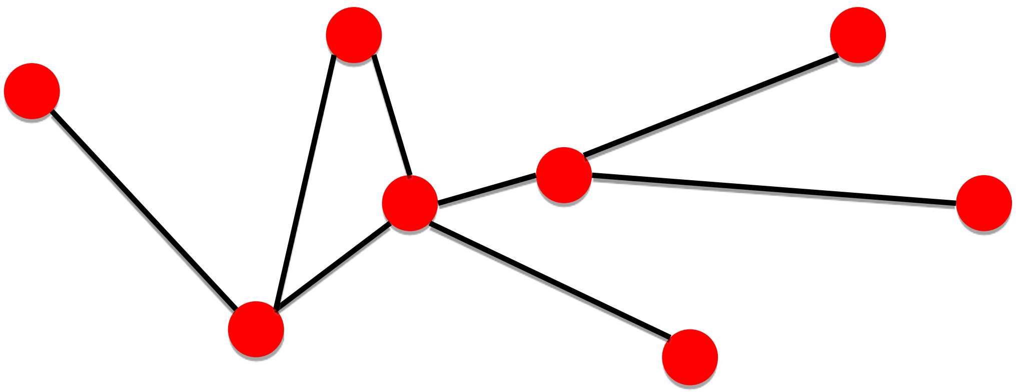
Ciência das Redes

Teoria de Grafos

Ricardo Luders
Thiago H Silva

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

Componentes de um sistema complexo



Componentes: nós, vértices   N
Interações: links, arestas   L
Sistema: rede, grafo   (N,L)

Redes ou grafos?

Redes geralmente se referem a sistemas reais
- WWW
- Rede social
- Rede metabólica

Linguagem: rede, nó, link

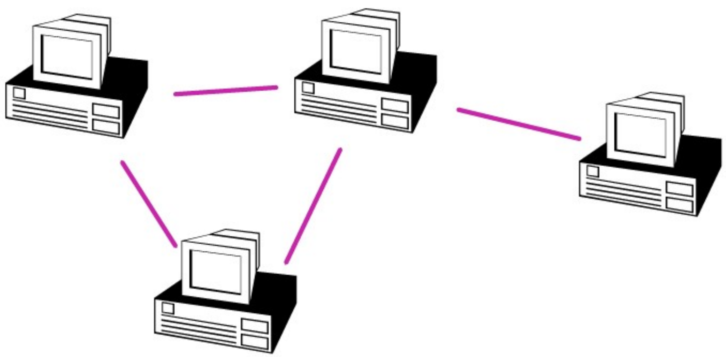
Grafo representação matemática de uma rede
- Web graph
- Grafo social (um termo do Facebook)

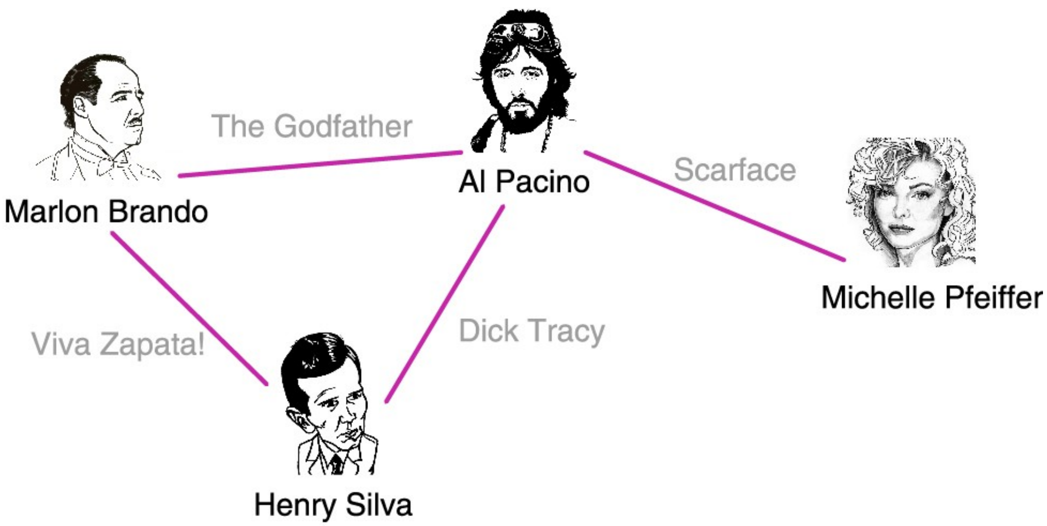
Linguagem: Grafo, vértice, aresta

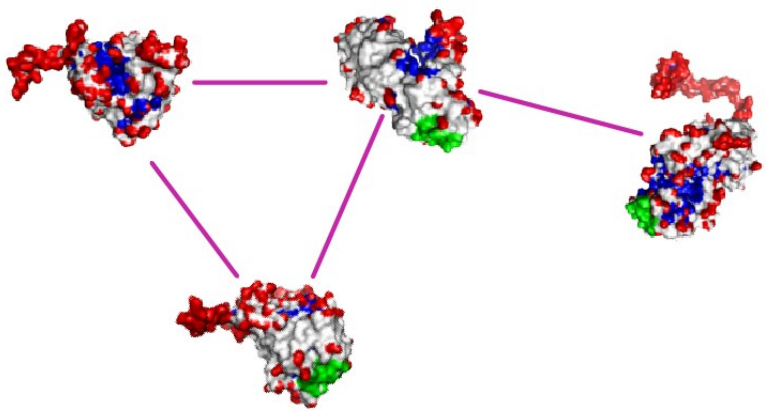
Neste curso, na maioria dos casos, usaremos os dois termos alternadamente.

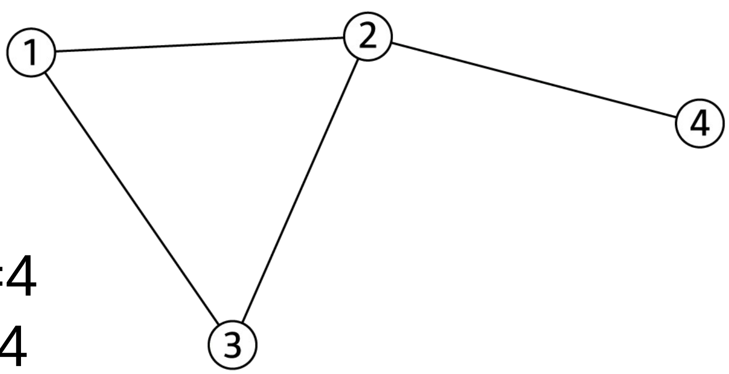
Uma linguagem comum

Várias redes diferentes, mas a mesma representação









N=4
L=4

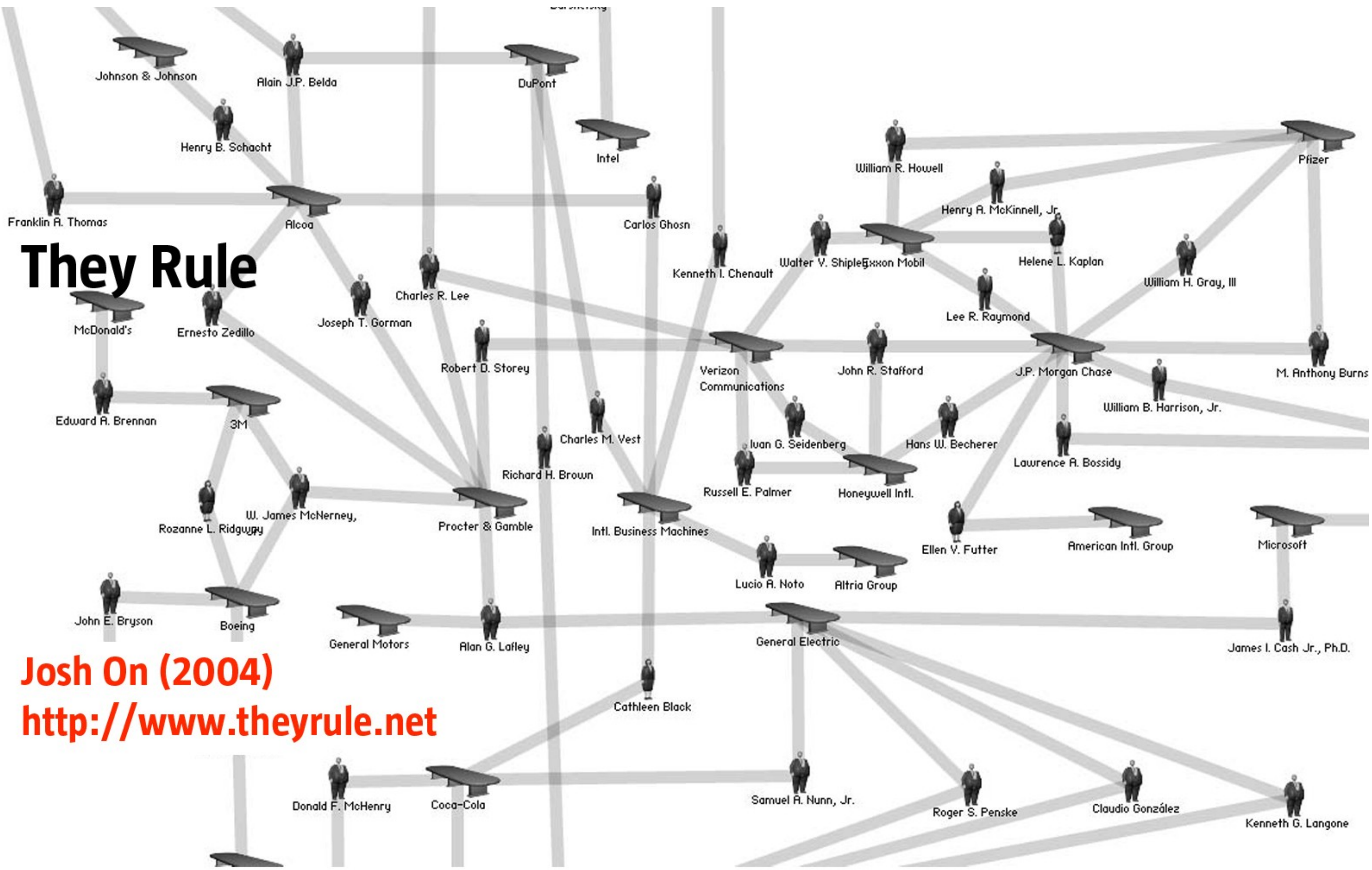
Representação apropriada

A escolha da representação de rede apropriada determina nossa habilidade de usar a teoria de rede com sucesso.

Em alguns casos, existe uma representação única e inequívoca. Em outros casos, a representação não é de forma alguma única.

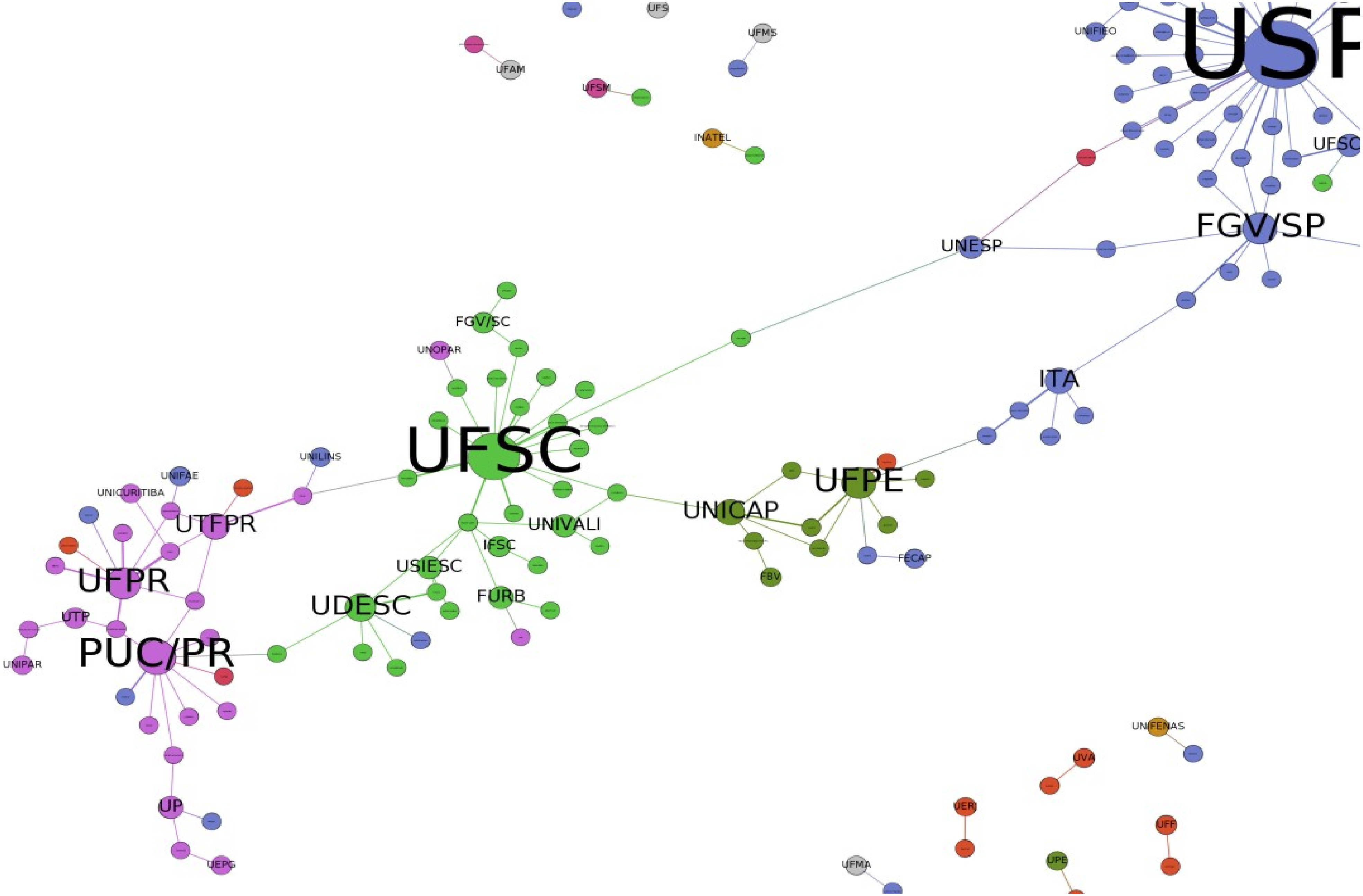
Por exemplo, a maneira como atribuímos os vínculos entre um grupo de indivíduos determinará a natureza da questão que podemos estudar.

Representação apropriada



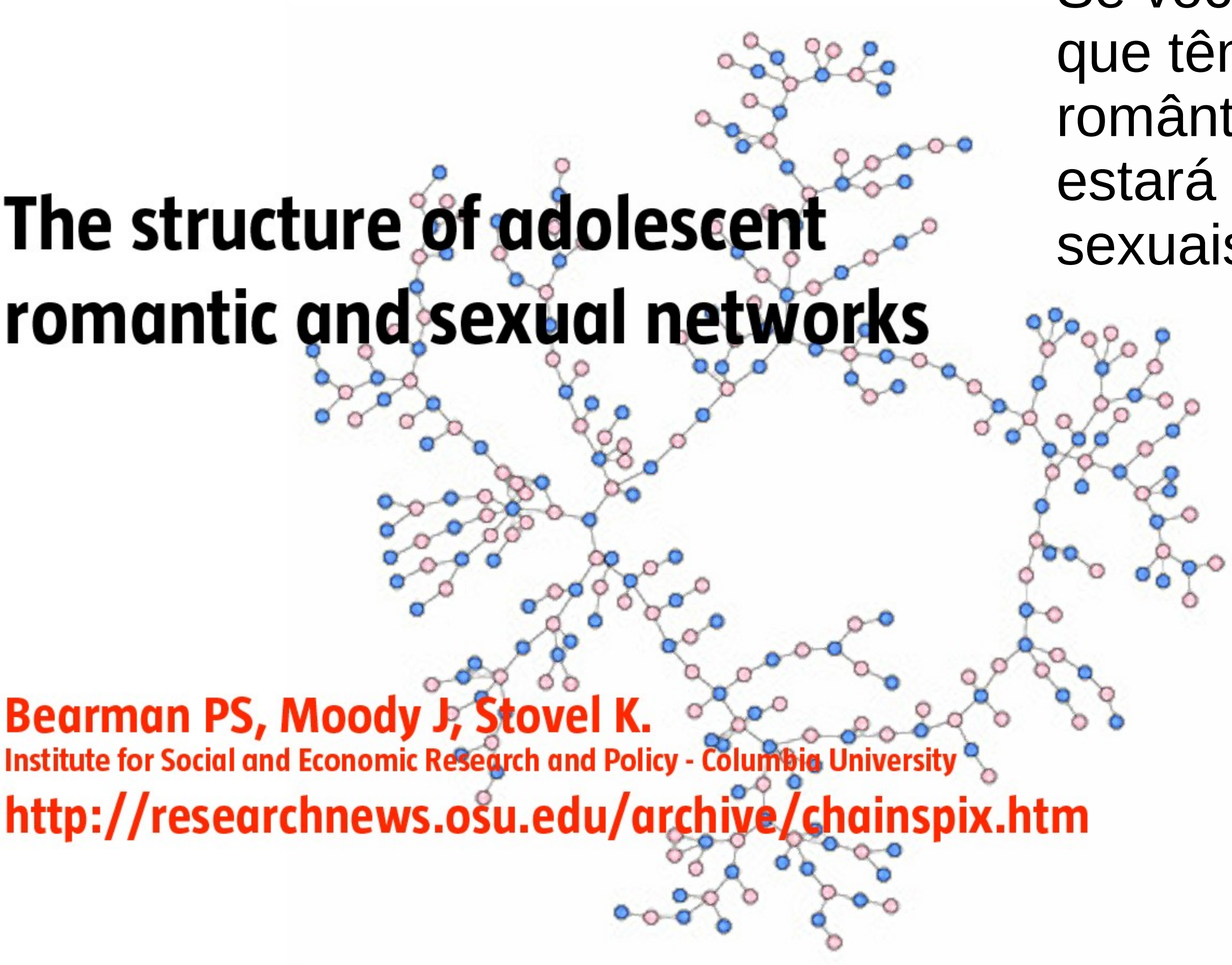
Se você conectar pessoas que trabalham umas com as outras, você explorará a rede profissional.

Representação apropriada



Representação apropriada

Se você conectar aqueles que têm um relacionamento romântico e sexual, você estará explorando as redes sexuais.



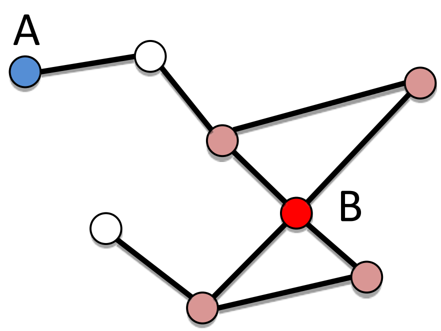
Representação apropriada

Se você conectar pessoas com base em seu primeiro nome (todos os Peters conectados uns aos outros), você explorará o quê?

É uma rede, no entanto.

Grau do nó

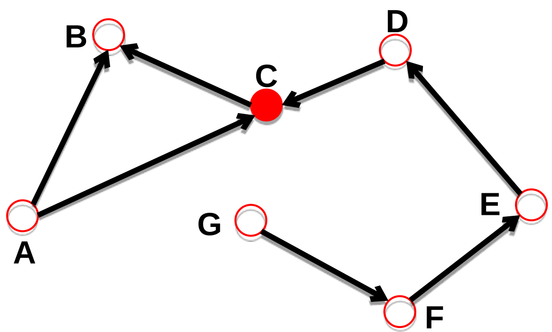
Não direcionado



Grau do nó: o número de links conectados ao nó.

$K_A = 1$ $K_B = 4$

Direcionado



Em redes direcionadas, podemos definir um grau de entrada e saída. O grau (total) é a soma do grau de entrada e saída.

$K_c^{in} = 2$

$K_c^{out} = 1$

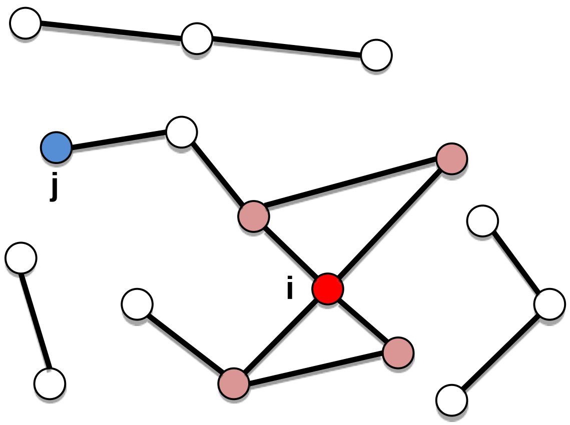
$K_c = 3$

Fonte (source): um nó com $K^{in} = 0$

Coletor (sink): um nó com $K^{out} = 0$

Grau médio

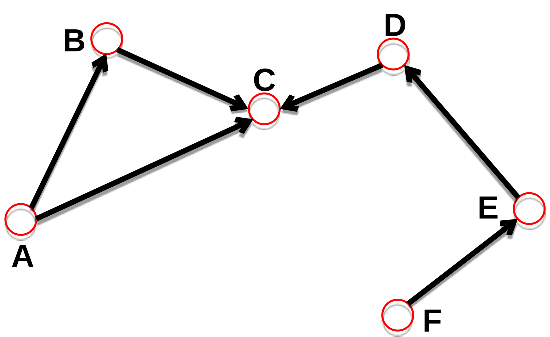
Não direcionado



$$\langle k \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i \qquad \langle k \rangle \equiv \frac{2L}{N}$$

N = número de nós no grafo

Direcionado

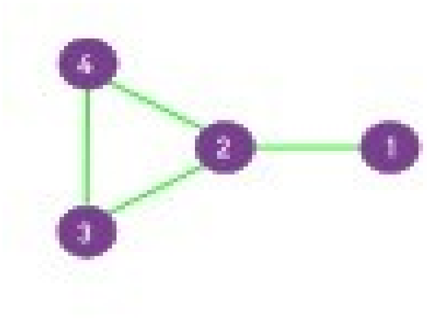


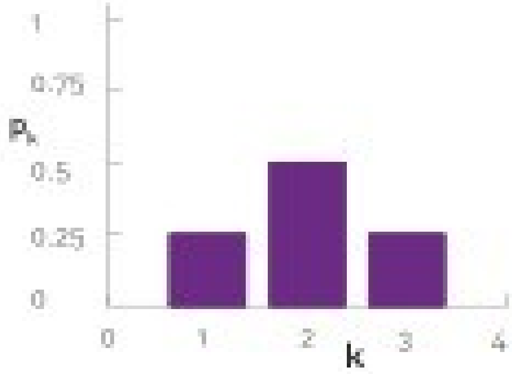
$$\langle k^{in} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{in}, \quad \langle k^{out} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{out}, \quad \langle k^{in} \rangle = \langle k^{out} \rangle$$

$$\langle k \rangle \equiv \frac{L}{N}$$

Distribuição de graus

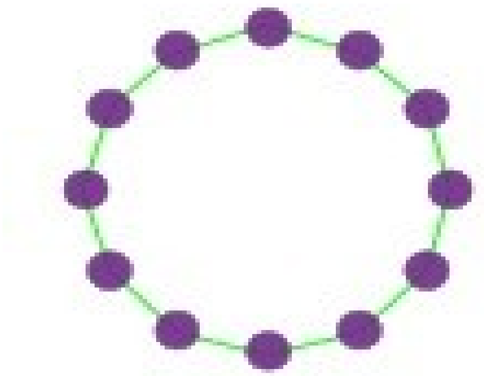
Distribuição de graus P(k): probabilidade de que um nó escolhido aleatoriamente tem grau k

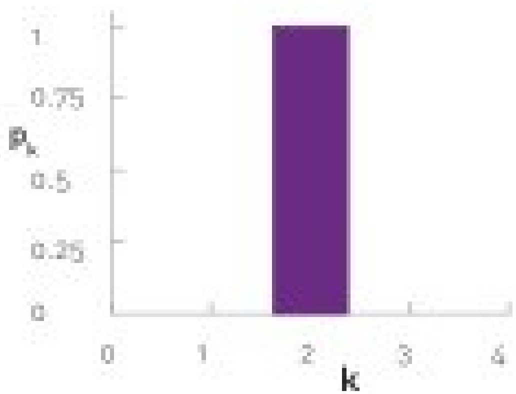




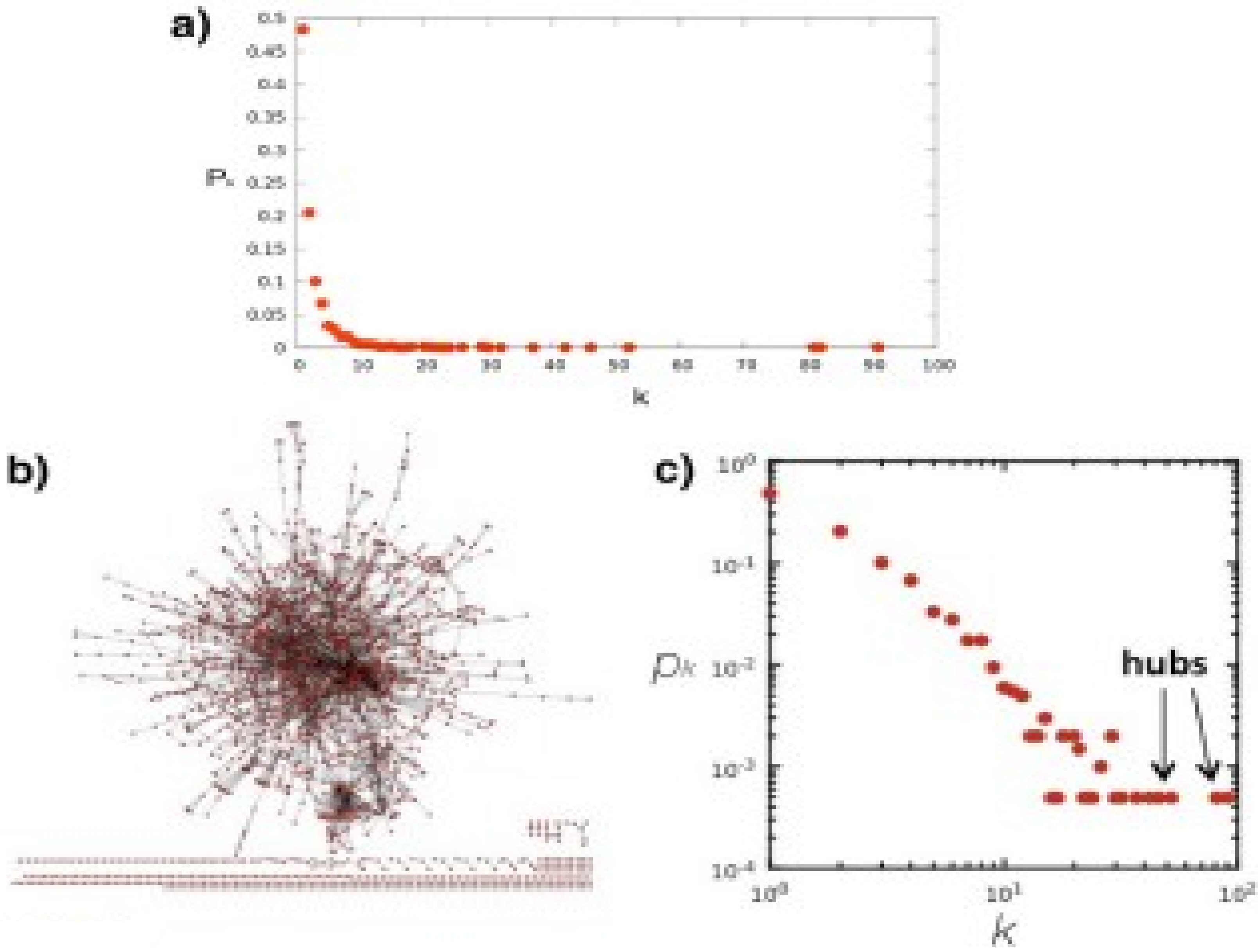
N_k = # de nós com grau k

$P(k) = N_k / N$





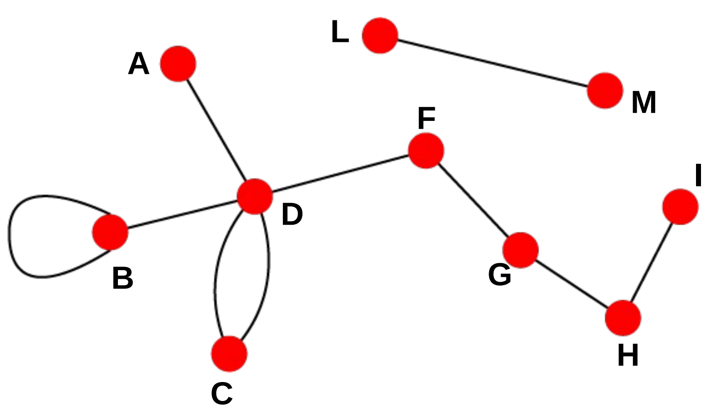
Distribuição de graus



Redes direcionadas vs direcionadas

Não direcionada

Links: não direcionados (simétricos)

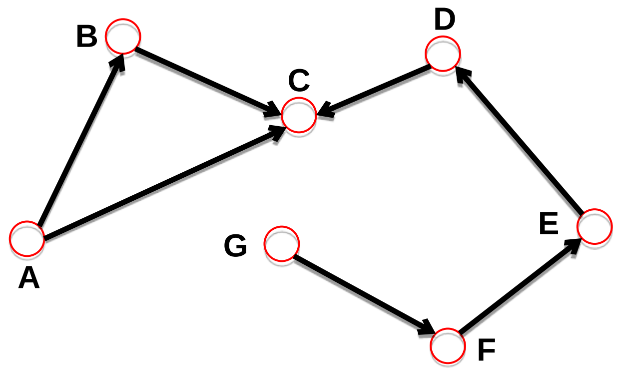


Exemplos de links:

Links de Coautoria
Rede de atores
Interações de proteínas

Direcionada

Links: direcionados (arcos)



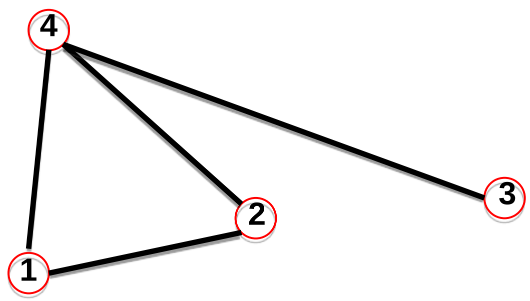
Exemplos de links:

URLs na WWW
Rede de chamadas telefônicas
Reações metabólicas

Grau médio e outras informações

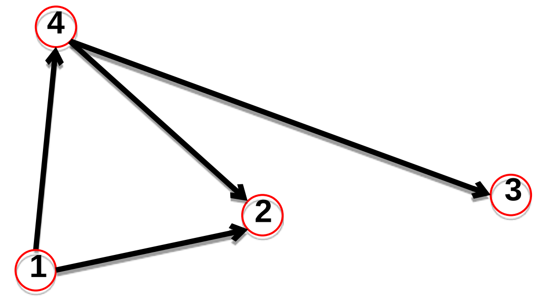
NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	⟨k⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826	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	2.90

Matriz de adjacência



A_{ij} = 1 se houver uma ligação entre o nó i e j
A_{ij} =0 se os nós i e j não estiverem conectados um ao outro.

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



A_{ij} = 1 se houver um link apontando do nó j e i

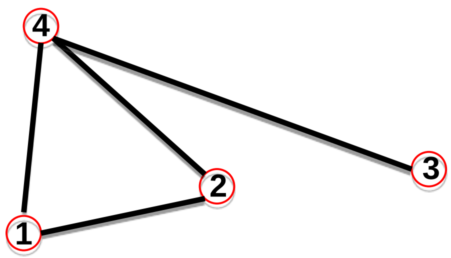
A_{ij} = 0 se não houver nenhum link apontando de j para i.

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

Observe que para um gráfico direcionado, a matriz não é simétrica.

Matriz de adjacência e grau dos nós

Não direcionado



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

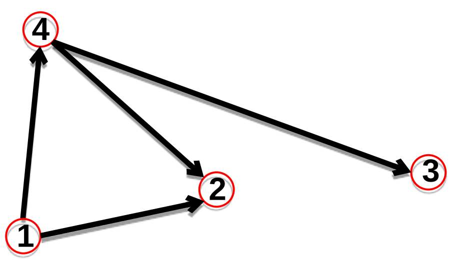
$$k_i = \sum_{j=1}^{N} A_{ij}$$

$$k_j = \sum_{i=1}^{N} A_{ij}$$

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

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

Direcionado



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

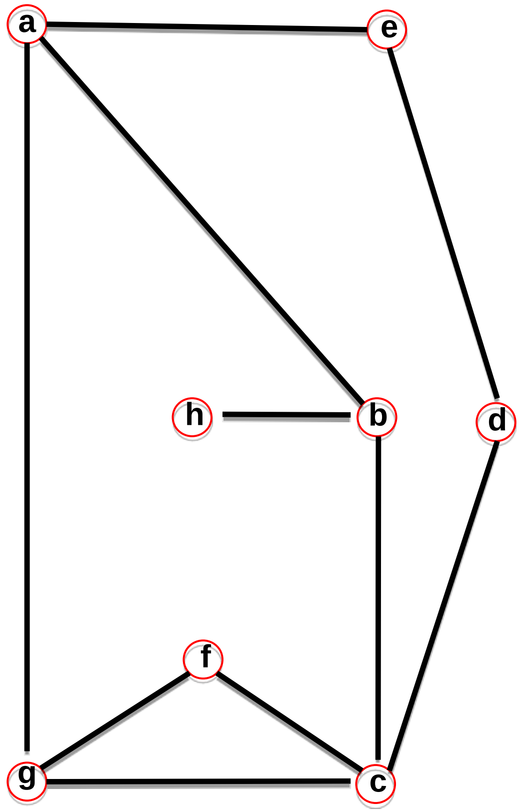
$$k_i^{in} = \sum_{j=1}^{N} A_{ij}$$

$$k_j^{out} = \sum_{i=1}^{N} A_{ij}$$

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

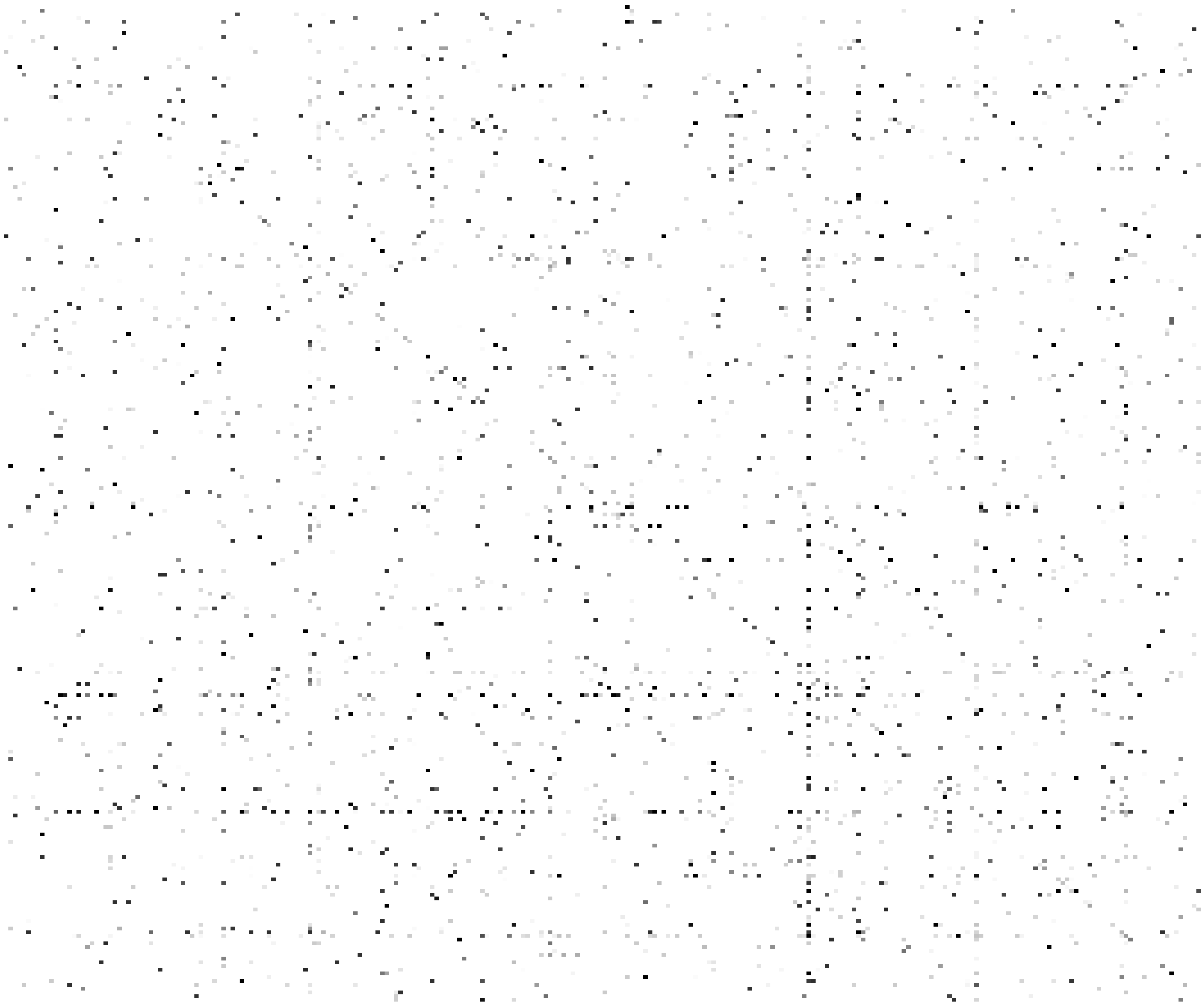
$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{j=1}^{N} k_j^{out} = \sum_{i,j}^{N} A_{ij}$$

Matriz de adjacência



	a	b	c	d	e	f	g	h
a	0	1	0	0	1	0	1	0
b	1	0	1	0	0	0	0	1
c	0	1	0	1	0	1	1	0
d	0	0	1	0	1	0	0	0
e	1	0	0	1	0	0	0	0
f	0	0	1	0	0	0	1	0
g	1	0	1	0	0	0	0	0
h	0	1	0	0	0	0	0	0

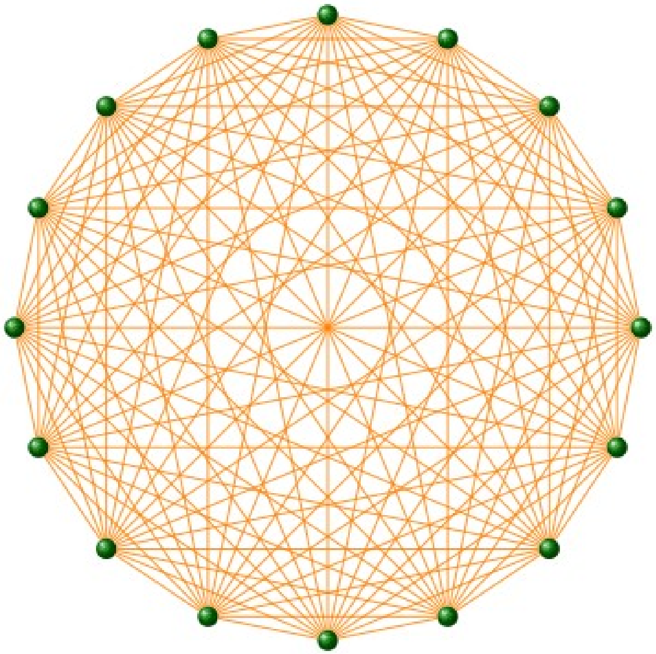
Matrizes de adjacência são esparsas



Grafos completos

O número máximo de links de uma rede de N nós pode ter é:

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



Um grafo com grau L = L_{max} é chamado de gráfo completo, e seu grau médio é **$\langle k \rangle = N-1$**

Redes reais são esparsas

A maioria das redes observadas em sistemas reais são esparsas:

$$L << L_{max}$$

ou

$$\langle k \rangle << N-1.$$

WWW (ND Sample):	N=325.729;	$L=1{,}4\ 10^{6}$	$L_{max}=10^{12}$	$\langle k \rangle=4{,}51$
Protein (*S. Cerevisiae*):	N= 1.870;	L=4.470	$L_{max}=10^{7}$	$\langle k \rangle=2{,}39$
Coauthorship (Math):	N= 70.975;	$L=2\ 10^{5}$	$L_{max}=3\ 10^{10}$	$\langle k \rangle=3{,}9$
Movie Actors:	N=212.250;	$L=6\ 10^{6}$	$L_{max}=1{,}8\ 10^{13}$	$\langle k \rangle=28{,}78$

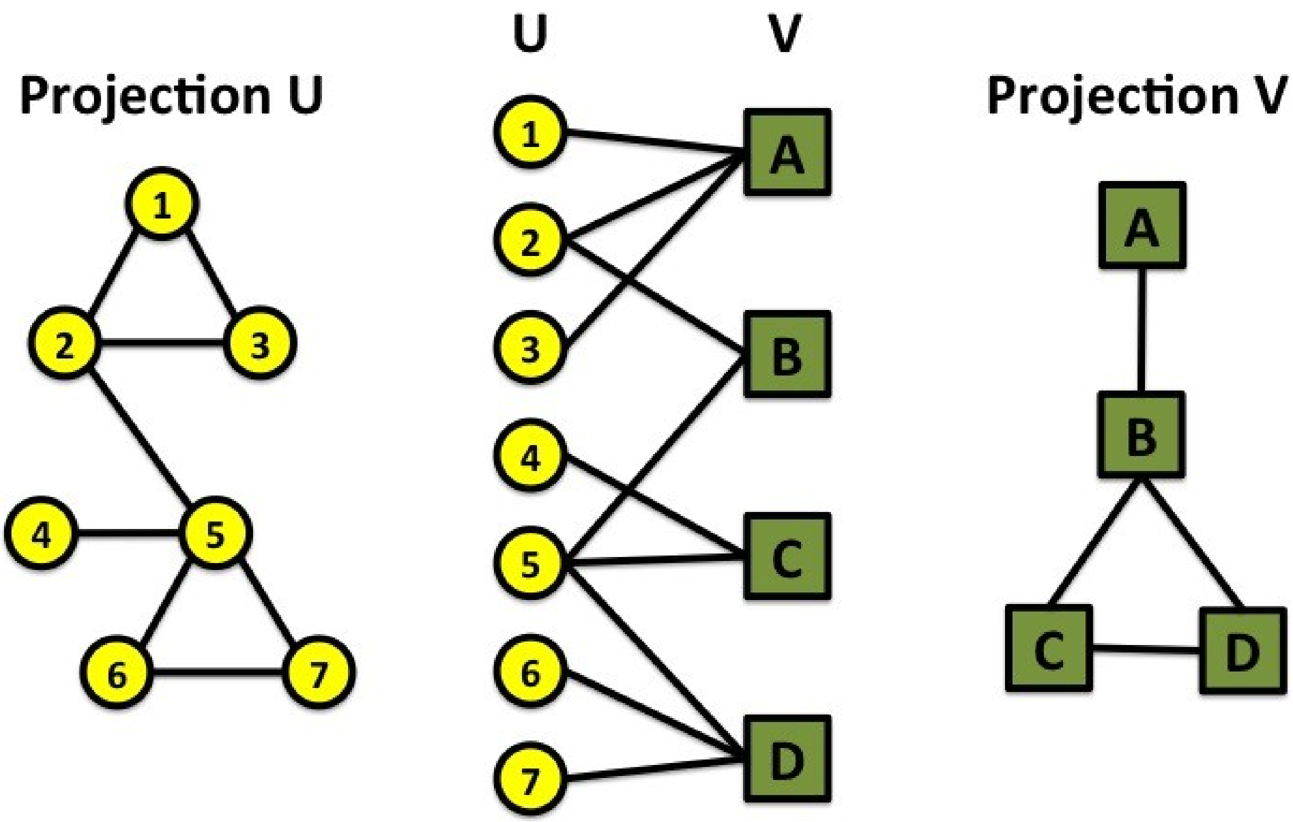
(Fonte: Albert, Barabasi, RMP2002)

Redes com pesos

$$A_{ij} = w_{ij}$$

Redes bipartidas

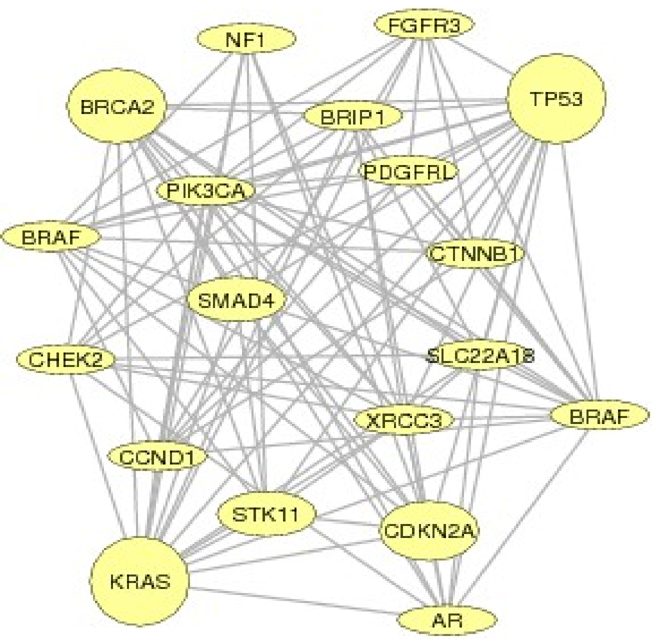
Grafo bipartido (ou bigraph) é um grafo cujos nós podem ser divididos em dois conjuntos disjuntos U e V, de modo que cada aresta conecta um nó em U a um em V; ou seja, U e V são conjuntos independentes.



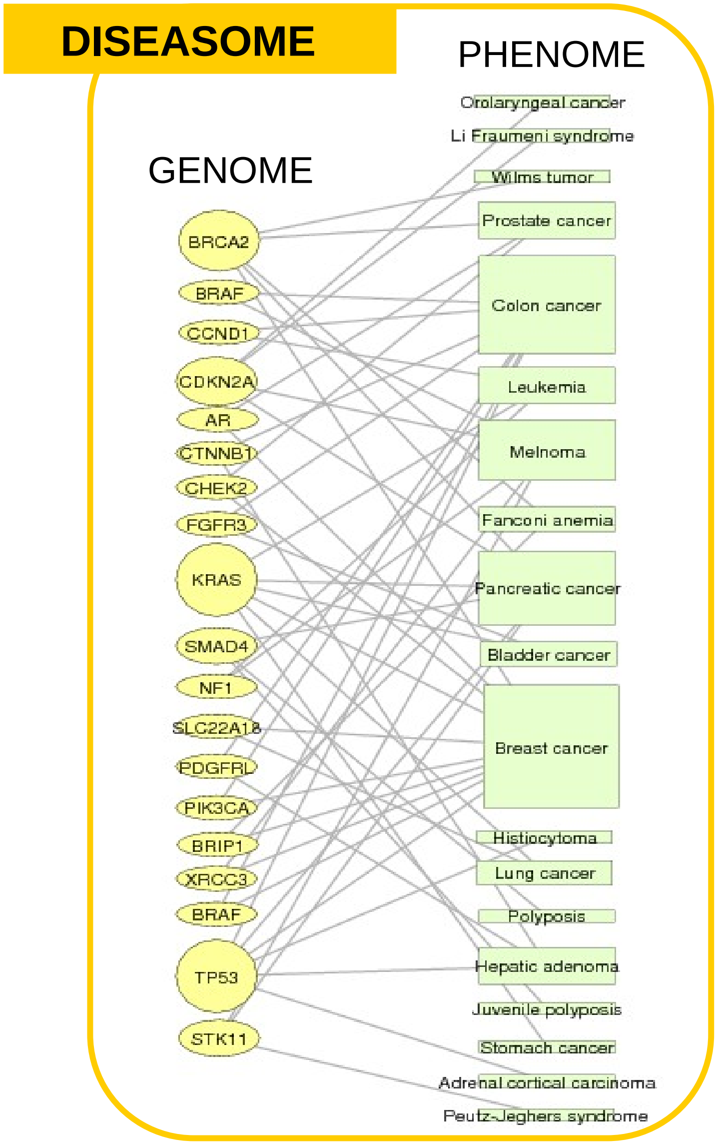
Exemplos:

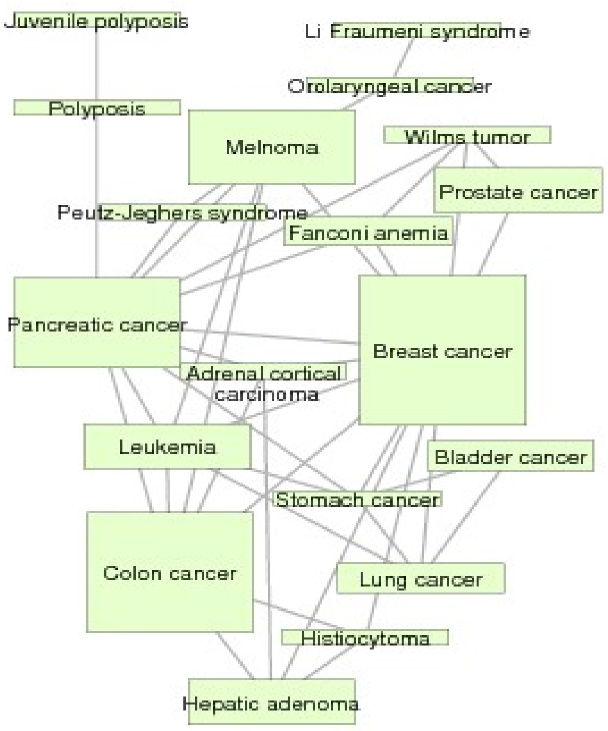
Rede de atores de Hollywood
Rede de doenças (*diseasome*)

Redes bipartidas



Gene network

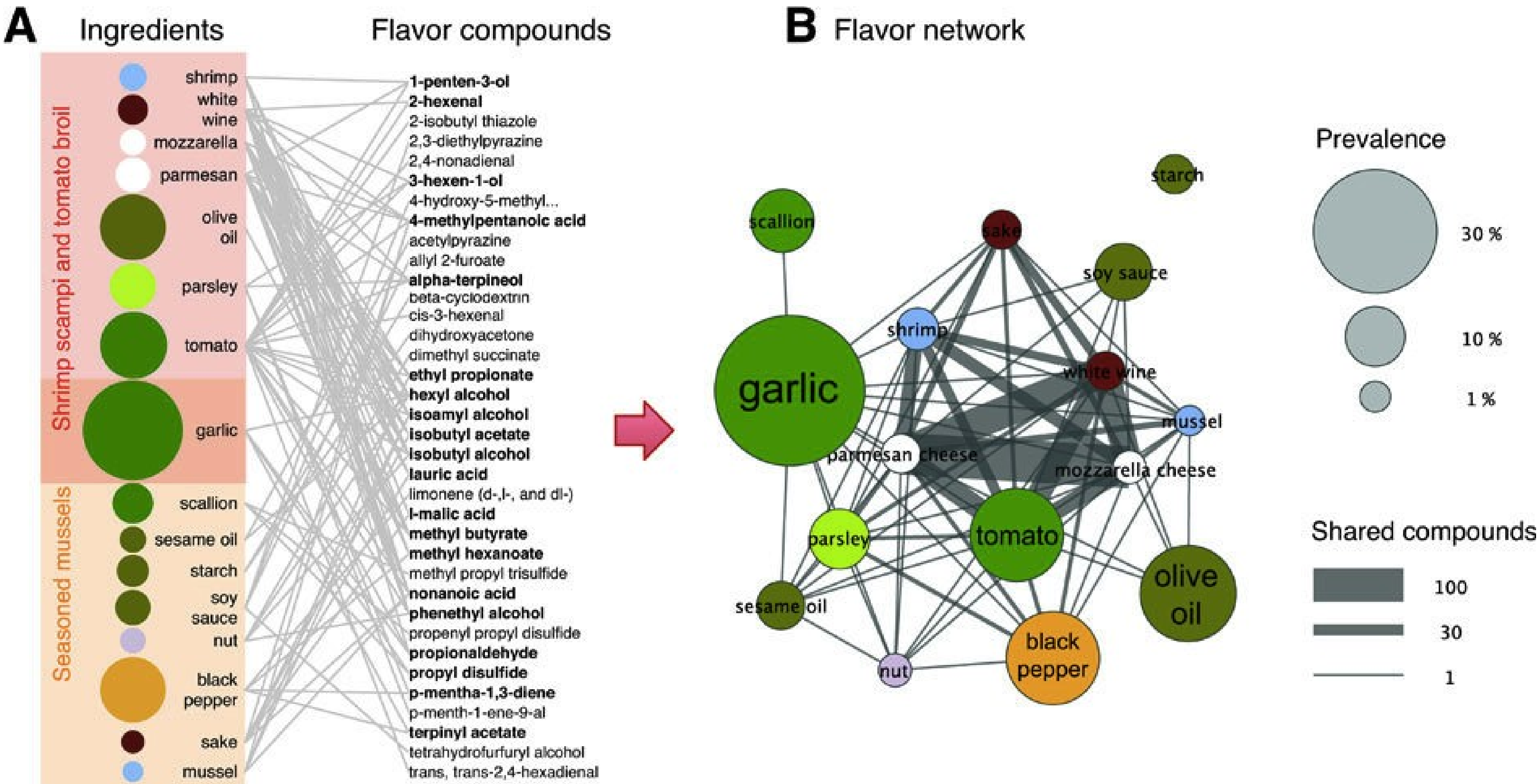




Disease network

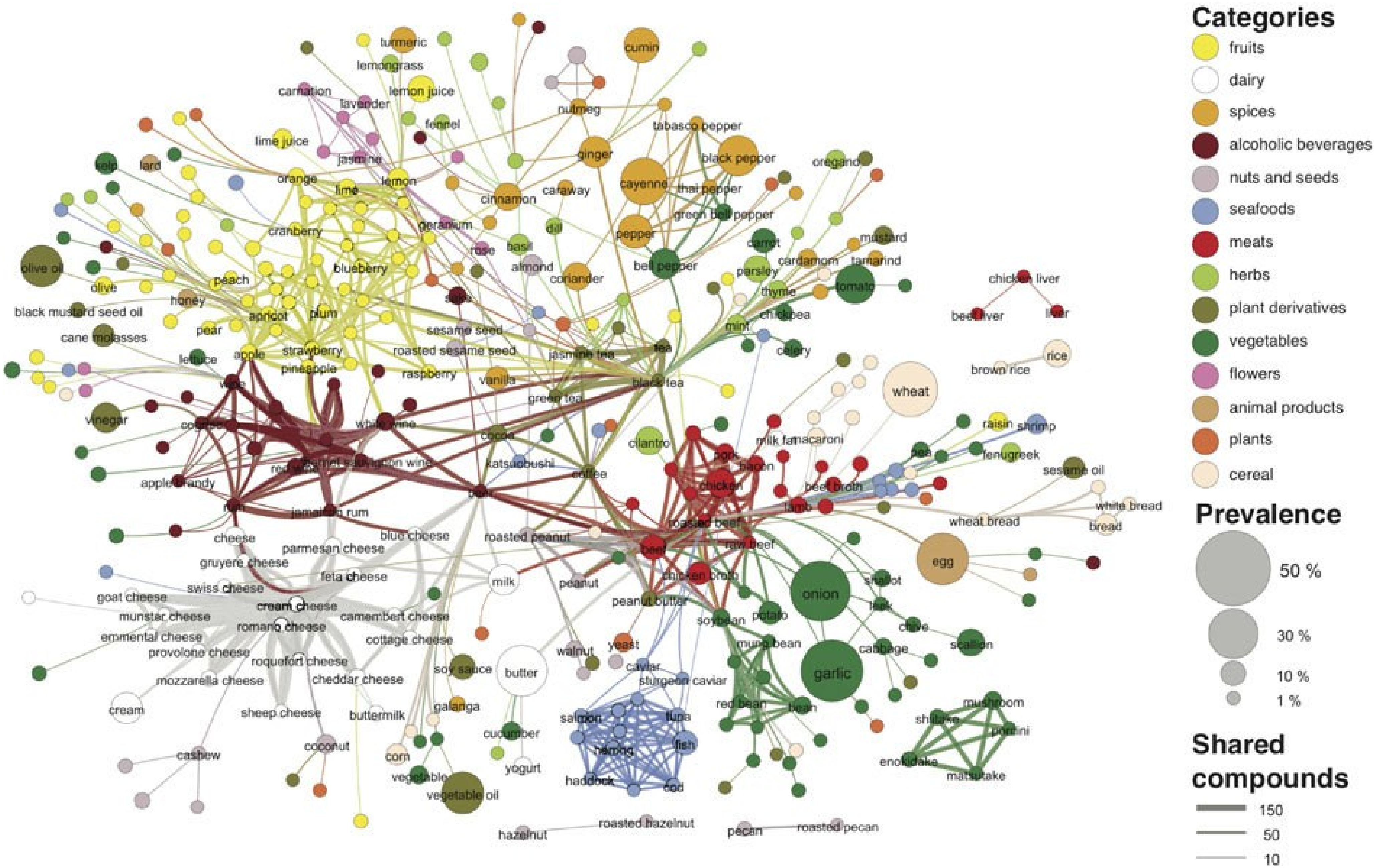
Goh, Cusick, Valle, Childs, Vidal & Barabási, PNAS (2007)

Redes bipartidas – ingrediente e sabor



Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, A.-L. Barabási Flavor network and the principles of food pairing , Scientific Reports 196, (2011).

Redes bipartidas – ingrediente e sabor

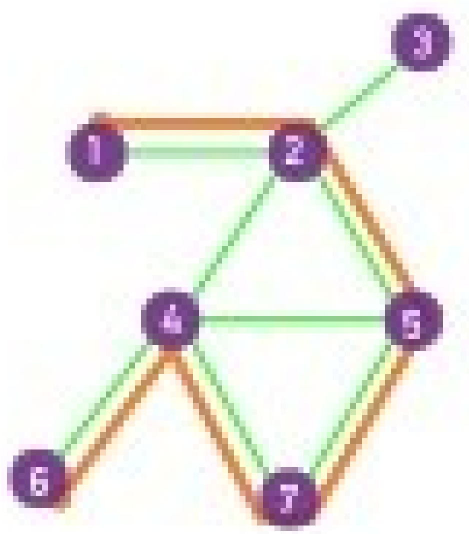


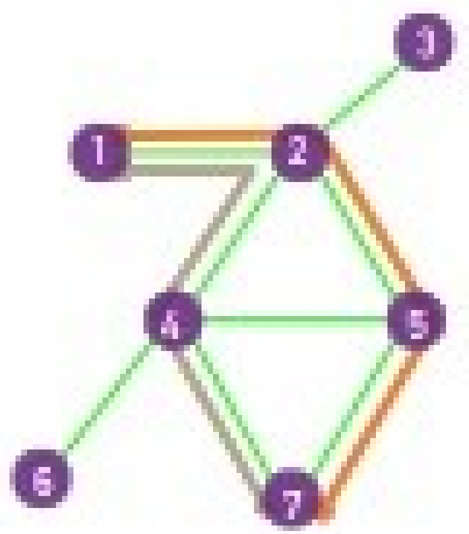
Caminhos

Um caminho é uma sequência de nós em que cada nó é adjacente ao próximo

$P_{i0,in}$ de comprimento n entre os nós i0 e in é uma coleção ordenada de n + 1 nós e n links

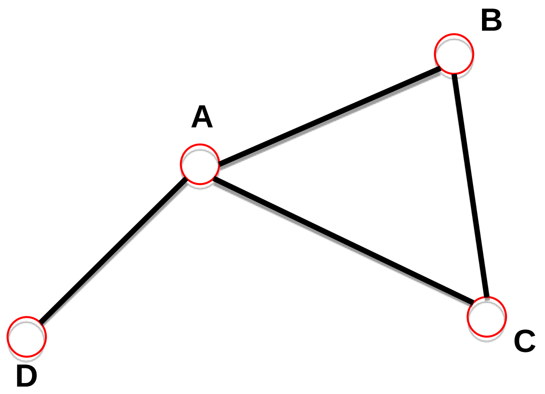
$$P_n = \{i_0, i_1, i_2, \ldots, i_n\} \qquad P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{n-1}, i_n)\}$$





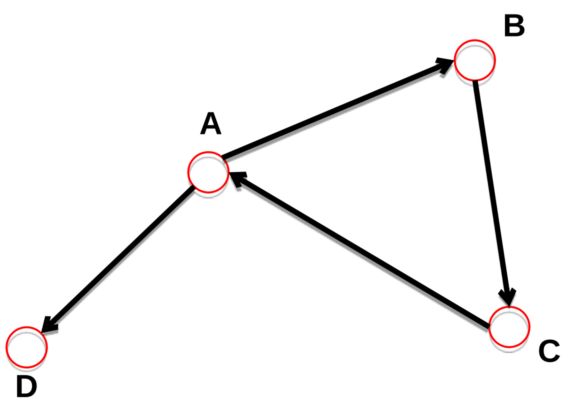
Em uma rede direcionada, o caminho pode seguir apenas a direção de uma seta.

Caminhos



A distância (caminho mais curto, caminho geodésico) entre dois nós é definida como o número de arestas ao longo do caminho mais curto que os conecta.

** Se os dois nós estiverem desconectados, a distância é infinita.*



Em grafos direcionados, cada caminho precisa seguir a direção das setas.

Assim, em um grafo direcionado, a distância do nó A a B é geralmente diferente da distância do nó B a A.

Diâmetro da rede e distância média

Diâmetro: d_{max} é a distância máxima entre qualquer par de nós no gráfico.

Comprimento / distância média do caminho, <d>, para um **gráfico conectado**:

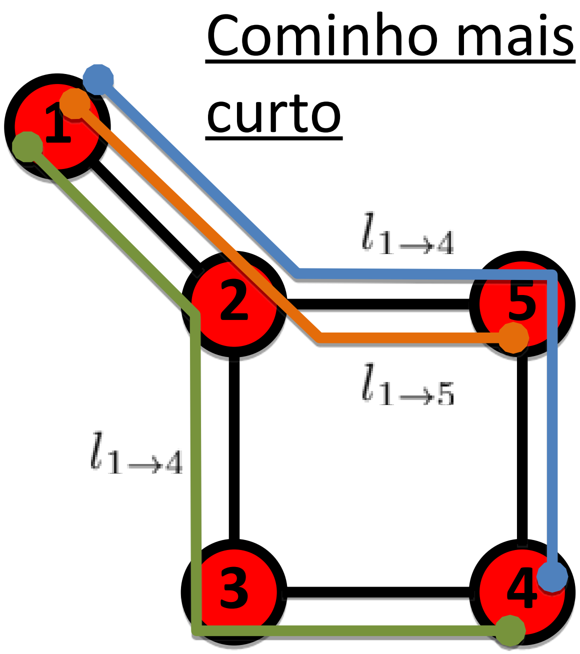
$$\langle d \rangle \equiv \frac{1}{2L_{max}} \sum_{i,j\neq i} d_{ij}$$

onde d_{ij} é a distância do nó i ao nó j

Em um gráfico não direcionado $d_{ij} = d_{ji}$, então só precisamos contá-los uma vez:

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

Caminho mais curso



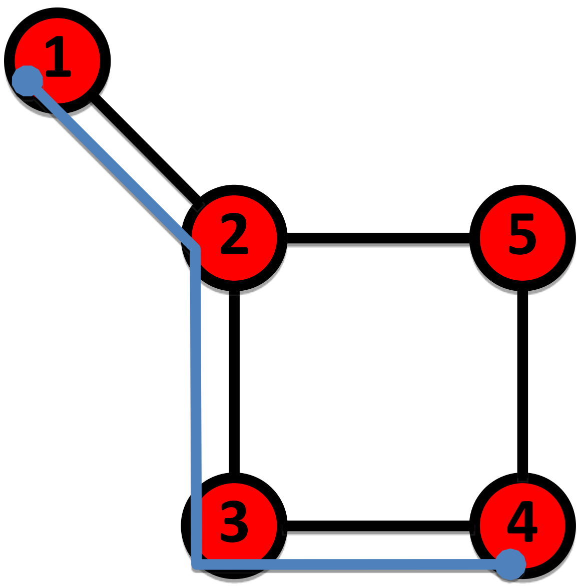
$$l_{1\to4} = 3$$

$$l_{1\to5} = 2$$

O caminho com o menor comprimento entre dois nós (distância).

Diâmetro e comprimento médio do caminho

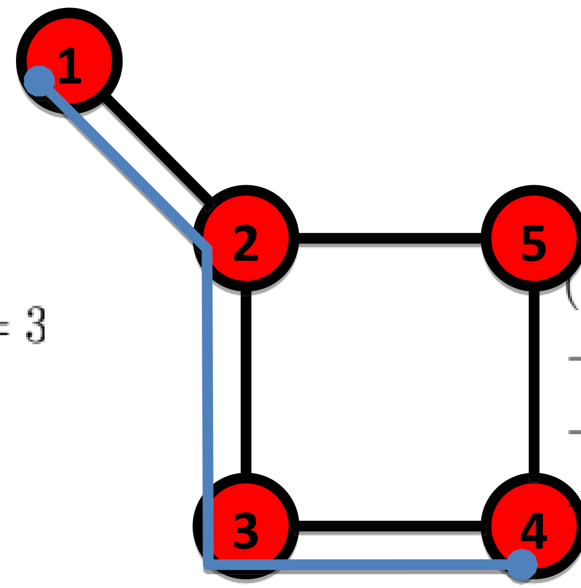
Diâmetro



$$l_{1\to4} = 3$$

O caminho mais curto mais longo em um gráfico

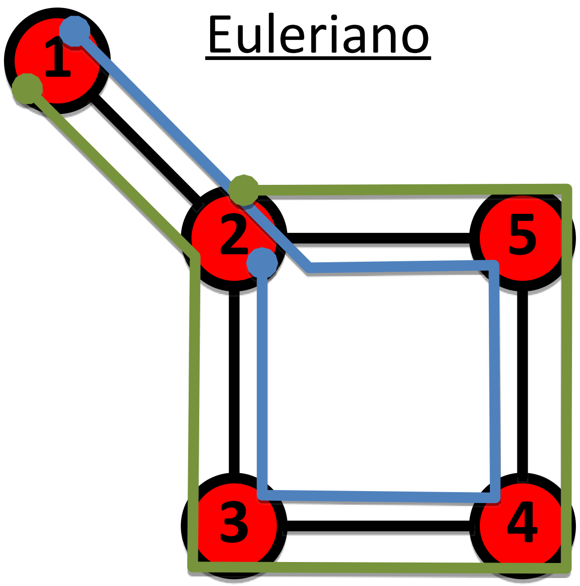
Comprimento Médio do Caminho



$$\begin{aligned}&(l_{1\to2} + l_{1\to3} + l_{1\to4} + \\ &+ l_{1\to5} + l_{2\to3} + l_{2\to4} + \\ &+ l_{2\to5} + l_{3\to4} + l_{3\to5} + \\ &+ l_{4\to5})/10 = 1.6\end{aligned}$$

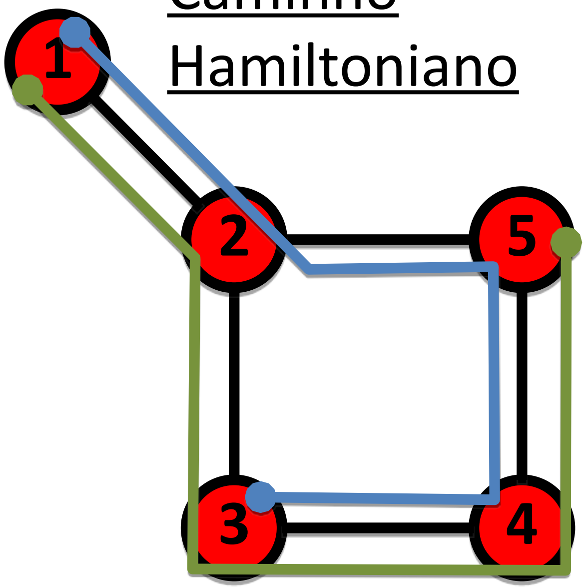
A média dos caminhos mais curtos para todos os pares de nós.

Outros caminhos



Caminho Euleriano

Um caminho que atravessa cada link exatamente uma vez.

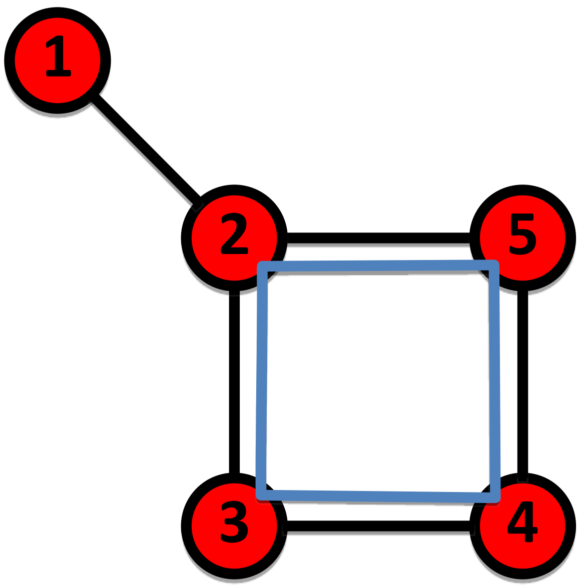


Caminho Hamiltoniano

Um caminho que visita cada nó exatamente uma vez.

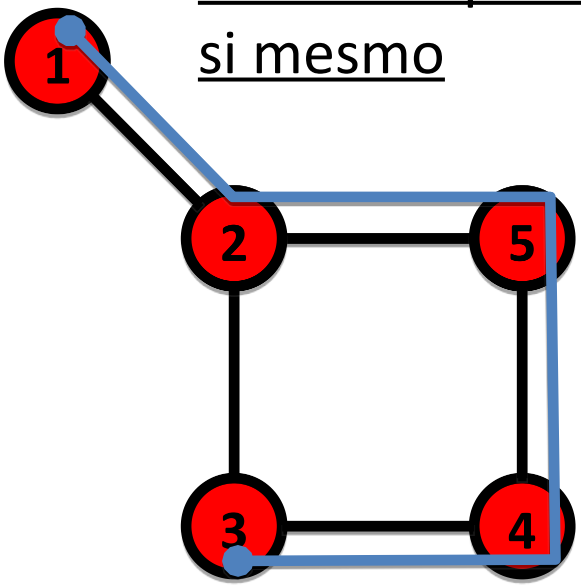
Ciclo e caminho que evita a si mesmo

<u>Ciclo</u>



Um caminho com o mesmo nó inicial e final.

<u>Caminho que evita a si mesmo</u>

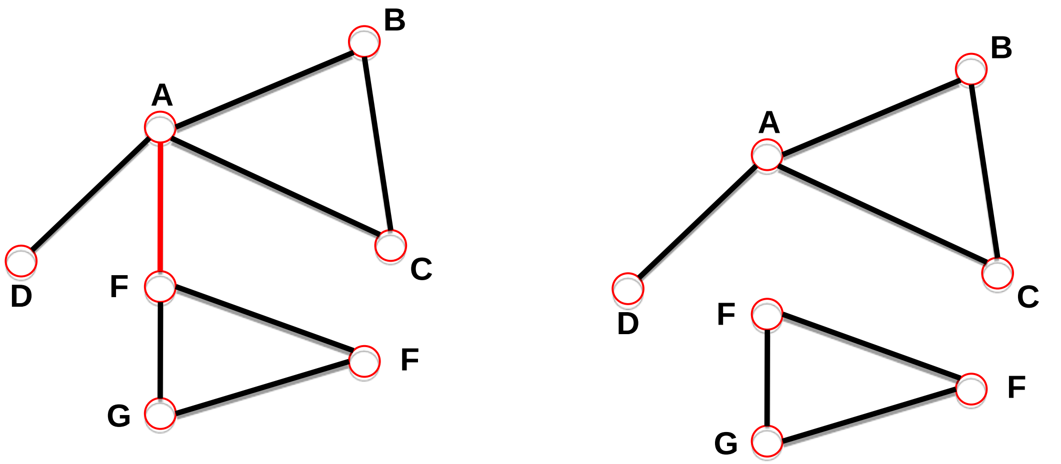


Um caminho que não se cruza.

Conectividade de grafos não direcionados

Grafo conectado (não direcionado): quaisquer dois vértices podem ser unidos por um caminho.

Um gráfico desconectado é composto por dois ou mais componentes conectados.



Maior componente: **Componente Gigante**

O resto: **Isolados**

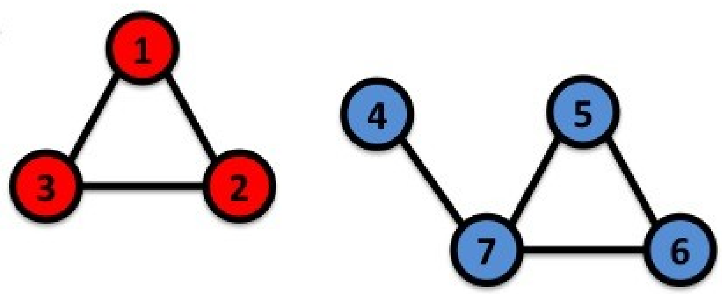
Ponte: se o apagarmos, o grafo se desconecta.

Conectividade de grafos não direcionados

UTFPR

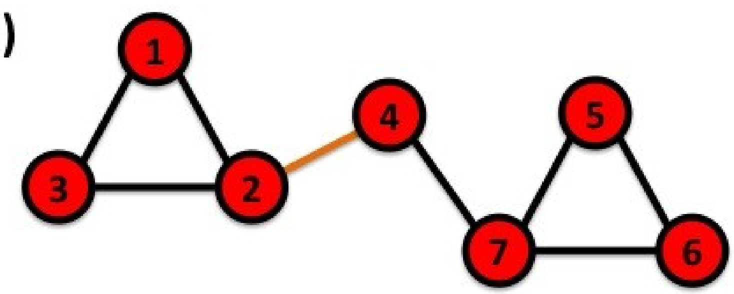
A matriz de adjacência de uma rede com vários componentes pode ser escrita na forma de bloco diagonal, de modo que elementos diferentes de zero sejam confinados a quadrados, com todos os outros elementos sendo zero:

(a)



$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

(b)



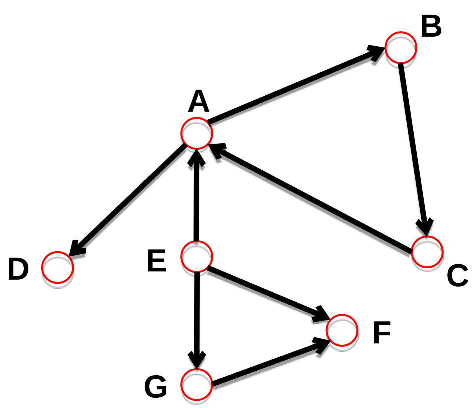
$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

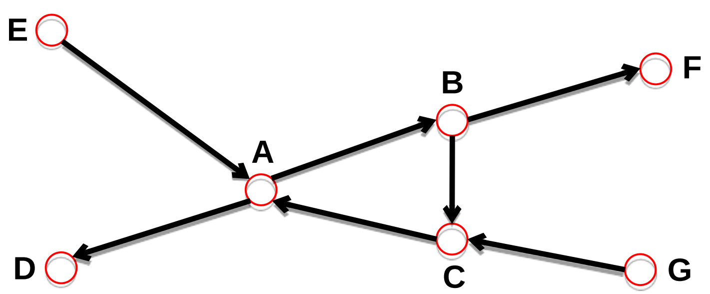
Conectividade de grafos direcionados

Grafo direcionado **fortemente conectado**: tem um caminho de cada nó para cada outro nó **e vice-versa** (por exemplo, caminho AB e caminho BA).

Gráfico direcionado **fracamente conectado**: ele está conectado se desconsiderarmos o direções das arestas.

Componentes fortemente conectados podem ser identificados, mas nem todo nó faz parte de um componente não trivial fortemente conectado.



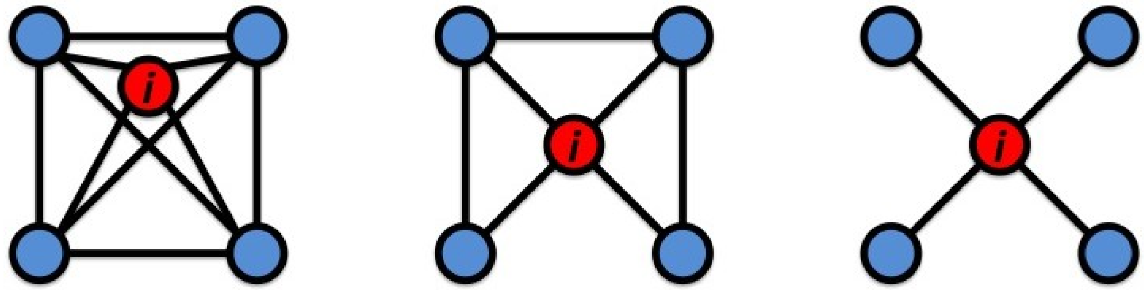


Componente interno: nós que podem alcançar o componente fortemente conectado,
Componente externo: nós que podem ser alcançados a partir do componente fortemente conectado.

Coeficiente de clusterização

Que fração de seus vizinhos está conectada?

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



$C_i = 1$ $C_i = 1/2$ $C_i = 0$

K_i = grau do nó i

e_i = número de links entre os k_i vizinhos do nó i

C_i em [0,1]

Watts & Strogatz, Nature 1998.

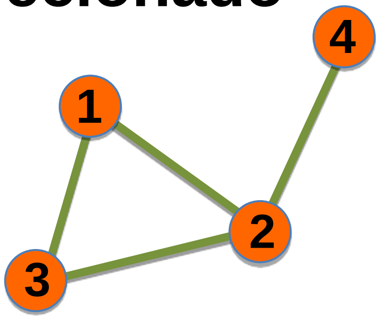
Três medidas centrais em ciência das redes

Distribuição dos graus: P (k)

Comprimento de caminho: <d>

Coeficiente de agrupamento: $C_i = \frac{2e_i}{k_i(k_i - 1)}$

Não direcionado



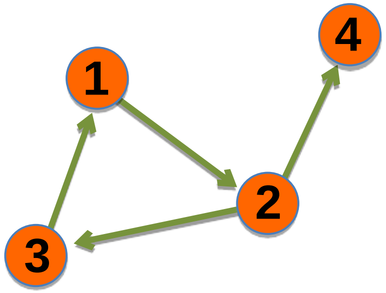
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

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

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{2L}{N}$$

Actor network, protein-protein interactions

Direcionado



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

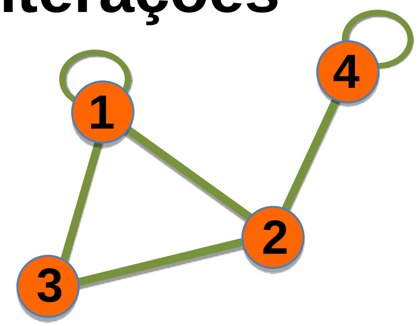
$$A_{ii} = 0 \qquad A_{ij} \neq A_{ji}$$

$$L = \sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{L}{N}$$

WWW, citation networks

Grafologia

Auto interações



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

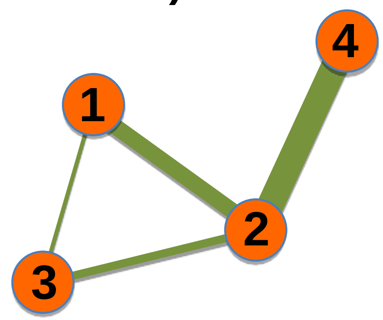
$$A_{ii} \neq 0 \qquad A_{ij} = A_{ji}$$

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

Protein interaction network, www

Ponderada

(não direcionada)



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

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

$$L = \frac{1}{2} \sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad <k> = \frac{2L}{N}$$

Call Graph, metabolic networks

Redes reais podem ter muitas características

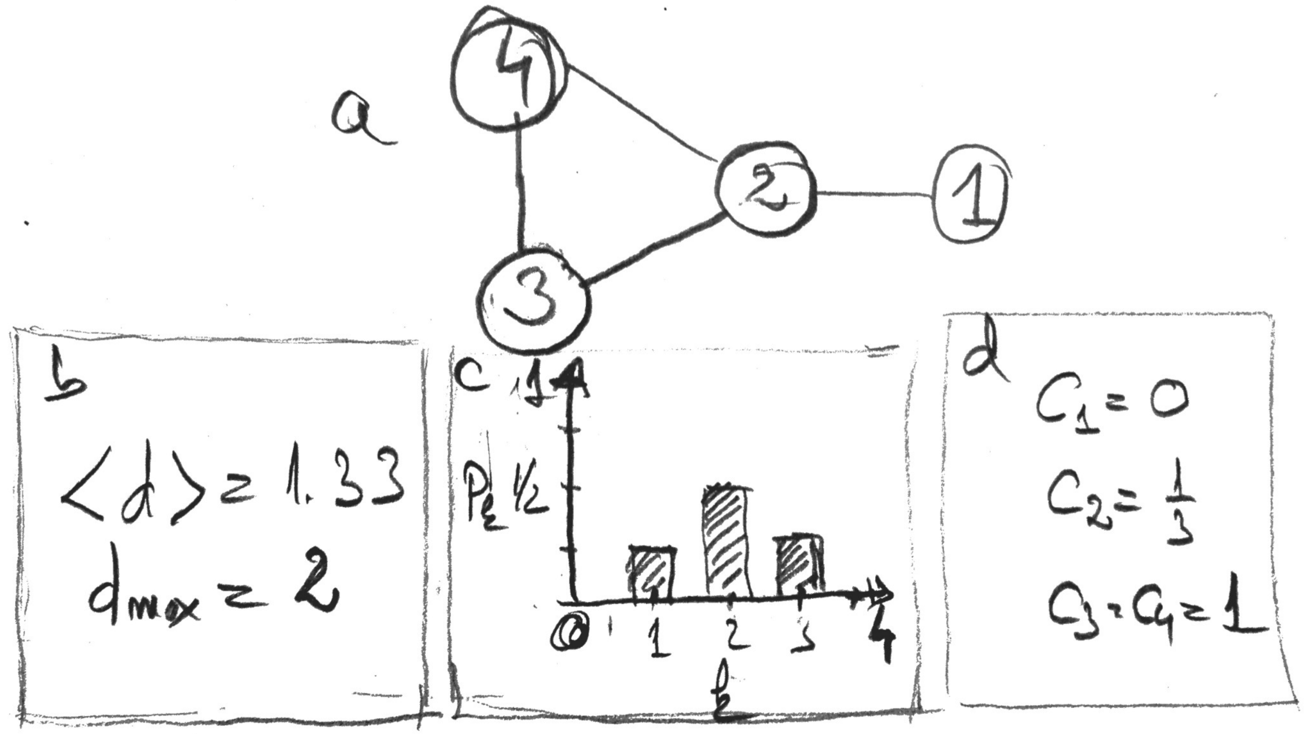
WWW -> dirigido com auto interações

Interações de proteínas -> não direcionadas não ponderadas com auto-interações

Chamadas de celular -> direcionadas, ponderadas.

Links de amizade do Facebook -> não direcionado, não ponderado.

Redes reais podem ter muitas características



Distribuição de graus: pk
Comprimento de caminho: <d>
Coeficiente de agrupamento: $C_i = \dfrac{2e_i}{k_i(k_i - 1)}$

Agradecimentos

Diversos trabalhos citados nos slides

Barabasi, Albert-László. Network Science. Cambridge University Press, 2016, 475p. ISBN 1107076269. Disponível online.

Newman, Mark. Networks. Oxford university press, 2018, 800p. ISSN 0198805098.

Alguns dos slides foram adaptados com autorização do Prof. Barabasi.