangle = N-1$.
- O maior número de arestas de em um grafo de ordem N:

$$L_{\max} = \binom{N}{2} = \frac{N!}{(N-2)!2!} = \frac{N(N-1)}{2}$$

- **Densidade:** número de arestas L em relação ao grafo completo L_{\max} .
- Dado um grafo de ordem **N** e tamanho **L**.
 - Grafo espaço: $L \sim N$.
 - Grafo denso: $L \sim N^2$.
 - Grafo completo: $L = N^2$.



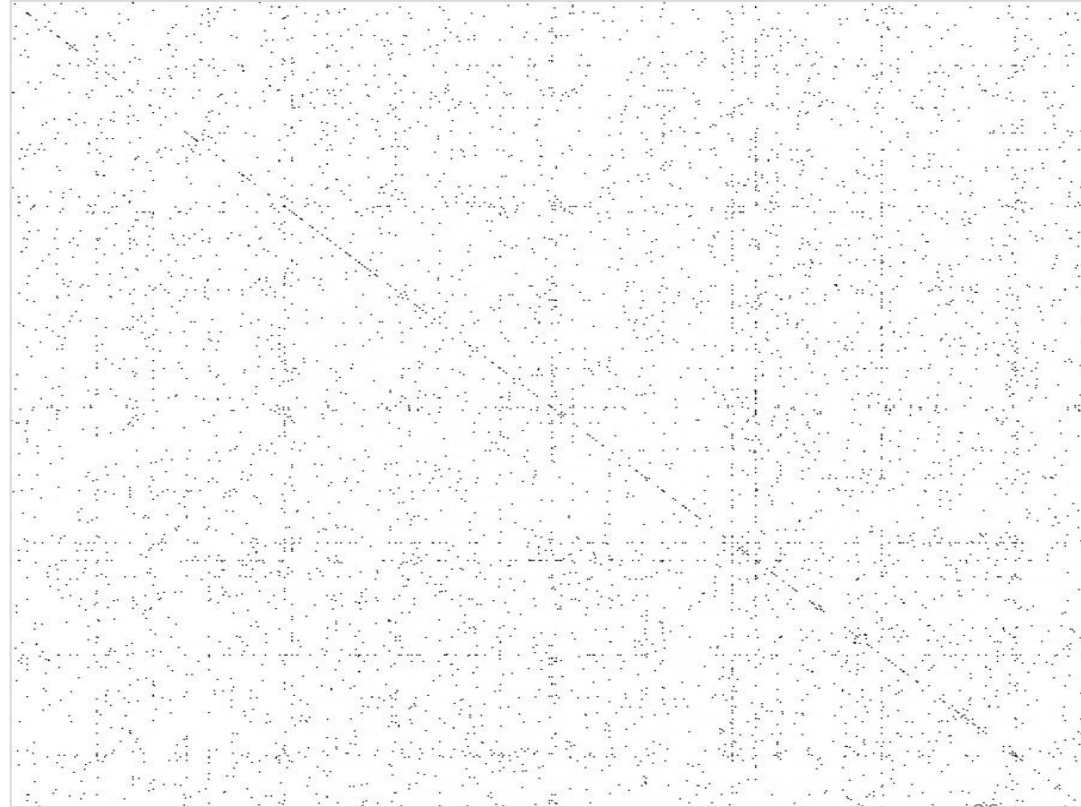
Redes Reais são Esparsas

$$L \ll L_{max} \text{ ou } \langle k \rangle \ll N-1$$

	Nodes	Links	Lmax	<K>
WWW (sample)	325,729	1.4×10^6	10^{12}	4.51
Proteína (<i>S. cerevisiae</i>)	1,870	4,470	10^7	2.39
Colaboração (math)	70,975	2×10^5	3×10^{10}	3.9
Atores em um filme	212,250	6×10^6	1.8×10^{13}	28.78

Redes Reais são Esparsas

A matriz de adjacência da rede de interação proteína-proteína de levedura, constituída por 2.018 vértices, cada um representando uma proteína.



Coeficiente de Clusterização Local

- Razão entre as **arestas existentes** e o **# máximo de arestas possíveis entre os vizinhos** de um dado vértice.
- Não está definido para vértices com grau 0 ou 1.

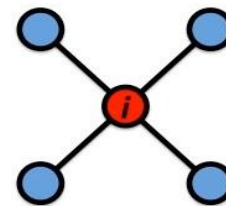
$$C_i = \frac{E_i}{\binom{d_i}{2}}$$

de arestas entre os vizinhos de i

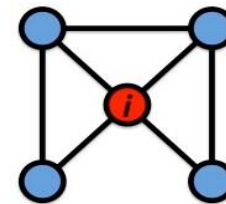
grau do vértice i

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

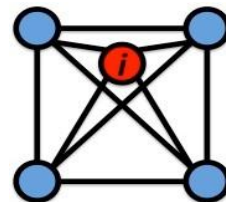
- CC não expressa uma propriedade do vértice e sim dos seus vizinhos!



$$CC = 0/12 = 0$$



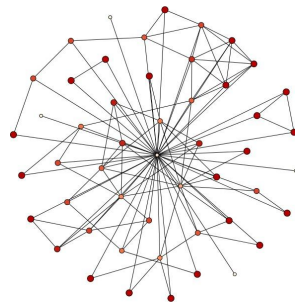
$$CC = 6/12 = 0,5$$



$$CC = 12/12 = 1$$

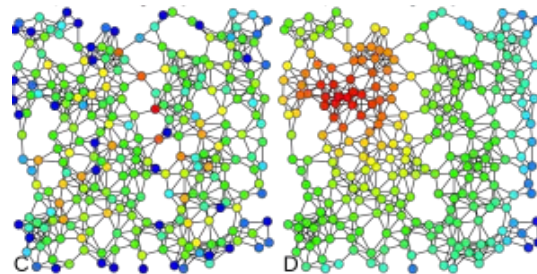
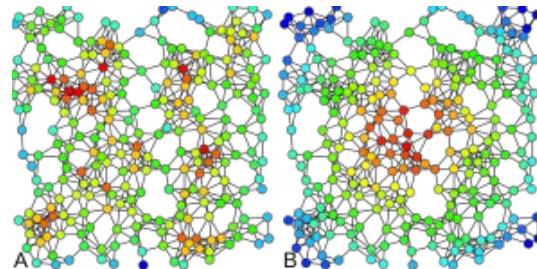
Coeficiente de Clusterização Global

- **Média aritmética dos CC de cada vértice:** mede o grau com que os vértices de um grafo tendem a agrupar.
- Nas **redes sociais**, o agrupamento refere-se **aos círculos de amigos ou conhecidos** onde os seus membros se conhecem, formando um grupo na rede.
- **Comportamentos assortivo:** vértices com alto grau se ligam preferencialmente a vértices com alto grau.
- **Comportamento dissasortivo:** o contrário.



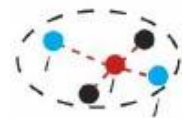
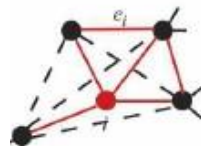
Medidas de Centralidade

- **Centralidade de Grau:** grau 'normalizado'.
- **Centralidade de Proximidade (*closeness*):** menor distância média.
- **Centralidade de Intermediação (*betweenness*):** pontes entre vértices, 'caminho do meio'.
- **Centralidade de Eigenvector:** conexão a vértices de alto grau.



Resumo: Propriedades de Redes

- Grau k_i
Número de arestas do vértice i
- Distribuição de grau $P(k)$
Probabilidade (frequência) dos vértices de grau k
- Distância Média $\langle L \rangle$
Média dos caminhos mínimos entre todos os vértices
- Diâmetro da rede
Maior caminho mínimo
- Coeficiente de Clusterização CC
Medida da 'organização' local ou global
- Centralidade de Intermediação (*Betweenness*)
Importância relativa do vértice na intermediação dos caminhos



Algumas questões...

Dada uma rede com V vértices e E arestas =>
rede com topologia estatisticamente idêntica!

Resultado: modela uma rede com topologia estática!

Problema: redes reais são sistemas dinâmicos!

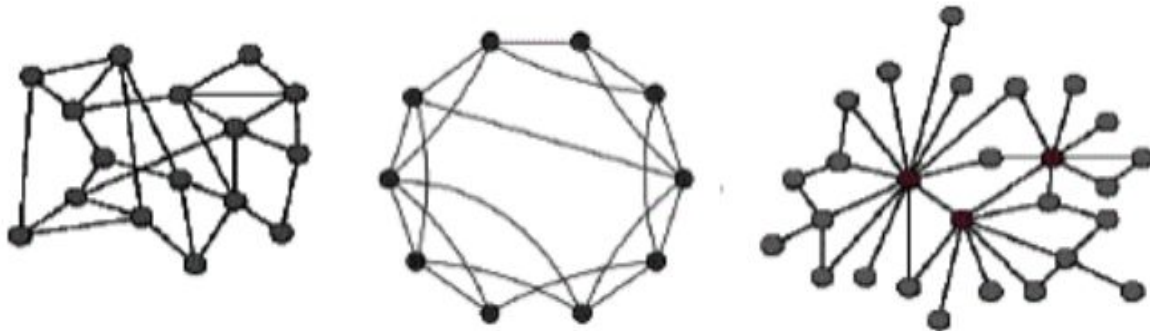
Redes Reais

- Dinâmica de vértices e de arestas (*rewiring*)
- Envelhecimento (*aging*)
- Respostas a estímulos e perturbações
- Efeitos não-lineares

Objetivo: identificar o **processo gerador** da dinâmica da rede.

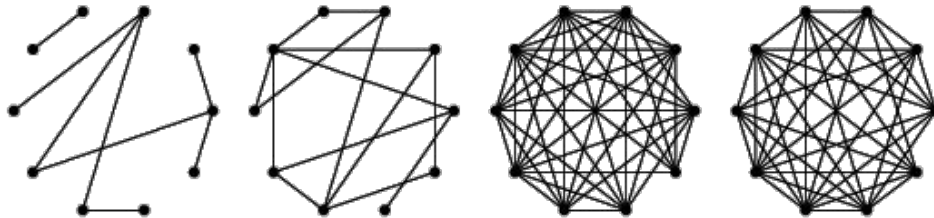
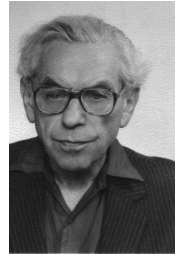
Bônus: representação correta da topologia da rede.

Modelos de Redes



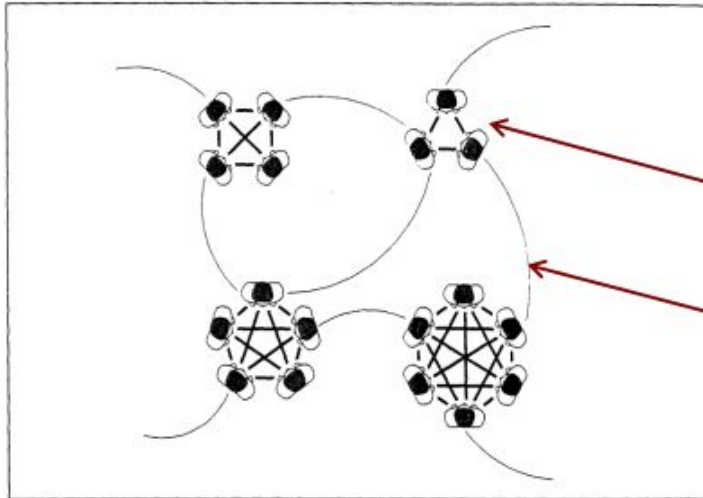
Redes Aleatórias

- **Paul Erdős e Alfred Rényi (1959)**: redes complexas naturais e sociais parecem seguir um padrão aleatório de formação => **Grafos Aleatórios**.
- **Formação**: novos nós são randomicamente adicionados à rede, gerando grafos estatisticamente homogêneos. Aparecimento de componente gigante quando $\langle k \rangle = 1$.
- $L \sim N^2$ => vértices com aproximadamente mesmo número de arestas.



Redes Aleatórias

- **Mark Granovetter (1970):** sociedade organizada em componentes bem agrupados, conectados por componentes mais fracos.
- O desafio foi reconciliar a teoria de grafos aleatórios com a realidade agrupada de Granovetter. Isto levou quase três décadas!



Conexões fortes: agrupamentos densamente conectados.

Conexões fracas: pontes para o mundo externo.

Redes Aleatórias

- **Distribuição de grau $\langle k \rangle$: Binomial ou Poisson.**

Apresentam pico em torno de $\langle k \rangle$. Se aumentar p , a rede se torna mais densa, aumentando $\langle k \rangle$ e movendo o pico para a direita.

A largura da distribuição (dispersão: quanto mais densa a rede, mais ampla é a distribuição e maiores as diferenças de graus.

- **Coeficiente de clusterização (CC)**

Diminui com o aumento da rede ($\sim 1/N$).

Independente do grau do nó $\Rightarrow C(k)$ é constante.

- **Distância média $\langle d \rangle$, $\langle L \rangle$**

Média dos caminhos pequenos.

$$\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle}$$

Seis graus de separação

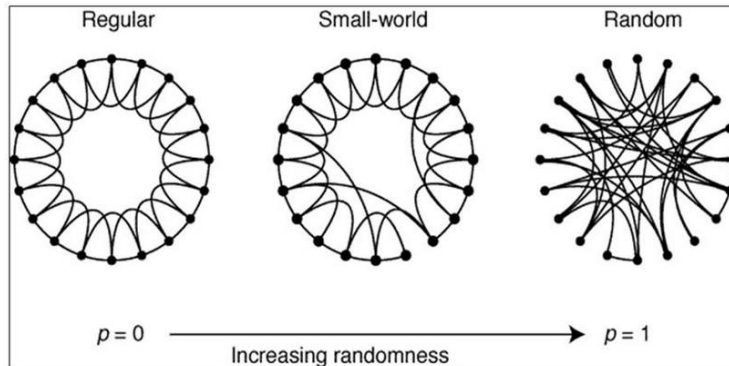
- Stanley Milgram (1967) realiza um experimento para **determinar a “distância” entre duas pessoas quaisquer** dos EUA.
- Envio de cartas partindo de **Nebraska KA**, com destino a uma pessoa em **Boston MA**, por intermédio de pessoas conhecidas.
- Das 160 cartas preparadas, 42 chegaram.
- O menor caminho foi de 2 conexões e o mais longo de 11.
- **O valor médio foi de 5,5 conexões!**

Efeito Mundo Pequeno: as informações se propagam rapidamente por toda a rede ($L \leq \log n$)

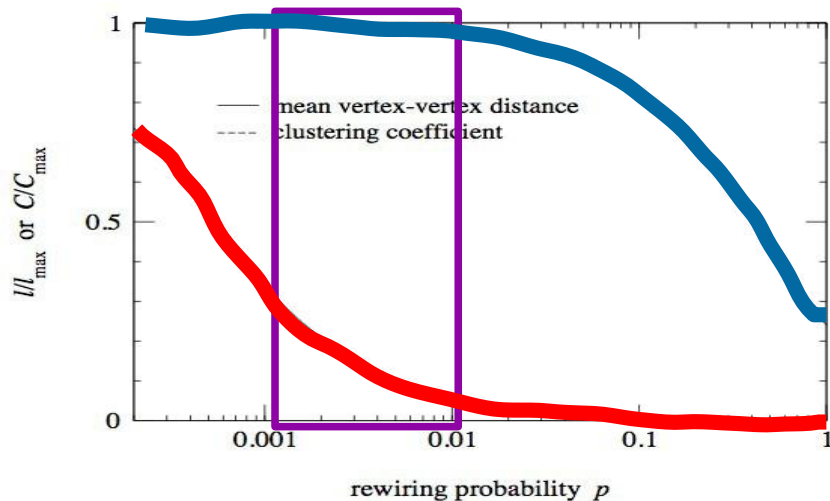


Redes Mundo Pequeno (*small-world*)

- **Duncan Watts e Steven Strogatz (1998)**: sistemas auto-organizáveis não são nem aleatórios nem regulares.
- **Formação**: a partir de um anel regular com N vértices e k arestas, reconecte cada vértice aleatoriamente com probabilidade p .



Redes Mundo Pequeno (*small-world*)



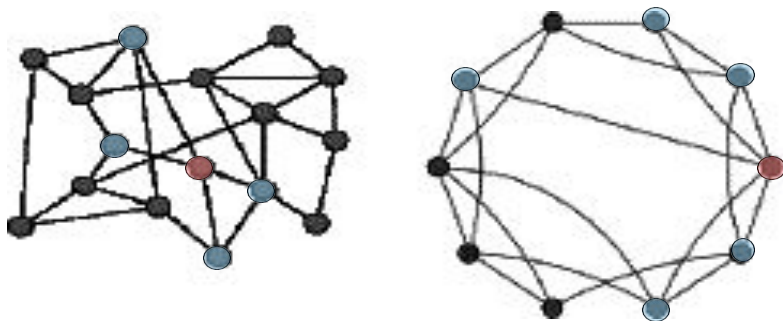
Baixo P => Distância média $\langle L \rangle$ pequena e coeficiente de clusterização $\langle C \rangle$ alto

Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
<i>C. elegans</i>	2.65	2.25	0.28	0.05

Modelos Erdos-Renyi e Watts-Strogatz

- **Proíbem a presença de vértices com um grau muito acima da média.**
- Redes com número de vértices fixo.
- Vértices com grau próximo da média.
- Arestas criadas aleatoriamente.



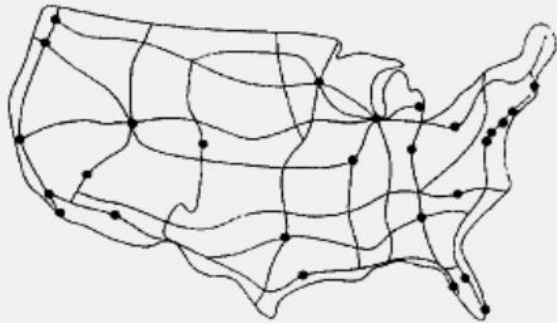
	Erdos- Renyi	Watts- Strogatz
Caminho $\langle L \rangle$	PEQUENO	PEQUENO
Coefficiente de Clusterização $\langle C \rangle$	PEQUENO	GRANDE

Redes Reais

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001
Movie actors	225 226	61	3.65	2.99	0.79	0.00027
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}
<i>E. coli</i> , substrate graph	282	7.35	2.9	3.04	0.32	0.026
<i>E. coli</i> , reaction graph	315	28.3	2.62	1.98	0.59	0.09
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006
Power grid	4941	2.67	18.7	12.4	0.08	0.005
<i>C. Elegans</i>	282	14	2.65	2.25	0.28	0.05

Redes Reais

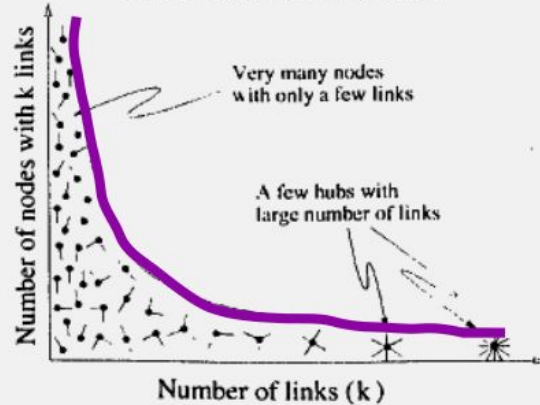
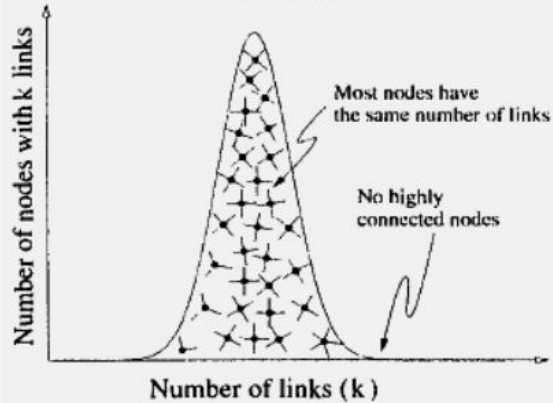
Hubs (conectores)



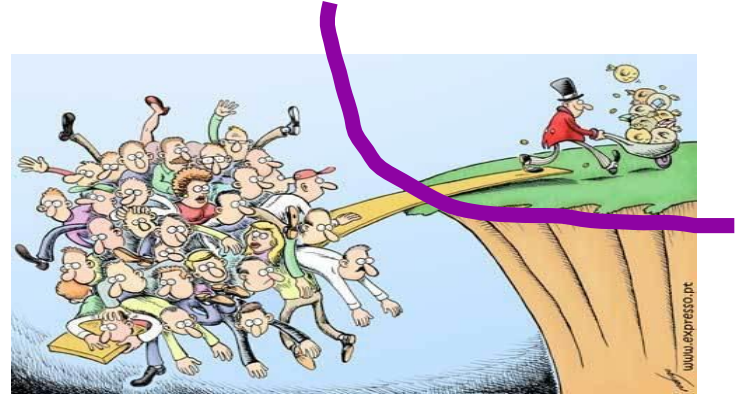
Bell Curve



Power Law Distribution



Redes Reais



Redes Sem Escala (*scale-free*)

Barabasi e Alberts (1999): a distribuição de graus em redes com um grande número de componentes segue uma **lei de potência**.

Princípios gerativos:

- Crescimento
A cada passo, um novo vértice é inserido na rede.
“Senioridade”: vértices mais antigos tem maior K .
- Conexão preferencial (“reforço”)
Probabilidade de conexão é proporcional ao grau do vértice.
“Rico-fica-mais-rico”: acelera o crescimento de vértices com alto K .



Redes Sem Escala (*scale-free*)

- **Distribuição de grau $\langle k \rangle$**

Lei de potência: aK^{-3}

Maior parte dos nós com graus baixos.

Existência de nós com alto grau (hubs).

- **Coeficiente de clusterização: baixo.**

$C(k)$ é constante.

Redes biológicas disassortivas.

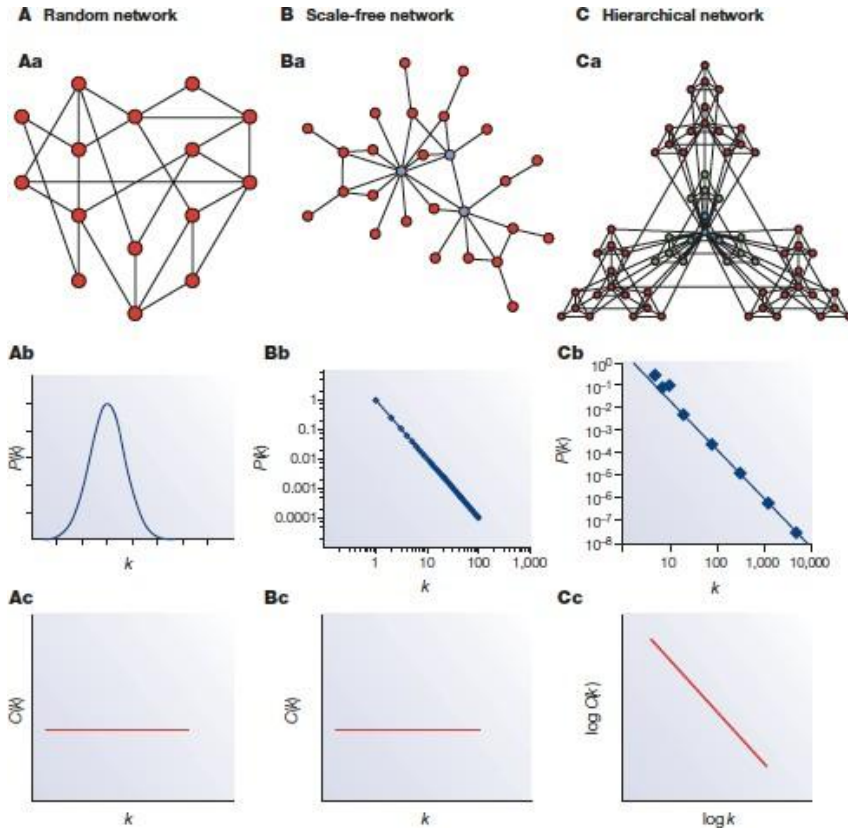
Redes sociais assortivas.

- **Distância média: $\sim \log N$ ou $\log \log N$**

Caminhos médios pequenos.

“Ultra small world”.

Resumo: Tipos de Redes











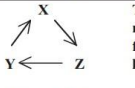

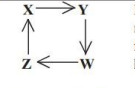





Motivos (*motiffs*)

- Uri Alon e cols. (2002): **padrões recorrentes em redes** parecem exercer **funções bem definidas**.
 - Redes gênicas *E. Coli* e *S. Serevisae*
 - Redes neuronais *C. elegans*
 - Cadeias alimentares
 - Circuitos eletrônicos
 - Subredes www
- **Hipótese:** *motifs* surgem devido a **requerimentos especiais** dos sistemas e podem ser utilizados para definir classes de redes ou homologias.



Motivos (*motiffs*)

Network	Nodes	Edges	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score
Gene regulation (transcription)			 Feed-forward loop			 Bi-fan					
<i>E. coli</i>	424	519	40	7 ± 3	10	203	47 ± 12	13			
<i>S. cerevisiae</i> *	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons			 Feed-forward loop			 Bi-fan			 Bi-parallel		
<i>C. elegans</i> †	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs			 Three chain			 Bi-parallel					
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			
Electronic circuits (forward logic chips)			 Feed-forward loop			 Bi-fan			 Bi-parallel		
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic circuits (digital fractional multipliers)			 Three-node feedback loop			 Bi-fan			 Four-node feedback loop		
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s838‡	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web			 Feedback with two mutual dyads			 Fully connected triad			 Uplinked mutual dyad		
nd.edu§	325,729	1,466,6	1.1e5	2e3 ± 1e2	800	6.8e6	5e4 ± 4e2	15,000	1.2e6	1e4 ± 2e2	5000

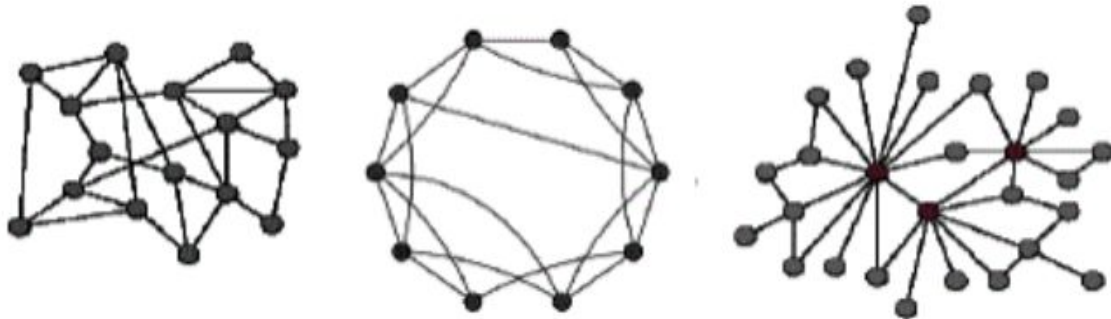
Technical Comment

Artzy-Randrup Y, Fleishman

S. J, Ben-Tal N, & Stone L.

Science, **305**, 2004

Interpretação biológica das propriedades de grafos



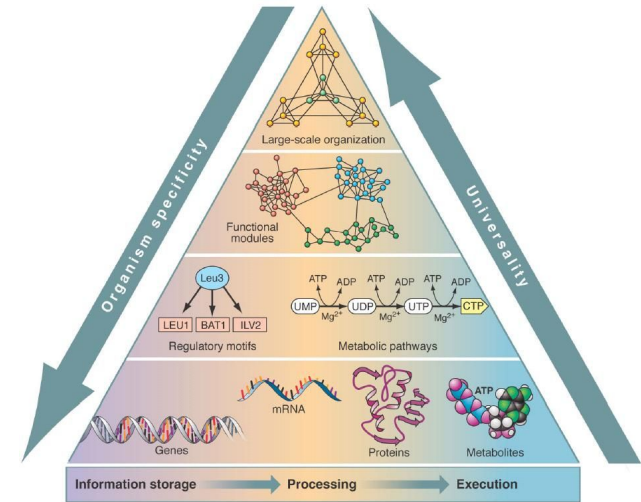
Propriedades de Redes Biológicas

MODULARIDADE

- Componentes com alta intraconectividade e baixa interconectividade.
- Alguma sobreposição e cruzamentos de módulos.
- **Organização hierárquica.**

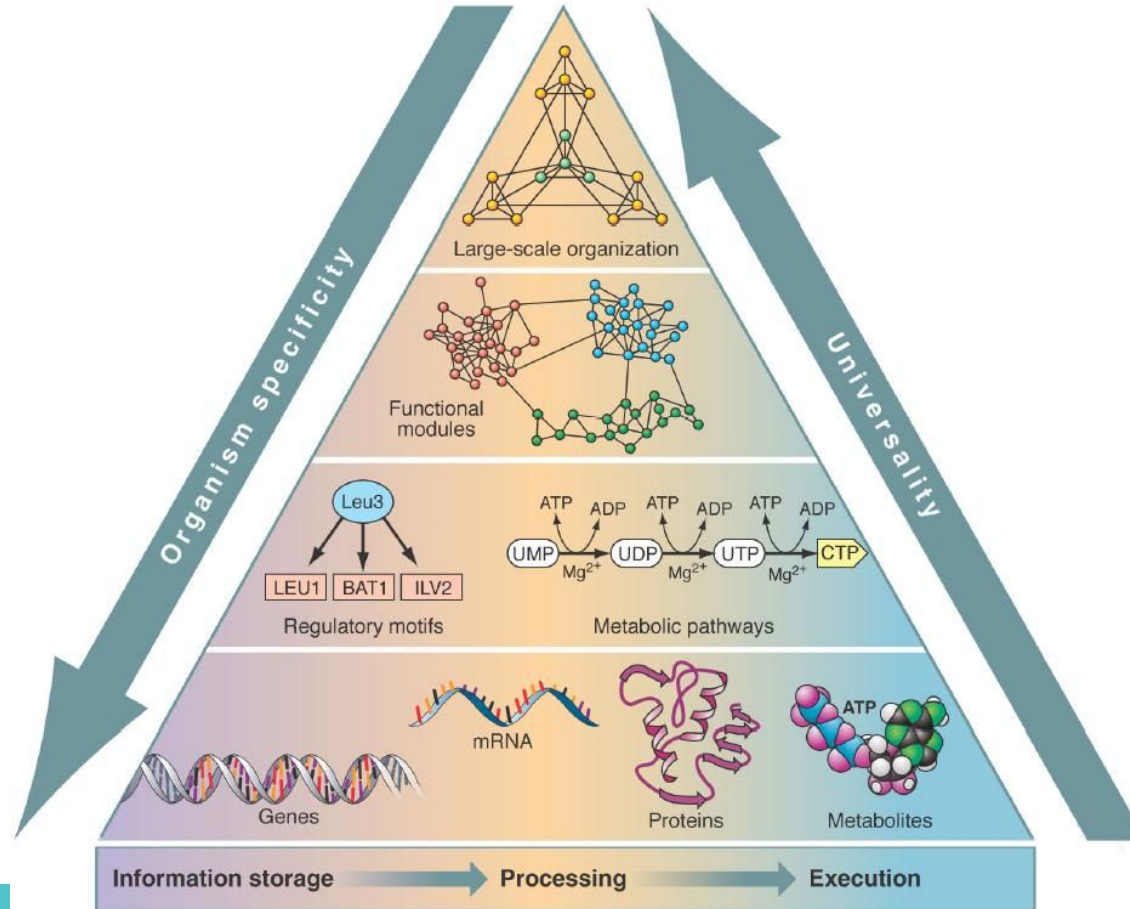
MOTIVOS & CLIQUES

- Subgrafos encontrados em frequência maior que a esperada.
- *Redes transcricionais*: alças de **feedback** e **feed-forward**.
- *PPIs*: pequenos **ciclos** e **subgrafos completos**.



Oltvai ZN, Barabási A-L. Science, **298**, 2002

Propriedades de Redes Biológicas



Propriedades de Redes Biológicas

HUBS

- Nós com grau bem maior que a média.
- Remoção aleatória vs. remoção dirigida.
- 73% dos genes *S. cerevisiae*: não-essenciais.
- **Mecanismos seletivos**: restrições funcionais e evolução!

CAMINHOS

- Caminho mínimos: **eficiência**.
- Caminhos redundantes: **robustez**.
- Centralidade de intermediação: identificação de hubs.



Nature Reviews | Genetics

Barabási A-L, Oltvai ZN. *Nature Review Genetics*, 5, 2004

Para Saber Mais...

A-L Barabasi. *Linked a Nova Ciência dos Networks: como tudo está conectado a tudo e o que isso significa para os negócios, relações sociais e ciência*. São Paulo: Leopardo Editora, 2009.

A-L Barabasi. **Network Science**. <http://networksciencebook.com/>

B H Junker & F Schreiber. *Analysis of Biological Networks*. New Jersey: Willey InterScience, 2008.

Complex Systems and Networks. *Science*, vol 325, 2009.

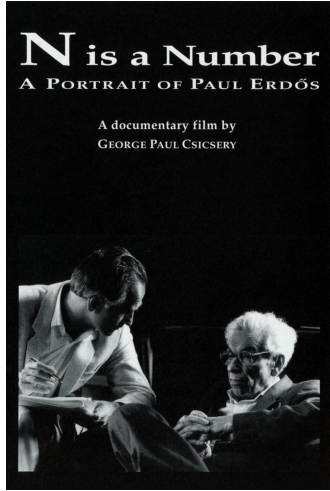
D J Watts. *Seis Graus de Separação*. São Paulo: Leopardo Editora, 2009.

D Noble. *The Music of Life – Biology Beyond Genes*. London: Oxford University Press, 2006.

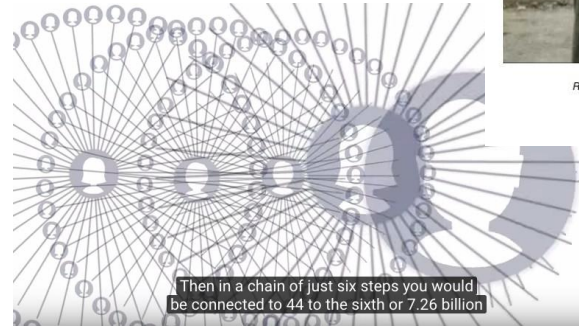
S Johnson. *Emergência – a vida integrada de formigas, cérebros, cidades e softwares*. Rio de Janeiro: Jorge Zahar, 2001

U Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Boca Raton: Chapman & Hall/CRC, 2007.





www.youtube.com/watch?v=dTzkrJKUo-l



www.youtube.com/watch?v=TcxZSmzPw8k



Read Aug. 1, 2014 [News at OU](#) article on the popularity of this website.

The Erdős Number Project

oakland.edu/enp/compute/



www.youtube.com/watch?v=BQ7UDWn_uw

S