# Análise de Redes Avançada

Aula 3

# Medidas

# TAMANHO E DENSIDADE DA REDE

- O **tamanho de uma rede** é caracterizado pelo número de nós e arestas que ela contém.

- A **densidade de uma rede** é a fração entre 0 e 1 que nos diz qual parte de todas as arestas possíveis são realmente realizadas na rede. Para uma rede $G$ feita de $n$ nós e $m$ arestas, a densidade de $G$ é dada por:

$$\rho(G) = \frac{m}{\frac{n(n-1)}{2}} = \frac{2m}{n(n-1)}$$

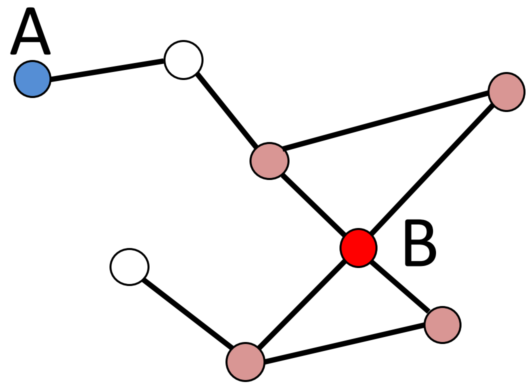
, para uma rede não direcionada

$$\rho(G) = \frac{m}{n(n-1)}$$

, para uma rede direcionada

# GRAU DE UM NÓ

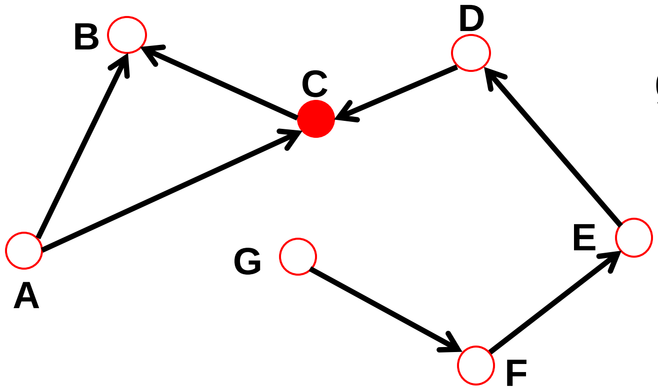
Não direcionado




Grau de um nó: o número de ligações desse nó

$$k_A = 1 \qquad k_B = 4$$

Direcionado




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

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

Fonte: um nó com $k^{in} = 0$; Ralo: um nó com $k^{out} = 0$.

# ALGUMAS MÉTRICAS ESTATÍSTICAS

## Média

$$\langle x \rangle = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

## N-ésimo momento

$$\langle x^n \rangle = \frac{x_1^n + x_2^n + \ldots + x_N^n}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i^n$$

## Desvio padrão

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \langle x \rangle\right)^2}$$

## Distribuição de $p_x$

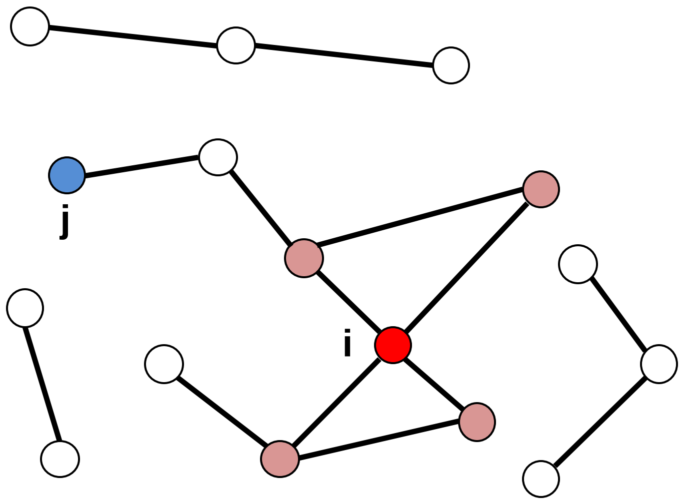
$$p_x = \frac{1}{N}\sum_i \delta_{x,x_i}$$

## Onde $p_x$ :

$$\sum_i p_x = 1 \left( \int p_x \, dx = 1 \right)$$

# GRAU MÉDIO

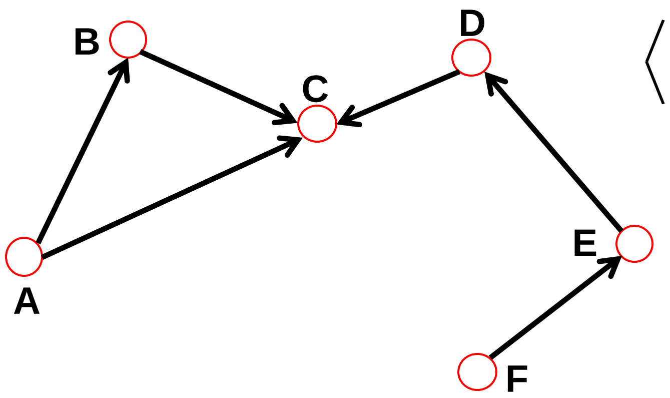
**Não direcionado**




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

N – número de nós no grafo

**Direcionado**




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

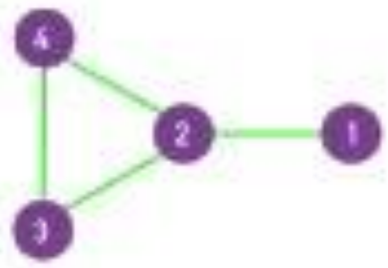
$$\langle k \rangle \text{ o } \frac{L}{N}$$

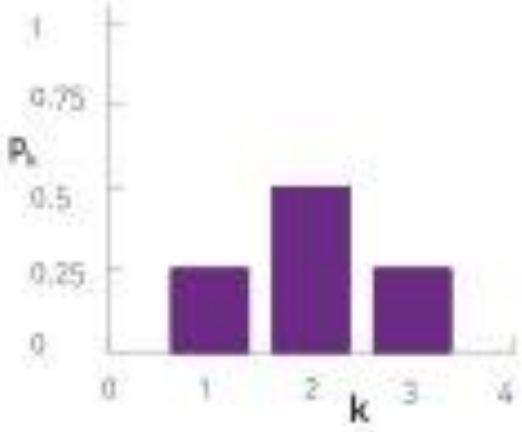
# GRAU MÉDIO

| 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 | 2.90 |

# Distribuição de grau

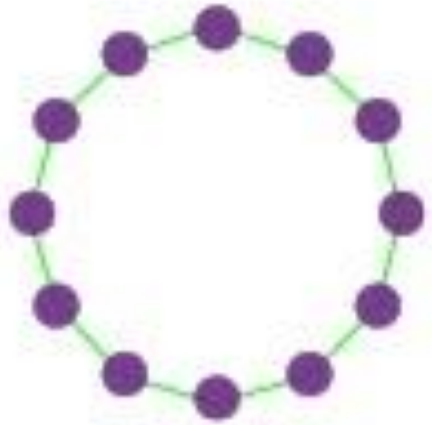
P(k): probabilidade de um
Nó escolhido aleatóriamente
tenha grau *k*

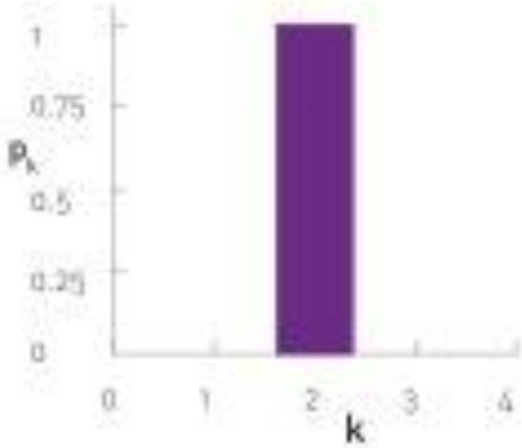



$N_k$ = número de nós com grau k

$P(k) = N_k / N$



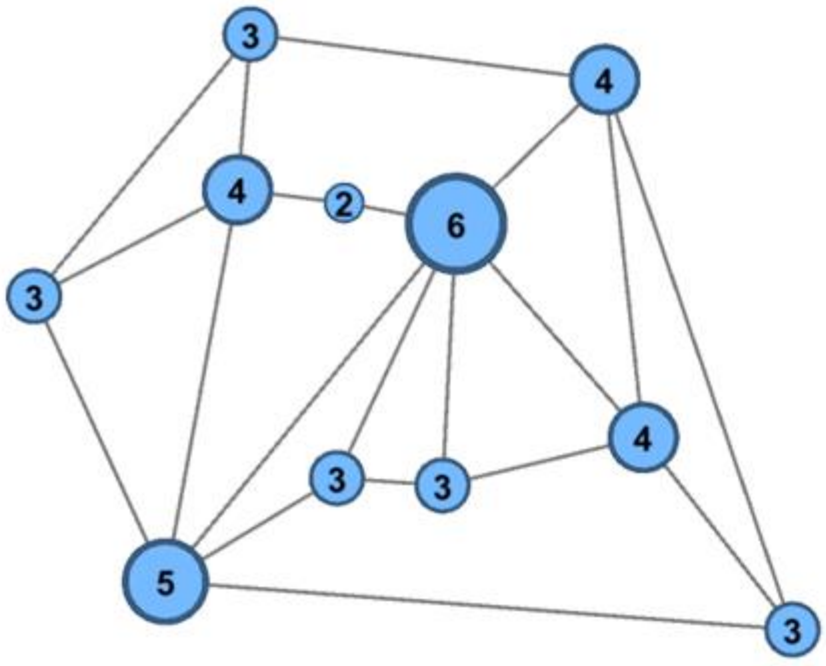


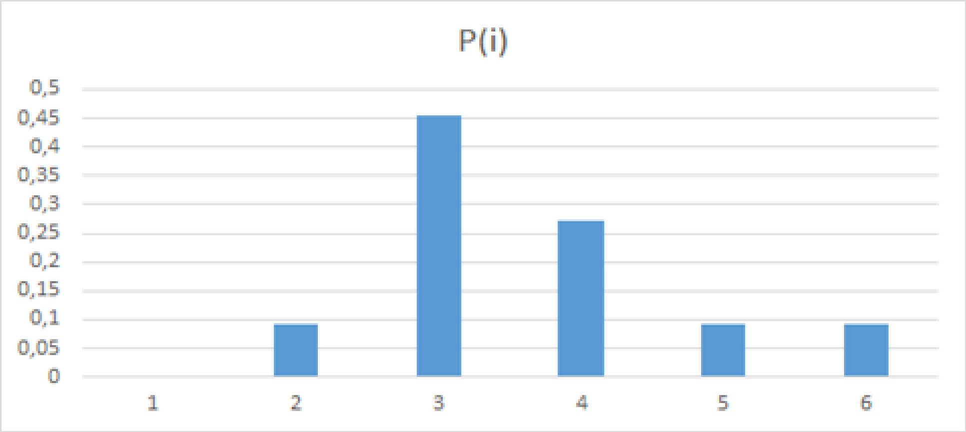
# DISTRIBUIÇÃO DE GRAU

A **distribuição de grau** de uma rede é a distribuição de probabilidade:

$$P(k) = \frac{|\{\, i \mid \deg(i) = k \}|}{n},$$

Que denota a probabilidade de um nó ter um grau $k$.



# DISTRIBUIÇÃO DE GRAU

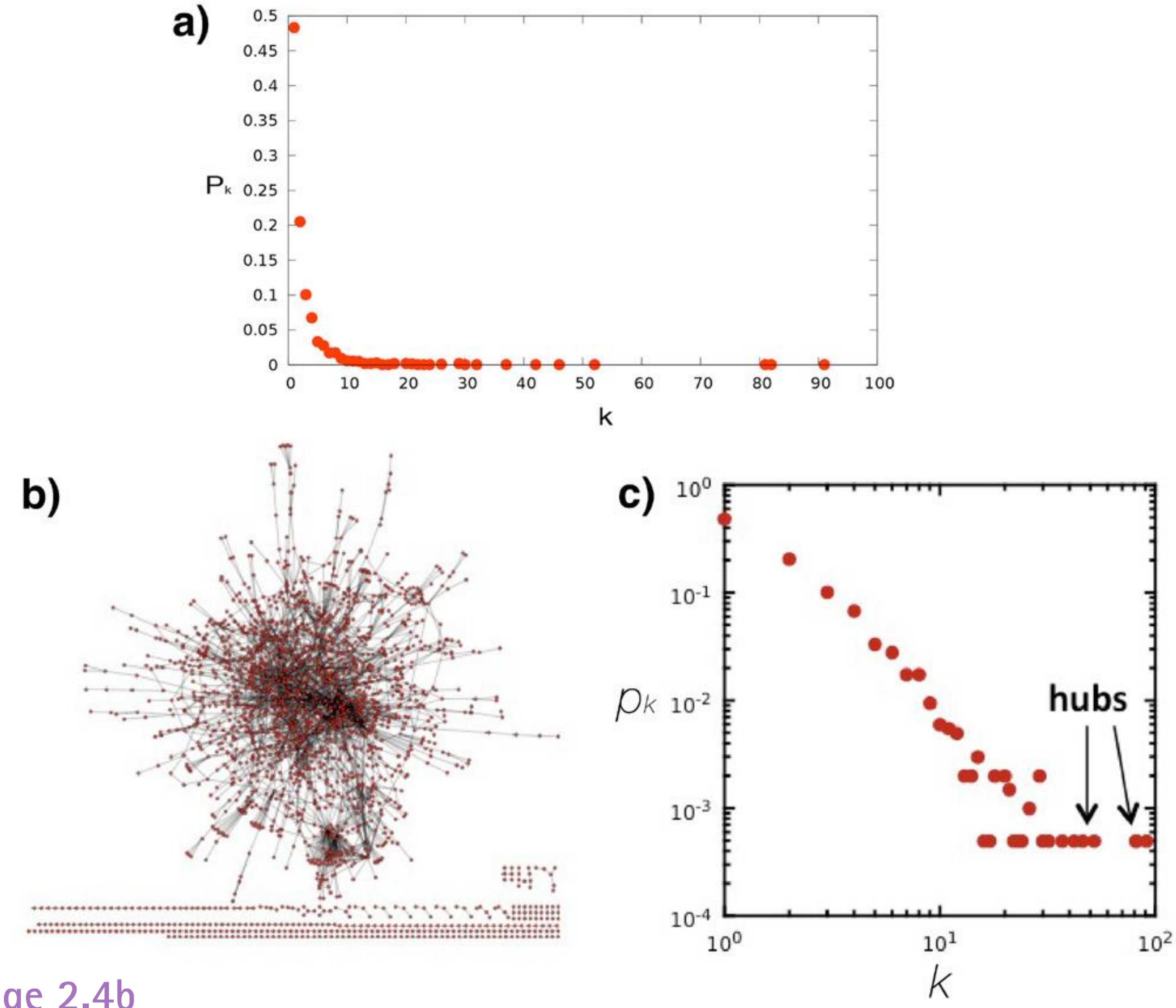



Image 2.4b

# DISTRIBUIÇÃO DE GRAU

**Representação discreta**: $p_k$ é a probabilidade que o nó tem o grau **k**

**Representação continua**: **p(k)** é a fdp do grau, onde

$$\int_{k_1}^{k_2} p(k)\,dk$$

representa a probabilidade que o grau do nó está entre $k_1$ e $k_2$.

**Condição de normalização:**

$$\sum_{0}^{\infty} p_k = 1 \qquad\qquad \int_{K_{\min}}^{\infty} p(k)\,dk = 1$$

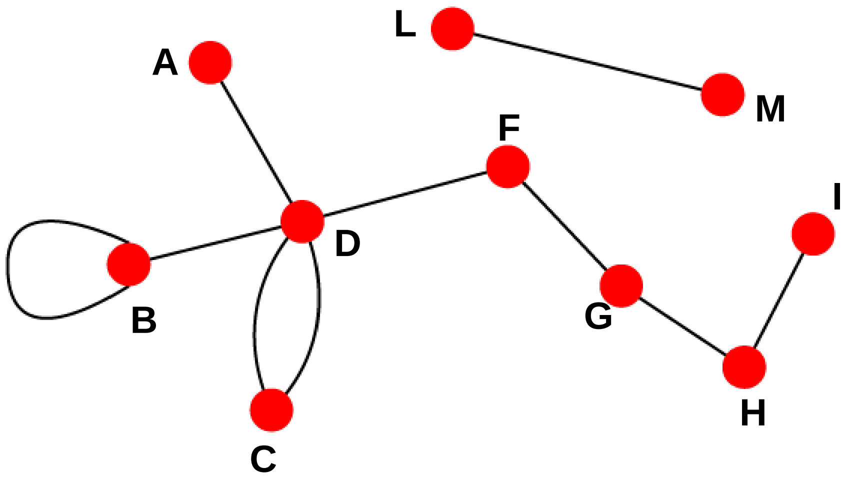
onde $K_{min}$ é o grau mínimo da rede/grafo.

# REDES DIRECIONADAS E NÃO DIRECIONADAS

## NÃO DIRECIONADAS

Ligações: não direcionadas (*simétricas*)

Grafo:

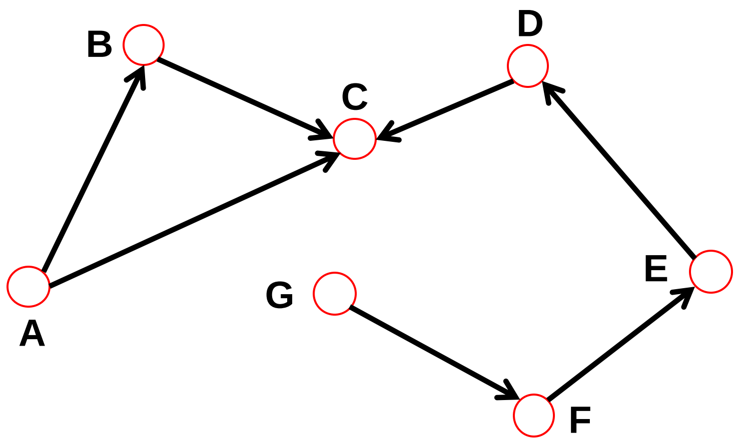



**Casos:**
Ligações de co-autoria
Redes de atores
Interações de proteinas

## DIRECIONADAS

Ligações: direcionadas (*arcos*).

Digrafo = grafo direcionado:




*Uma ligação não direcionada é a sobreposição de duas ligações direcionadas opostas.*

**Casos :**
URLs na www
Chamadas telefónicas
Relações metabólicas

# REDES DE REFERÊNCIA

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

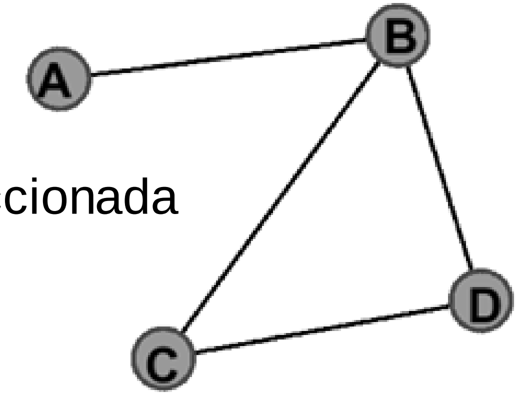
# Matriz de Adjacência

# Matriz de Adjacência

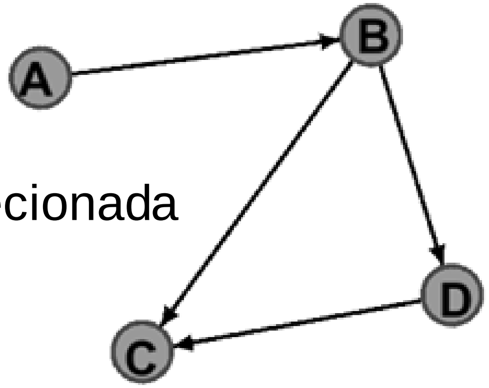
| | A | B | C | D |
|---|---|---|---|---|
| A | 0 | 1 | 0 | 0 |
| B | 1 | 0 | 1 | 1 |
| C | 0 | 1 | 0 | 1 |
| D | 0 | 1 | 1 | 0 |

| | A | B | C | D |
|---|---|---|---|---|
| A | 0 | 1 | 0 | 0 |
| B | 0 | 0 | 1 | 1 |
| C | 0 | 0 | 0 | 0 |
| D | 0 | 0 | 1 | 0 |

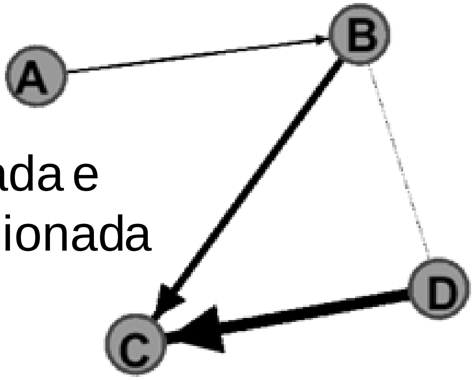
| | A | B | C | D |
|---|---|---|---|---|
| A | 0 | 0.8 | 0 | 0 |
| B | 0 | 0 | 0.9 | 0.75 |
| C | 0 | 0 | 0 | 0 |
| D | 0 | 0 | 1.0 | 0 |




Não direccionada

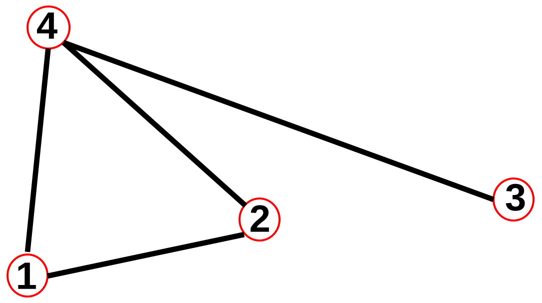


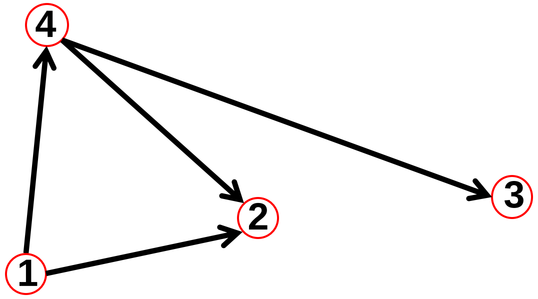

direcionada




Pesada e direcionada

# ADJACENCY MATRIX




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

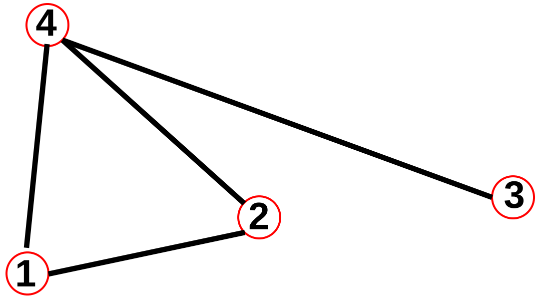
$A_{ij}=1$ se há uma ligação entre os nós *i* e *j*

$A_{ij}=0$ se os nós *i* e *j* não estão conectados entre si.

Notar que para um grafo direcionado (à direita) a matriz não é simétrica.

# MATRIZ DE ADJACÊNCIA E GRAU DE UM NÓ

**Não direcionado**




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

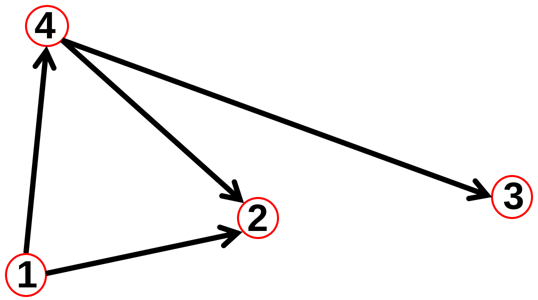
$$A_{ij} = A_{ji}$$
$$A_{ii} = 0$$

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

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

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

**Direcionado**




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

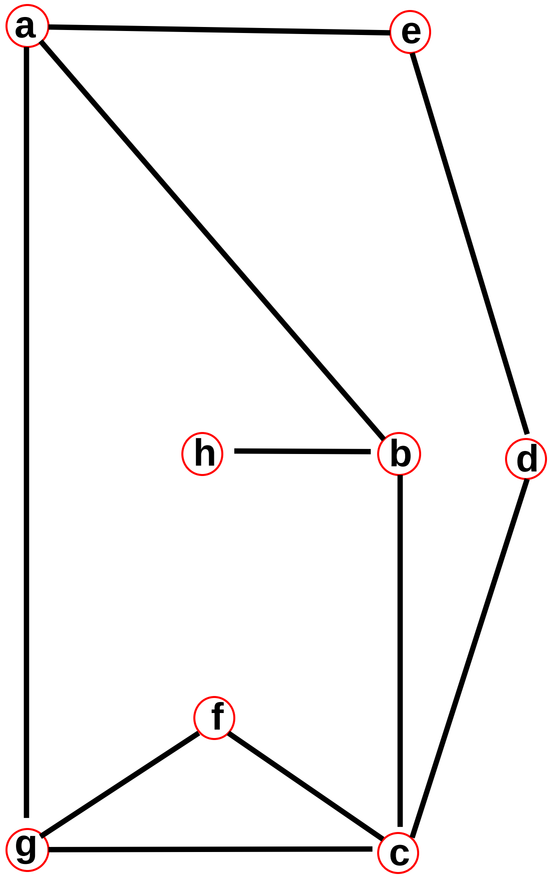
$$A_{ij} \neq A_{ji}$$
$$A_{ii} = 0$$

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

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

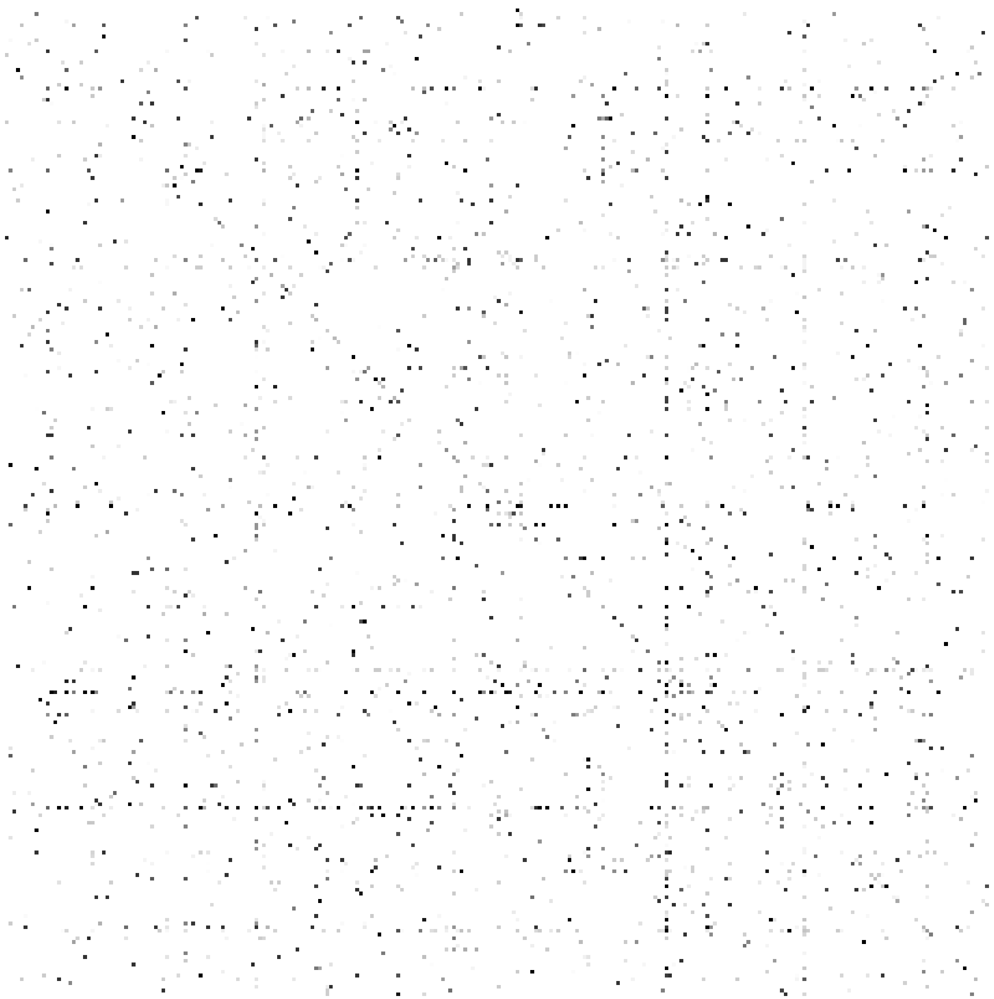
$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{j=1}^{N} k_j^{out} = \sum_{i,j}^{N} A_{ij}$$

# A MATRIZ DE ADJACÊNCIA É NORMALMENTE ESPARSA




| | 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 |

# A MATRIZ DE ADJACÊNCIA É NORMALMENTE ESPARSA



# AS REDES REAIS SÃO ESPARSAS

**A maioria das redes observadas nos sistemas reais são esparsas**

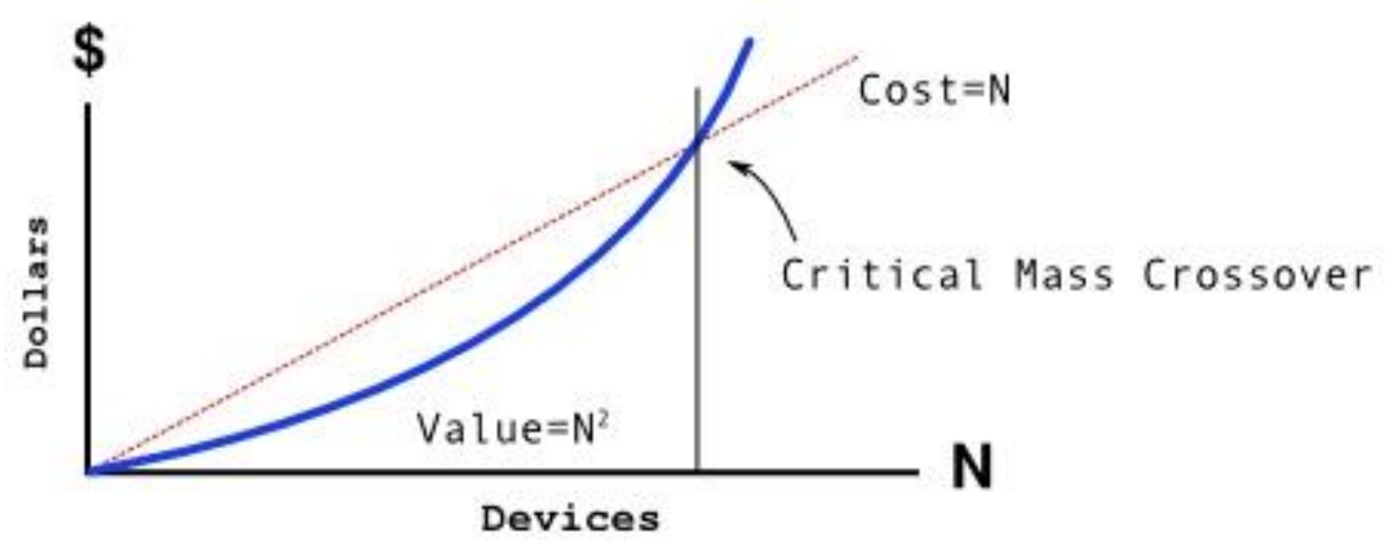
$$L \ll L_{max} \quad \text{or} \quad \langle k \rangle \ll N-1$$

| | | | | |
|---|---|---|---|---|
| WWW(amostra ND) | N=325,729 | $L=1.4\ 10^{6}$ | $L_{max}=10^{12}$ | $\langle k \rangle=4.51$ |
| Proteinas(S. Cerevisiae) | N=1,870 | L=4,470 | $L_{max}=10^{7}$ | $\langle k \rangle=2.39$ |
| Coautoria(Math) | N=70,975 | $L=2\ 10^{5}$ | $L_{max}=3\ 10^{10}$ | $\langle k \rangle=3.9$ |
| Actores | N=212,250 | $L=6\ 10^{6}$ | $L_{max}=1.8\ 10^{13}$ | $\langle k \rangle=28.78$ |

*(Fonte: Albert, Barabasi, RMP2002)*

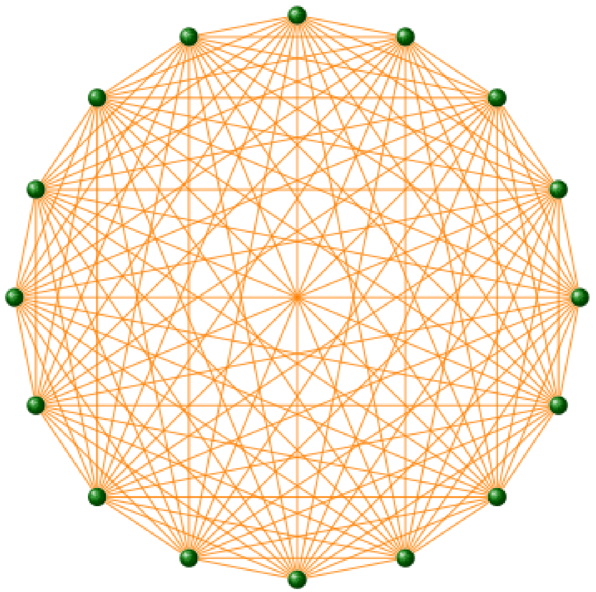
# LEI DE METCALFE

*O valor de um sistema de comunicações cresce na razão do quadrado do número de utilizadores do sistema*




O número máximo de linques é dado por:

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

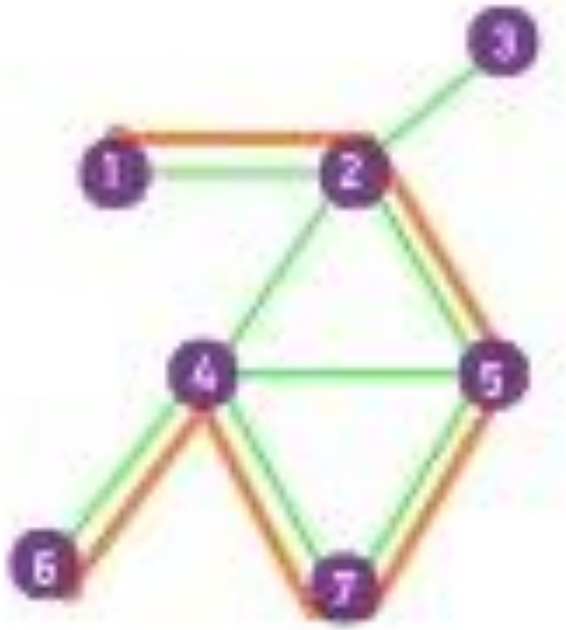


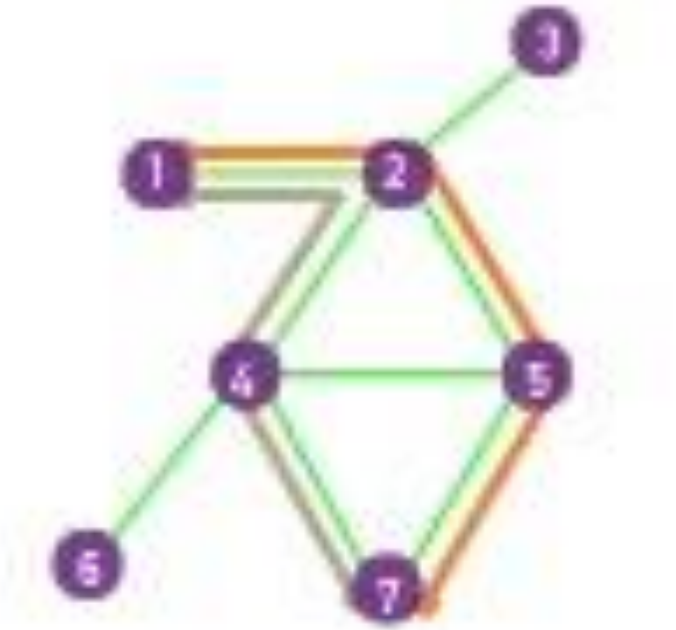
# CAMINHOS

Um *caminho* é uma sequência de nós em que cada nó é adjacente de um outro

$P_{i0,in}$ de tamanho *n* entre os nós $i_0$ e $i_n$ é uma coleção ordenada de *n+1* nós e *n* linques

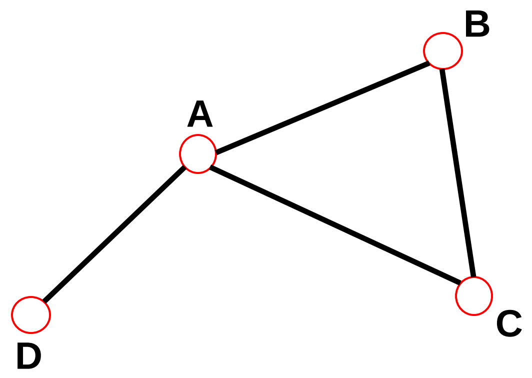
$$P_n = \{i_0, i_1, i_2, \ldots, i_n\} \qquad P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{n-1}, i_n)\}$$




- Numa rede direccionada um cominho só pode tomar uma direção.

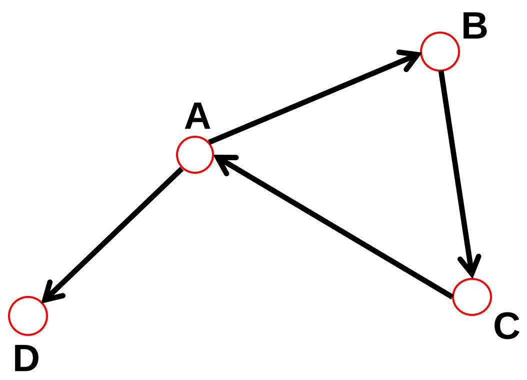
# DISTÂNCIA NUM GRAFO




A distância $d$ (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.

$$d(i \to j)$$




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

Assim, em um dígrafo, a distância do nó A ao B (em um caminho AB) é geralmente diferente da distância do nó B a A (em um caminho BCA).

# NÚMERO DE CAMINHOS ENTRE DOIS NÓS

## $N_{ij}$, número de caminhos entre quaisquer dois nós *i* e *j*:

***Tamanho n=1*:** se há uma ligação entre *i* e *j*, então $A_{ij}=1$ e $A_{ij}=0$ senão.

***Tamanho n=2:*** se há um caminho de tamanho dois entre *i* e *j*, então $A_{ik}A_{kj}=1$, e $A_{ik}A_{kj}=0$ senão.

Número de caminhos de tamanho 2:

$$N_{ij}^{(2)} = \sum_{k=1}^{N} A_{ik}A_{kj} = [A^2]_{ij}$$

***Tamanho n:*** em geral, e há um caminho de tamanho *n* entre *i* e *j*, então $A_{ik}\ldots A_{lj}=1$ e $A_{ik}\ldots A_{lj}=0$ senão.

O número de caminhos de tamanho *n* entre *i* e *j* é*

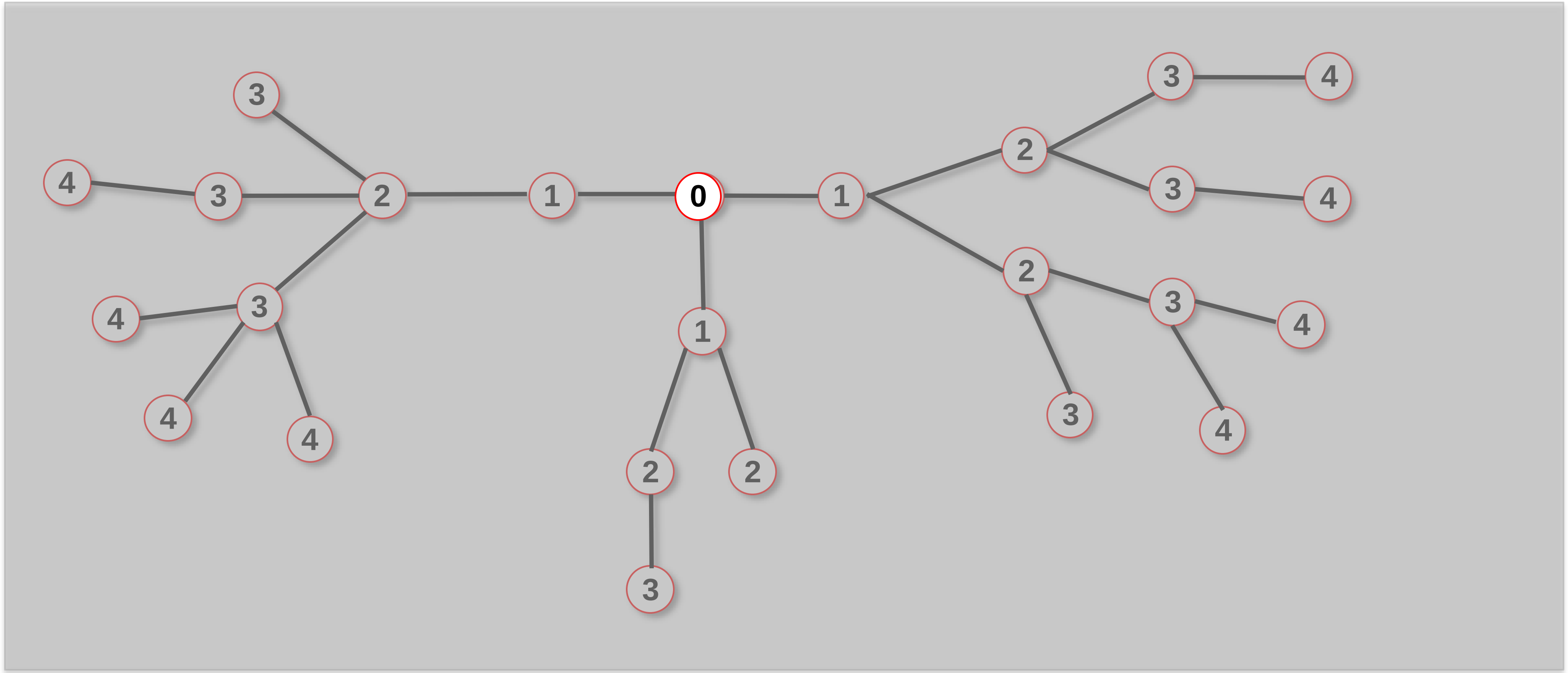
$$N_{ij}^{(n)} = [A^n]_{ij}$$

**não só nas redes direcionadas como nas não direcionadas.*

# CALCULAR DISTÂNCIAS : PESQUISA *BREATH FIRST*

**Distância entre nó 0 e nó 4:**

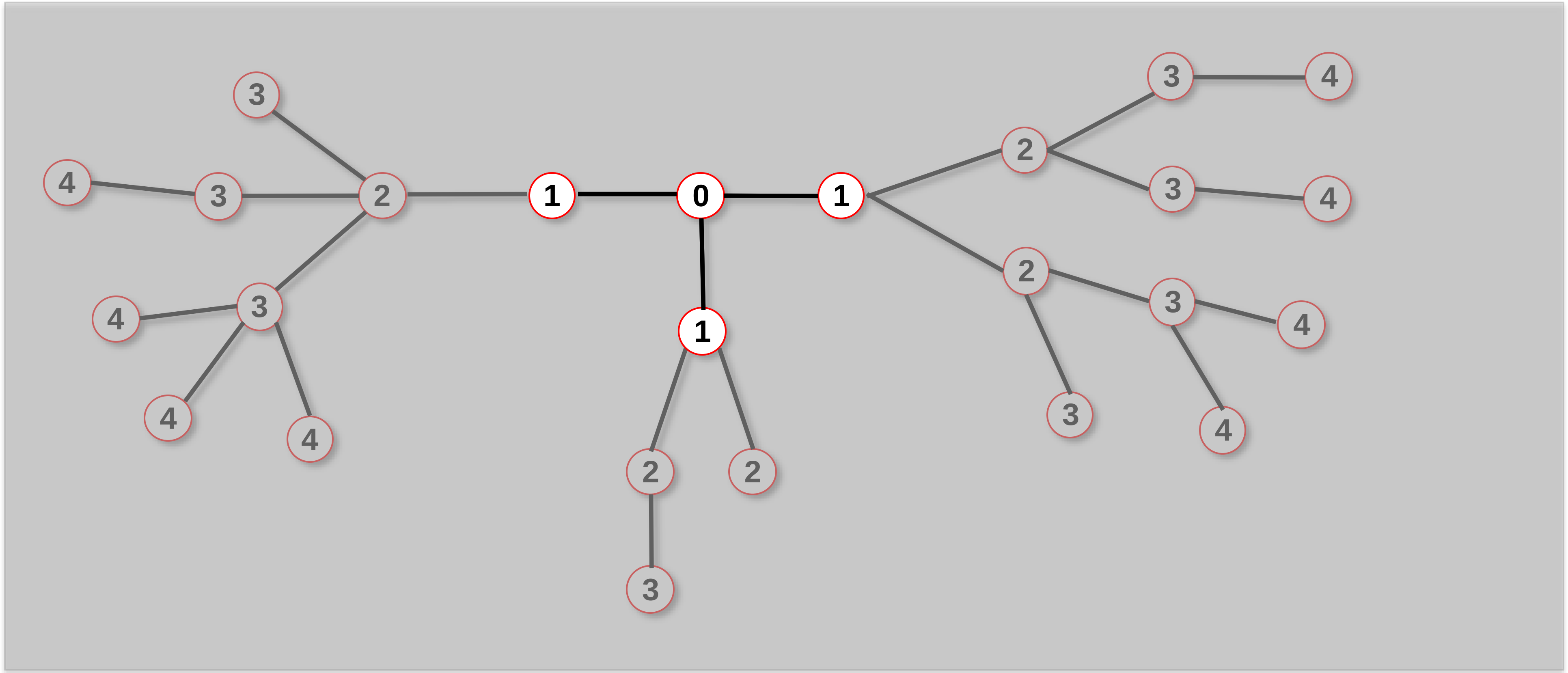
1. Começar em 0.



# CALCULAR DISTÂNCIAS : PESQUISA *BREATH FIRST*

**Distância entre nó 0 e nó 4:**

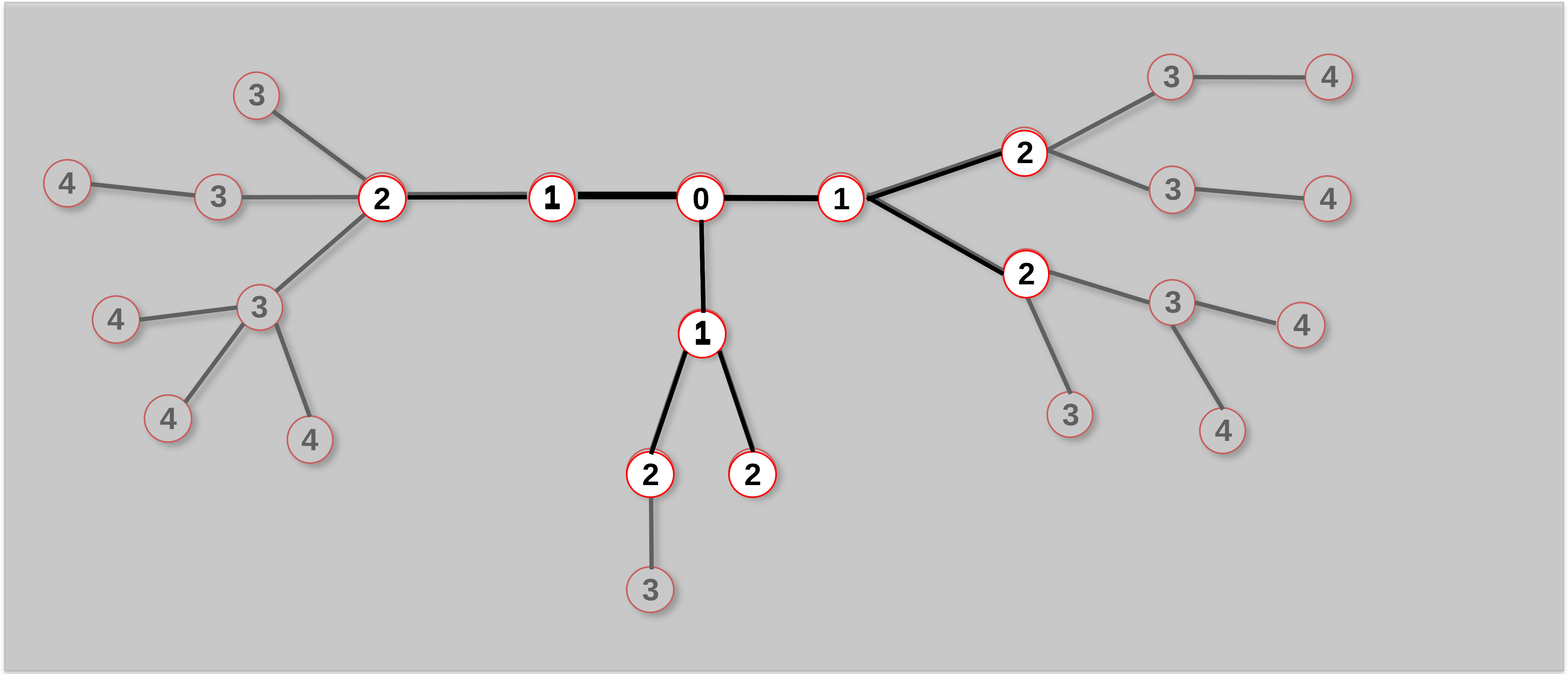
1. Começar em 0.
2. Encontrar os nós adjacentes. Marcar com distância *1*. Colocá-los numa fila.



# CALCULAR DISTÂNCIAS : PESQUISA *BREATH FIRST*

**Distância entre nó 0 e nó 4:**

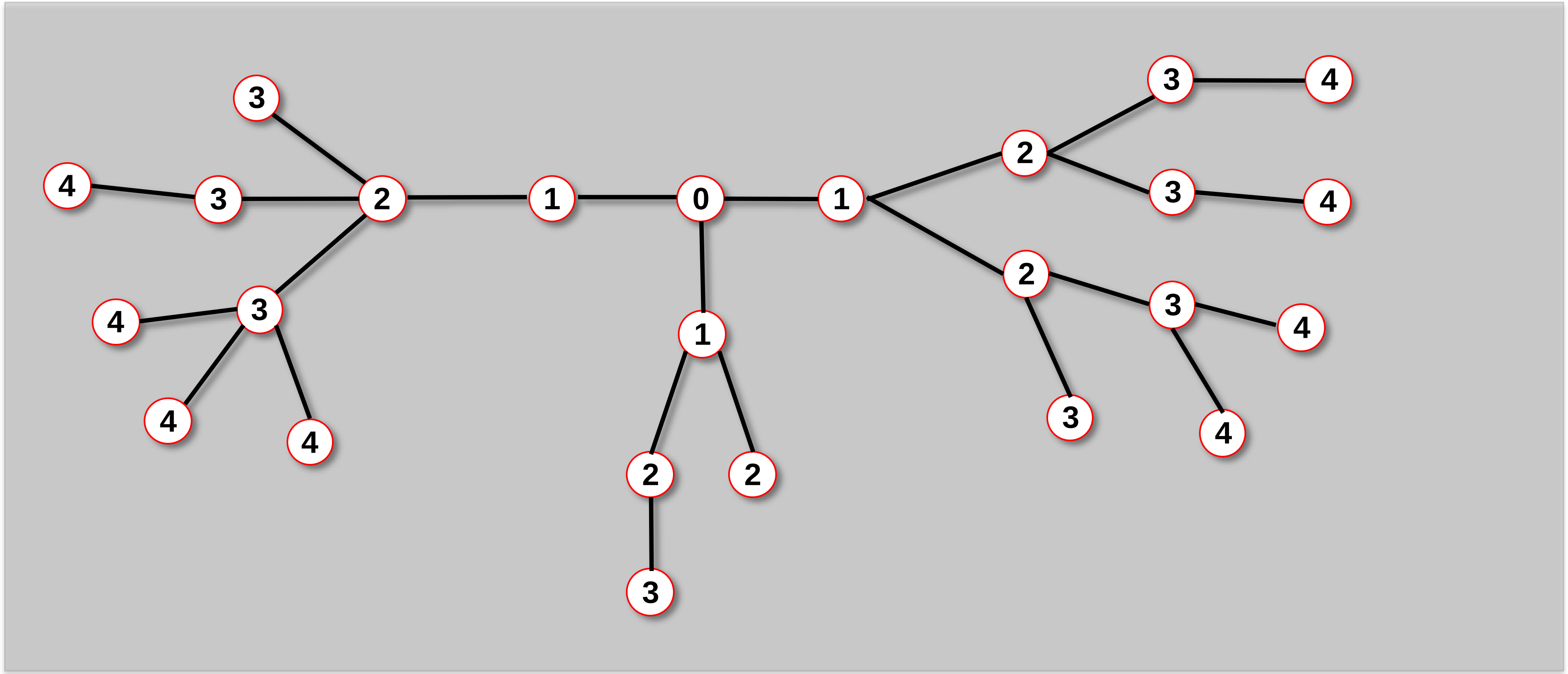
1.Começar em 0.
2.Encontrar os nós adjacentes. Marcar com distância *1*. Colocá-los numa fila.
3.Tirar o primeiro nó da fila. Encontrar os nós não marcados adjacentes a ele no grafo. Marcar esses nós com rótulo *2*. Colocá-los na fila.



# CALCULAR DISTÂNCIAS : PESQUISA *BREATH FIRST*

**Distância entre nó 0 e nó 4:**

1. Repetir até encontrar o nó *4* ou não haver mais nós na fila.
2. A distância entre *0* e *4* é o rótulo de *4* ou, se *4* não tiver rótulo, infinito.



# DIÂMETRO DA REDE E DISTÂNCIA MÉDIA

*Diâmetro*: $d_{max}$ a distância máxima entre qualquer par de nós no grafo.

*Raio*: $d_{min}$ a distância mínima entre qualquer par de nós no grafo.

*Excentricidade*: ***e(i)*** é a distância máxima a um dos outros nós na rede

*Distância média*, **<d>**, para um grafo conectado:

(também chamada distância característica)

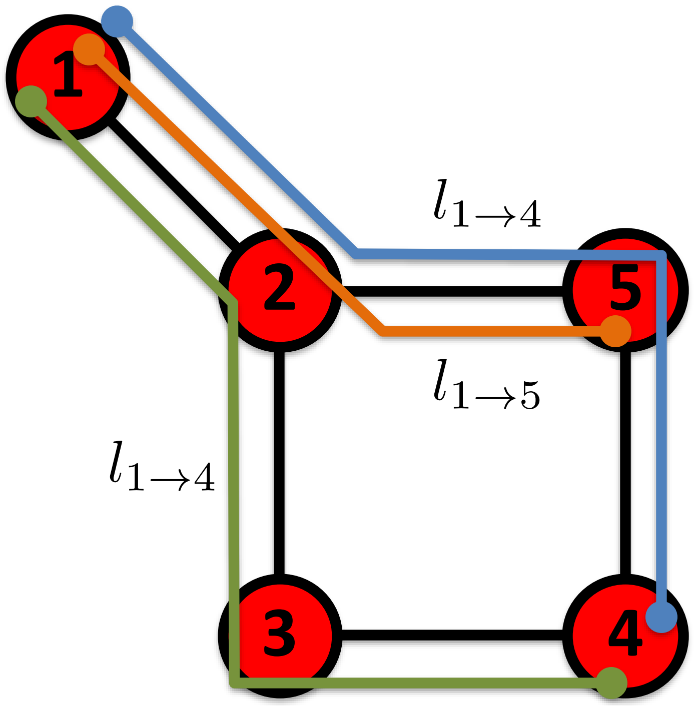
$$\langle d \rangle \equiv \frac{1}{2L_{\max}} \sum_{i,j \neq i} d_{ij}$$

Num grafo não direcionado $d_{ij} = d_{ji}$, só temos que contar uma vez:

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

## CAMINHOS: sumário

Caminho mais curto



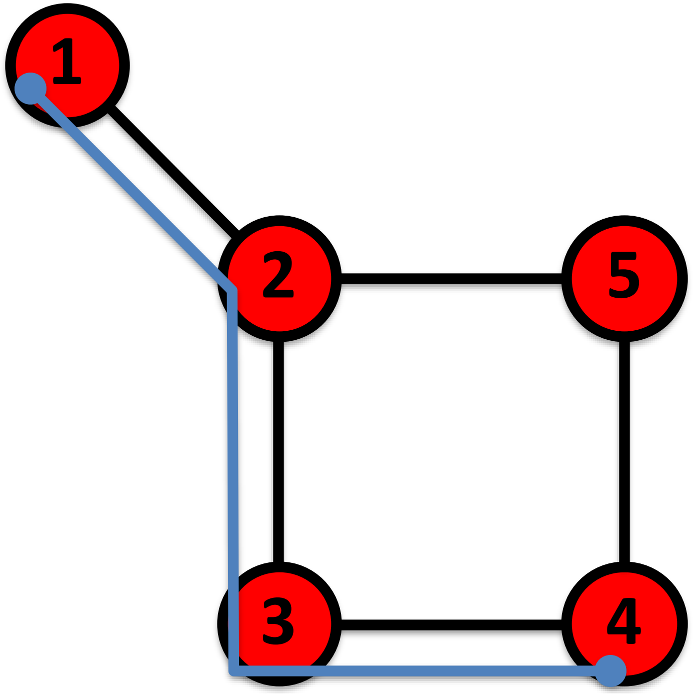

$$l_{1\to4} = 3$$

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

O caminho com a menor distância entre os nós

# CAMINHOS: sumário

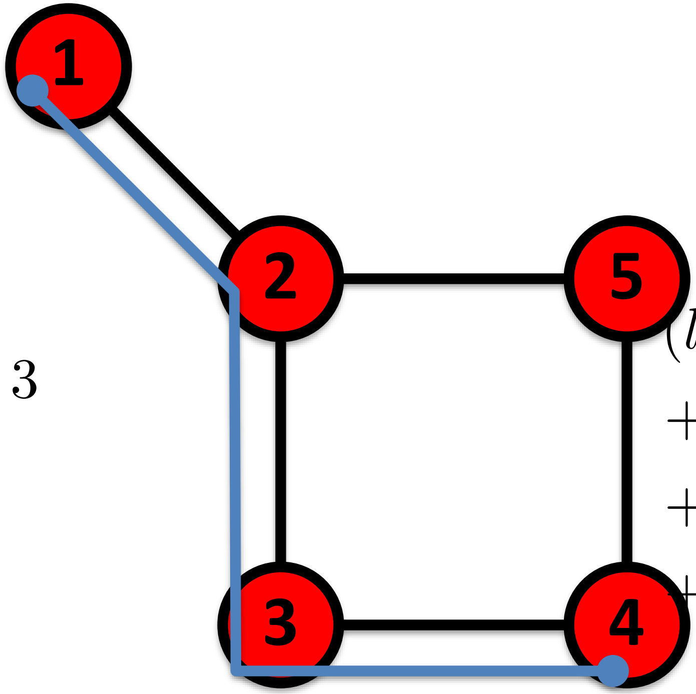
## Diâmetro




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

O maior caminho mais curto

## Tamanho médio dos caminhos

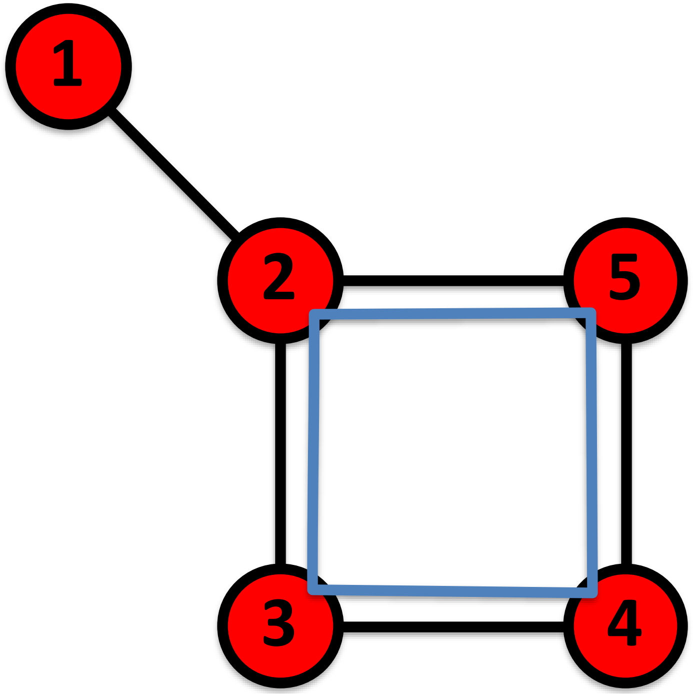



$$\begin{aligned}(l_{1\to 2} + l_{1\to 3} + l_{1\to 4} +\\ + l_{1\to 5} + l_{2\to 3} + l_{2\to 4} +\\ + l_{2\to 5} + l_{3\to 4} + l_{3\to 5} +\\ + l_{4\to 5})/10 = 1.6\end{aligned}$$

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

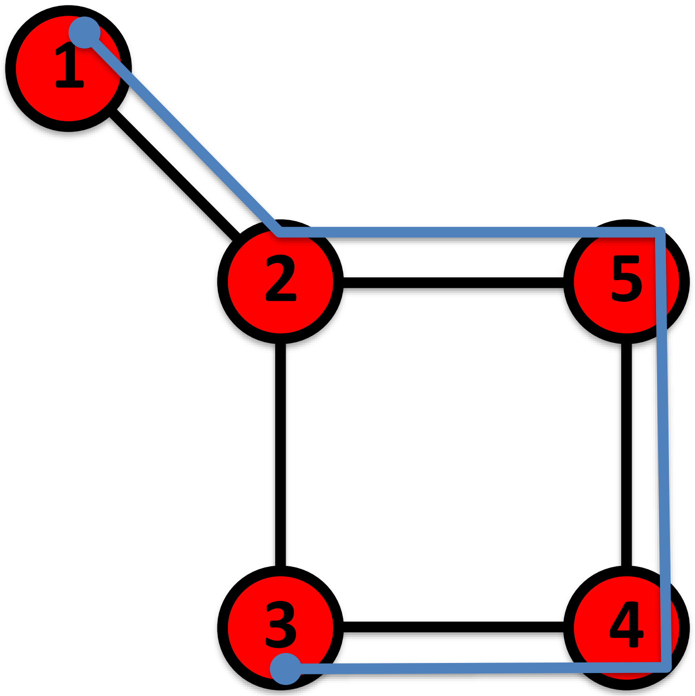
# CAMINHOS: sumário

## Ciclo




Um caminho com o mesmo nó de partida e chegada

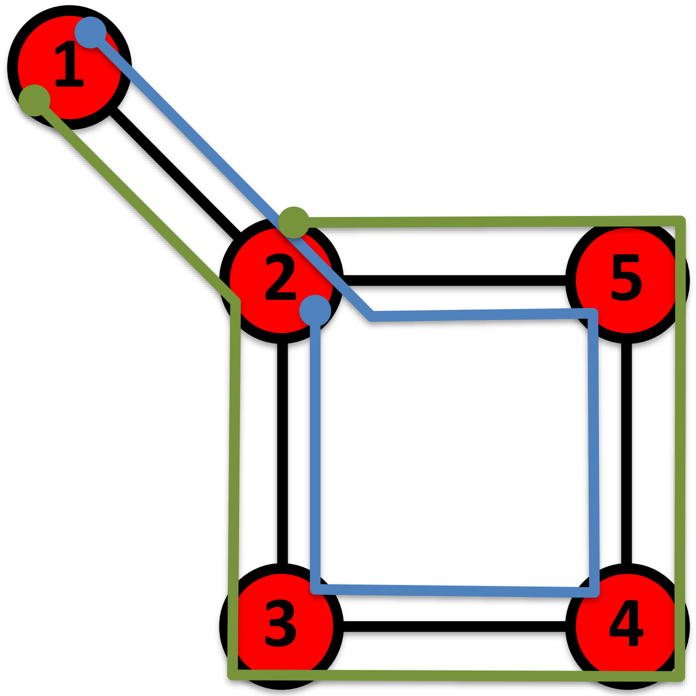
## Caminho auto-exclusivo




Um caminho que não se intersecta a si próprio.

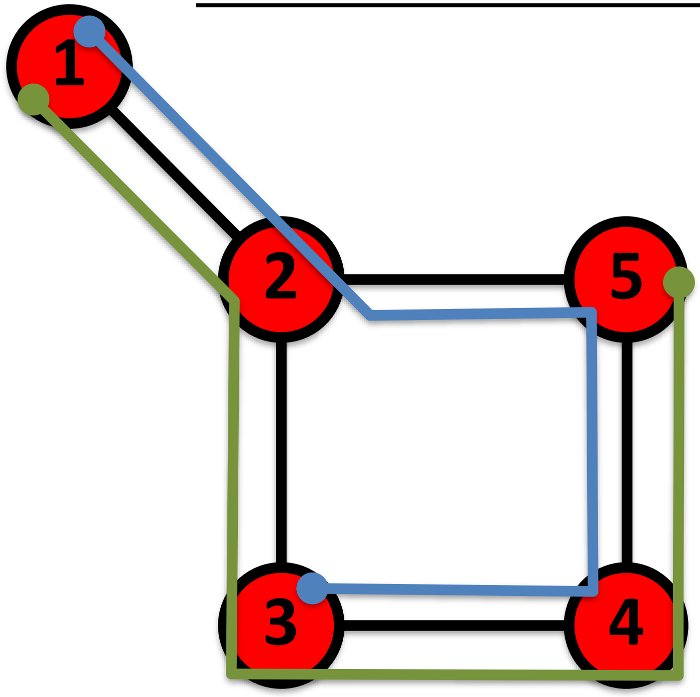
# CAMINHOS: sumário

## Caminho Euleriano




Caminho que atravessa cada ligação apenas uma vez

## Caminho Hamiltoniano

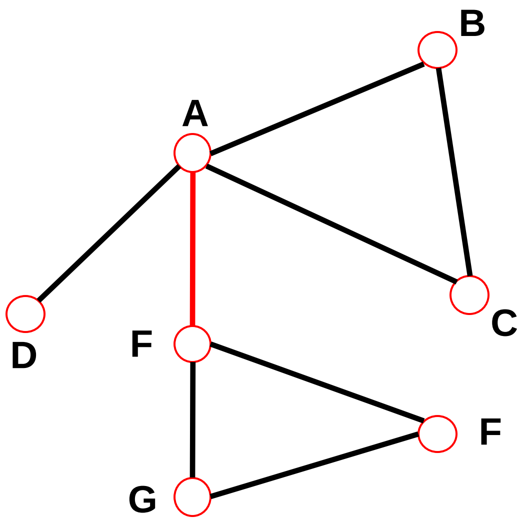


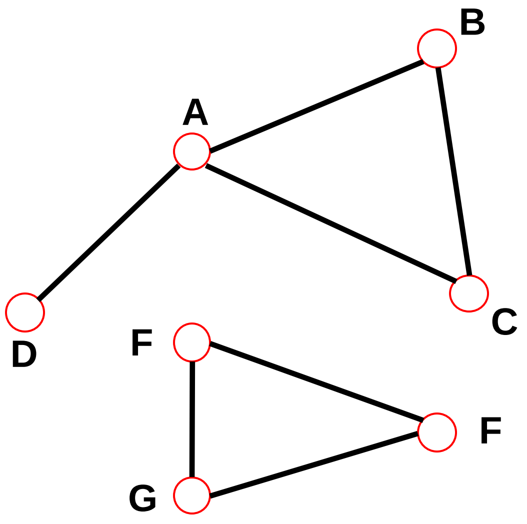

Caminho que visita todos os nós apenas uma vez.

# CONETIVIDADE

# CONETIVIDADE DE GRAFOS NÃO DIRECIONADOS

Grafo conetado (não direcionado) : quaisquer dois nós podem ser unidos por um caminho.
Um grafo desconetado contém dois ou mais grafos conetados.




Maior componente:

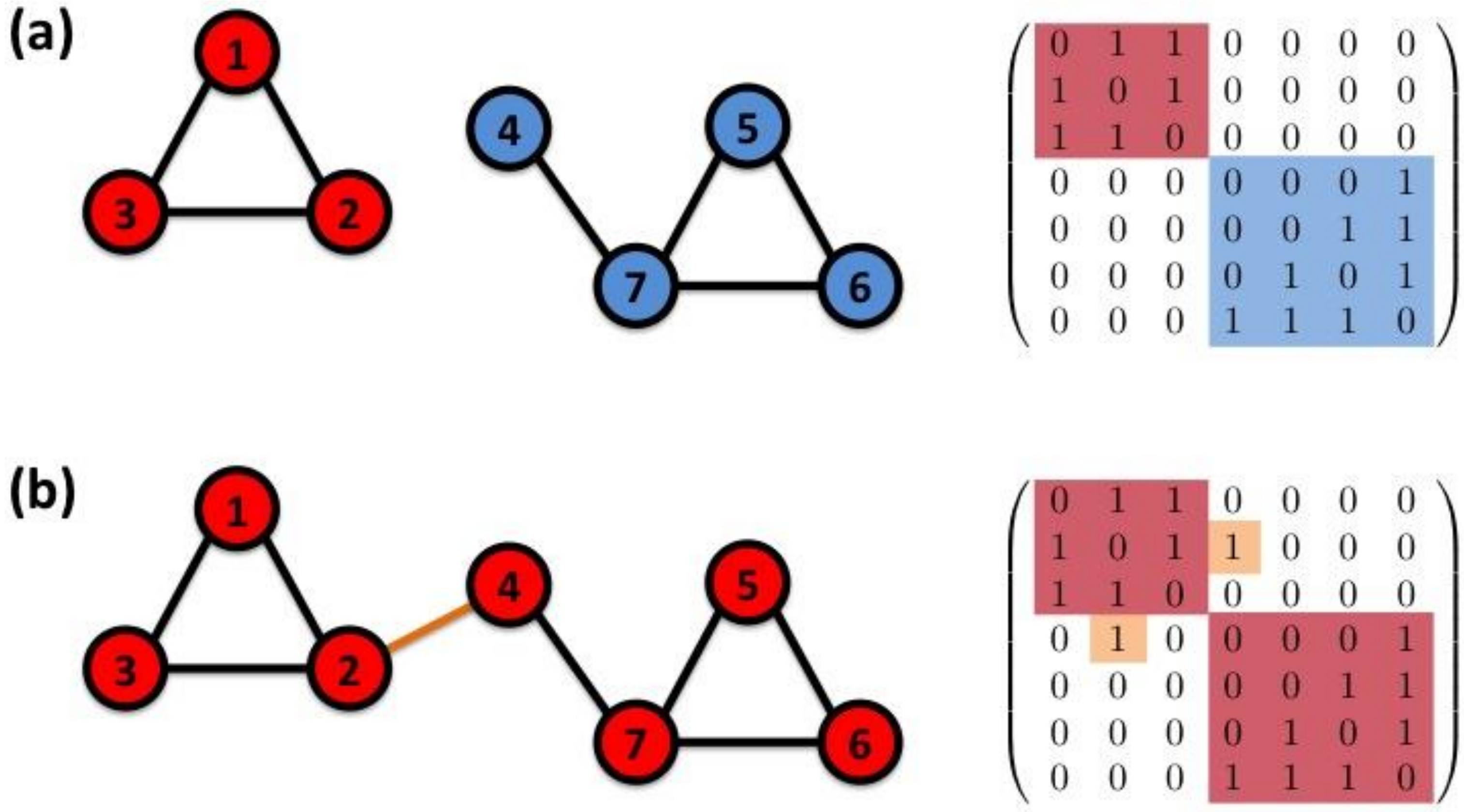
**Component Gigante**

O resto: **Isolados**

Ponte: se a removemos o grafo fica desconetado.

# CONETIVIDADE DE GRAFOS NÃO DIRECIONADOS

A matriz de adjacência de uma rede com vários componentes pode ser escrita numa forma de blocos-diagonais, tal que os elementos não nulos estão confinados a quadrados, sendo todos os outros nulos:

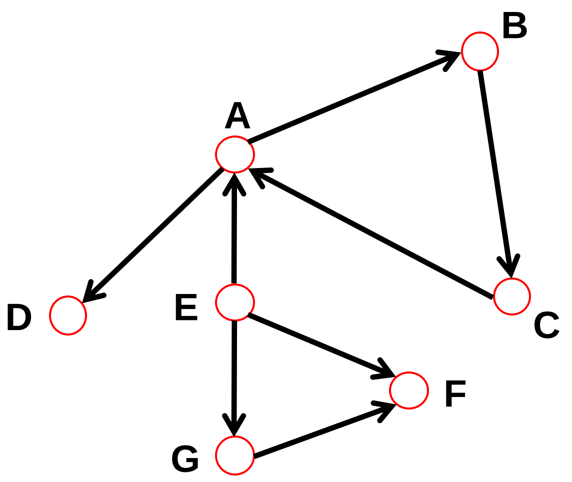


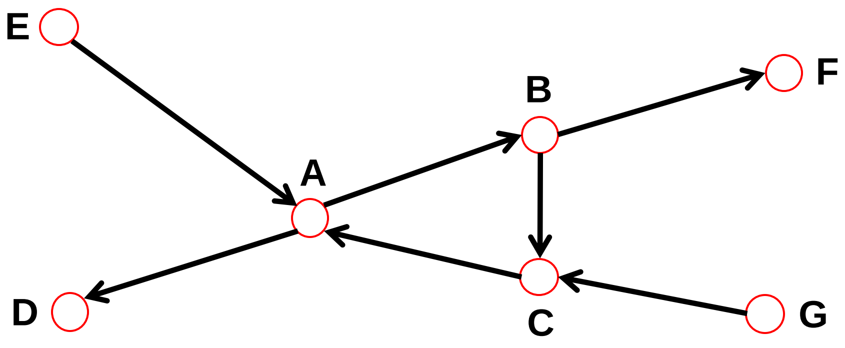
# CONNECTIVIDADE DE GRAFOS DIRECIONADOS

**Grafo direcionado fortemente conetado** : tem um caminho de um nó para qualquer outro nó e vice versa (e.x. caminho AB e caminho BA).

**Grafo direcionado fracamente conetado**: é o grafo conectado não considerando as direções

Os componentes fortemente conetados podem ser identificados, embora nem todos os nós possam fazer parte de um componente conectado não trivial.




**Componente-in**: nós que conseguem chegar à fonte.

**Componente-out**: nós a que se consegue chegar a partir da fonte.

## ENCONTRAR OS COMPONENTES CONETADOS DE UMA REDE

1. Começar por um nó *i* escolhido ao acaso e efetuar uma pesquisa *Breath First*. Rotular todos os nós alcançados como 1.
2. Se o número total de nós rotulados for N, então a rede é conectada. Se o número for menor, a rede contém vários componentes.
3. Aumente o rótulo de *n* para *n+1*. Escolher um nó arbitrário não rotulado *j* e rotulá-lo n+1. Usar pesquisa *Breath First* para encontrar todos os nós alcançáveis a partir de *j*, rotulá-los todos com *n+1*. Retornar ao passo 2.

# Coeficiente de *clustering*

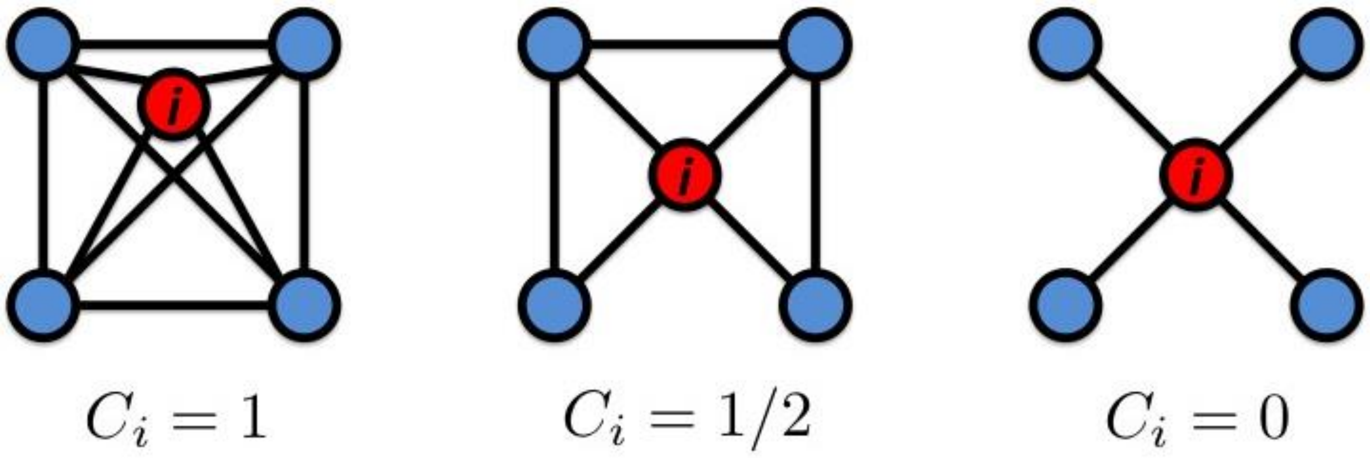
# COEFICIENTE DE *CLUSTERING*

## Coeficiente de *clustering*:

Que fração dos seus vizinhos está ligada?

- Nó i com grau $k_i$
- $C_i$ em [0,1]

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



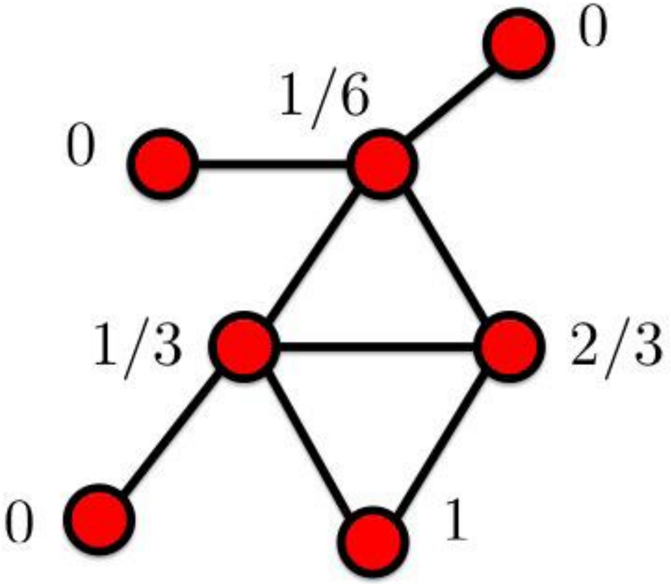

**Watts & Strogatz, Nature 1998.**

## Coeficiente de *clustering*:

Que fração dos seus vizinhos está ligada?

- Nó i com grau $k_i$
- $C_i$ em [0,1]

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



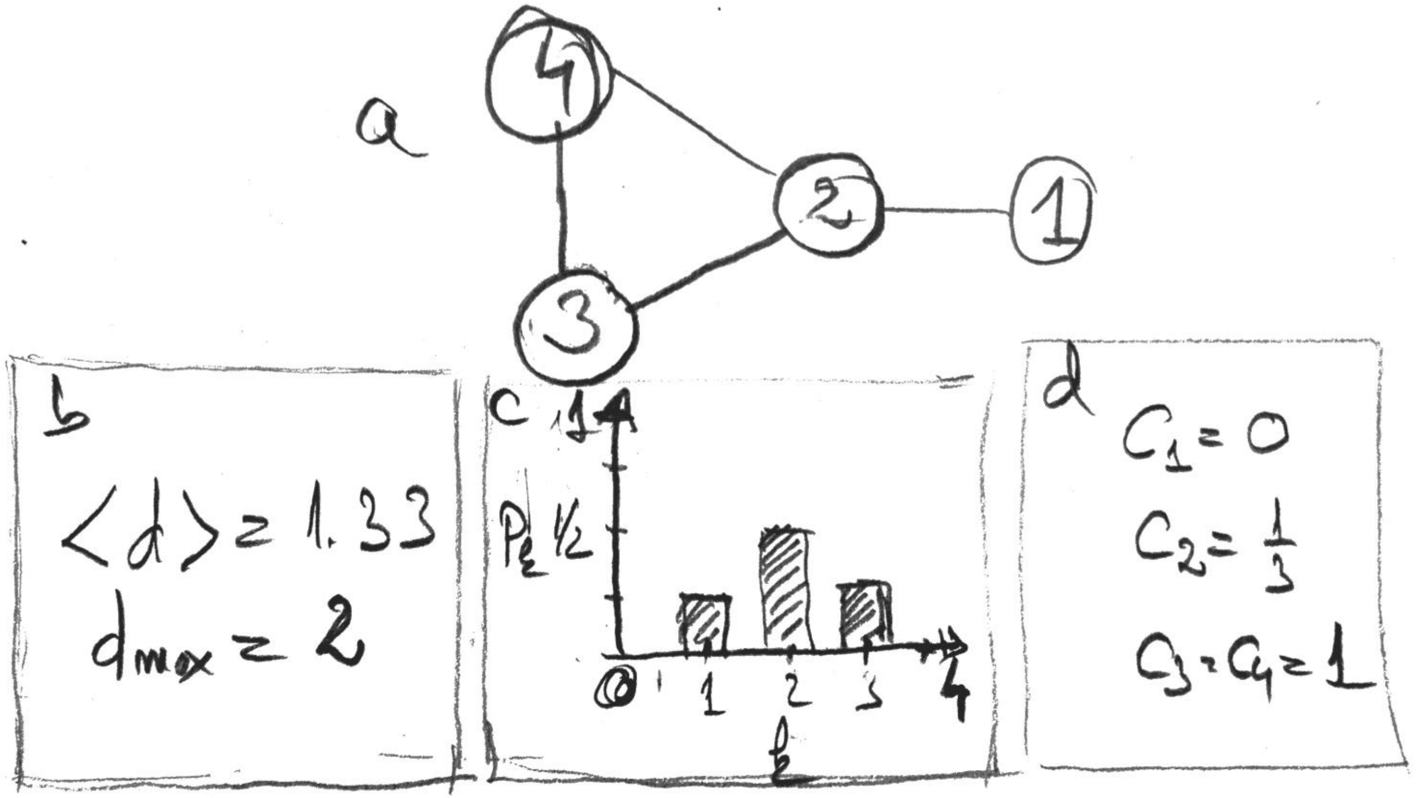

$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

$$C = \frac{3}{8} = 0.375$$

**Watts & Strogatz, Nature 1998.**

# SUMÁRIO

# TRÊS QUANTIDADES CENTRAIS NA CIÊNCIA DE REDES

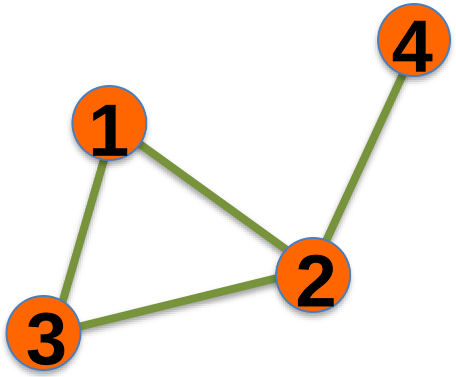



A. **Distribuição de grau:** $p_k$

B. **Distância média:** $\langle d \rangle$

C. **Coeficiente de *clustering*:** $C_i = \frac{2e_i}{k_i(k_i - 1)}$

## Não direcionado



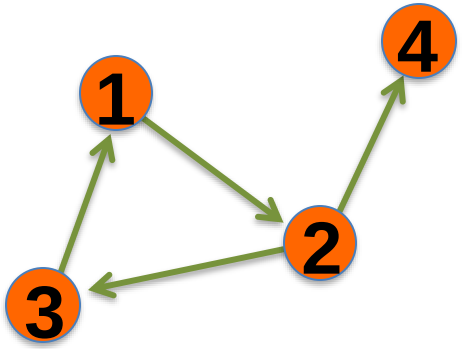

$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

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

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

*Actores, interação proteina-proteina*

## Direcionado




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

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

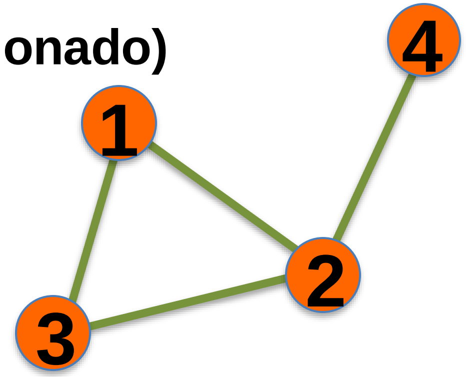
$$L = \sum_{i,j=1}^{N} A_{ij} \qquad \langle k \rangle = \frac{L}{N}$$

*WWW, redes de citações*

# GRAFOLOGIA

## Não ponderado

**(não direcionado)**




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

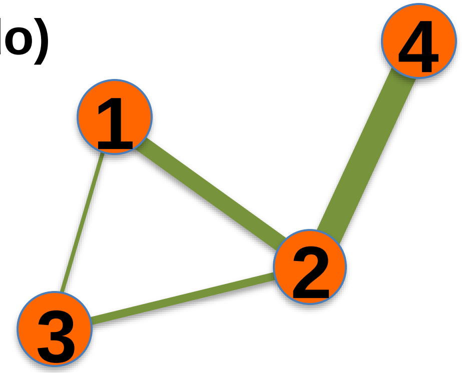
$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

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

*Interação proteina-proteina, www*

## Ponderado

**(direcionado)**




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

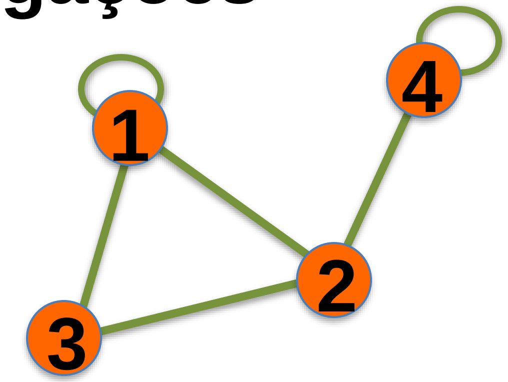
$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

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

*Grafo de chamadas, redes metabolicas*

# GRAFOLOGIA

## Auto-ligações




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

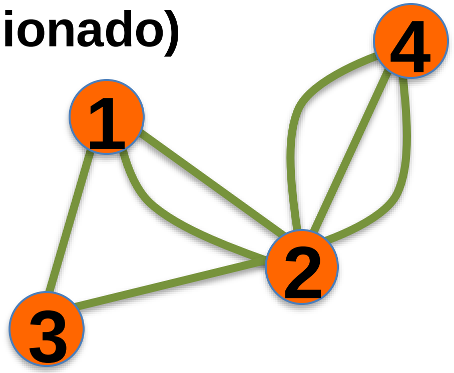
$$A_{ii} \neq 0 \qquad A_{ij} = A_{ji}$$

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

*Interação proteina-proteina, www*

## Multigrafo

**(não direcionado)**




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

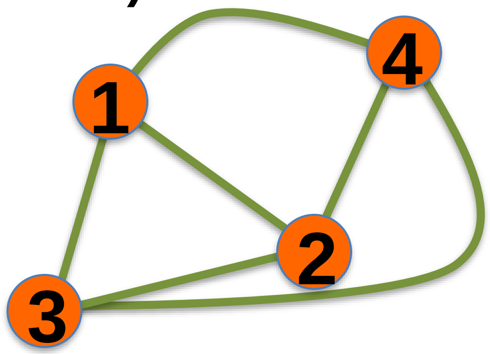
$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

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

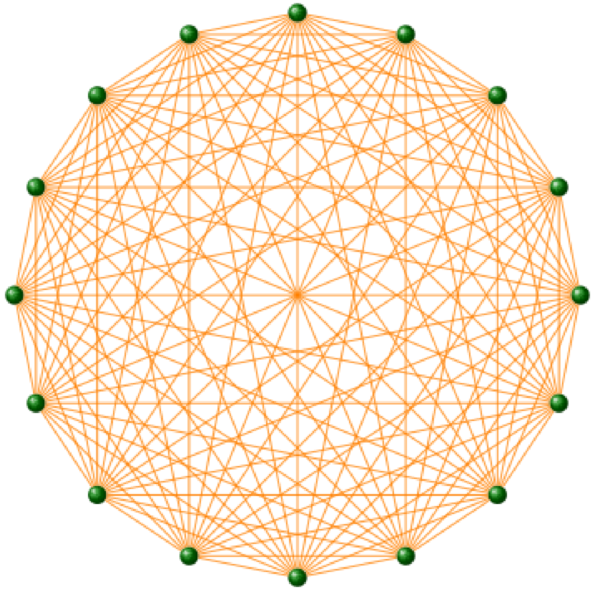
*Redes sociais, redes de colaboração*

# Grafo Completo

**(não direcionado)**




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



$$A_{ii} = 0 \qquad A_{i \neq j} = 1$$

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

*Actores, interação proteina-proteina*