Análise de Redes

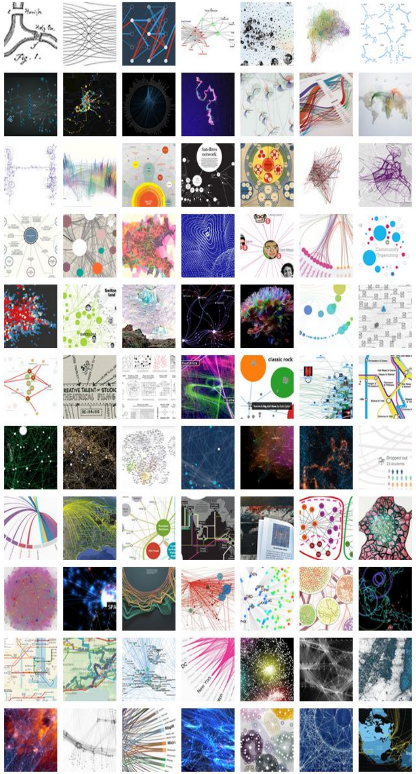
Aula 02 – Introdução a Grafos

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

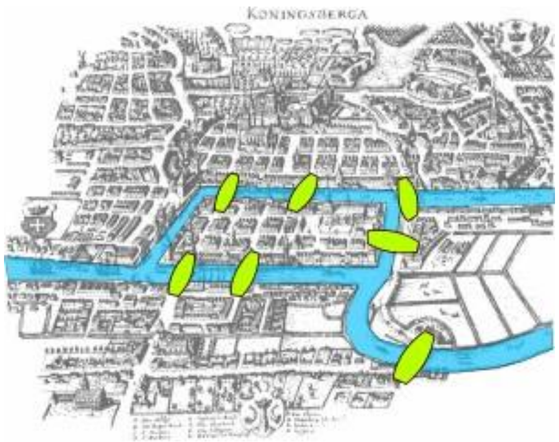


http://www.visualcomplexity.com/vc/

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.

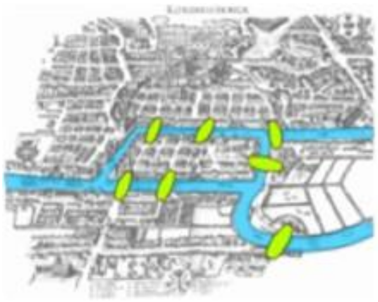
- **Problema**:
  *Há um **caminho** que passe por todas as pontes, visitando cada ponte uma única vez?*

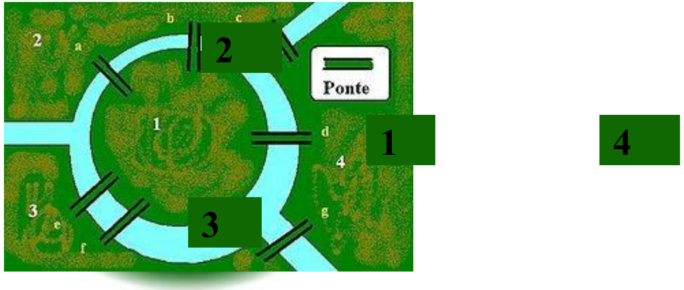


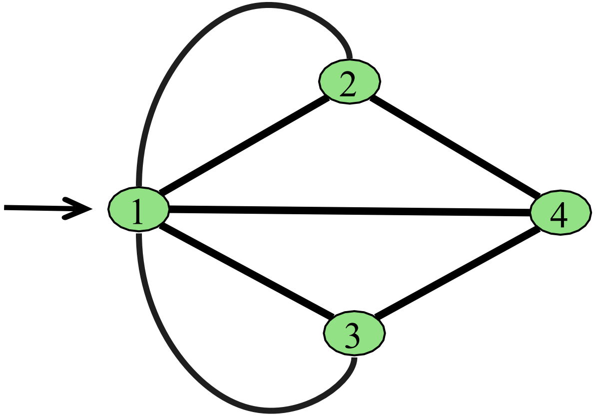
Em 1735, o matemático suíço **Leonard Euler** mostrou que **não existe** uma solução para o problema.

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.



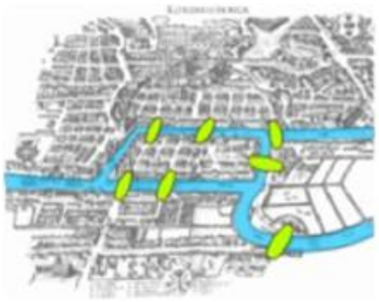


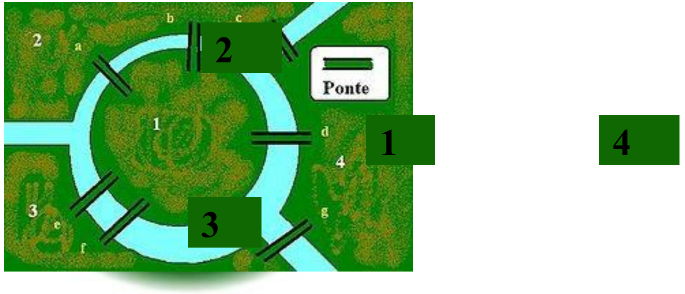


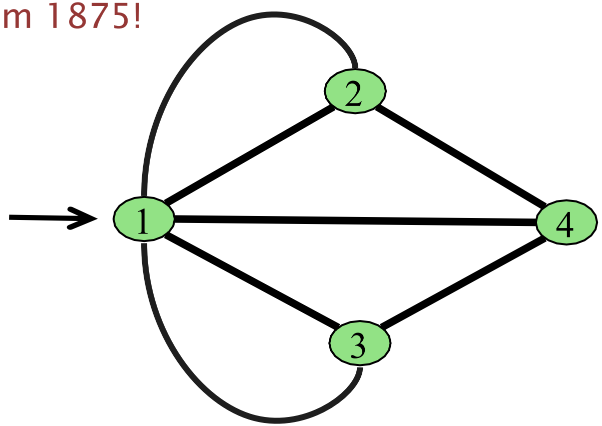
www.numericana.com/answer/graphs.htm#euler

As Pontes de Königsberg

- **Teorema de Euler**: Se um grafo não-direcionado tiver dois, ou nenhum vértice impar, 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!

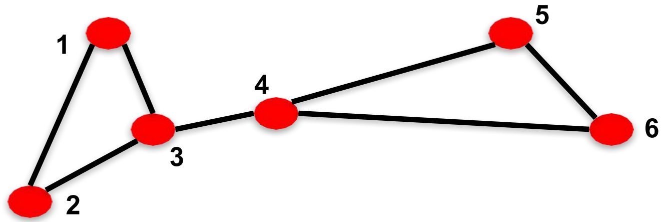






www.numericana.com/answer/graphs.htm#euler

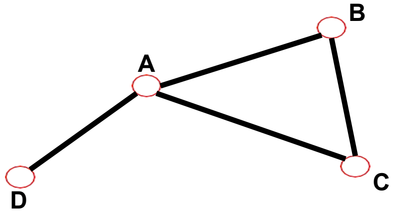
Grafos



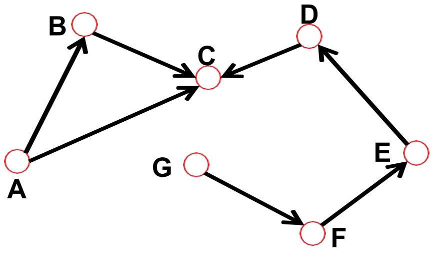
Conjunto composto pelo **par ordenado *G* = (*N* , *L*)**

- **Ordem**: # vértices n(G) = 6
- **Tamanho**: # arestas l(G) = 7

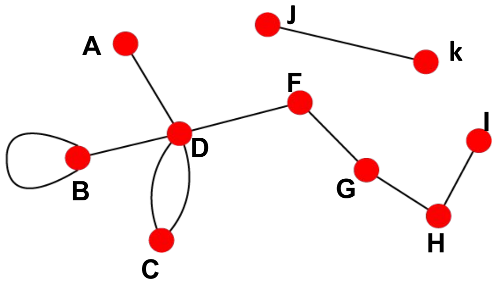
- Dado G = (N, L), o maior número de arestas de G = onde *n* é a ordem do grafo. $\binom{n}{2} = \frac{n(n-1)}{2} \leq n^2$



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 ou Redes?

REDES – sistemas reais

World Wide Web

Rede metabólica

Rede social

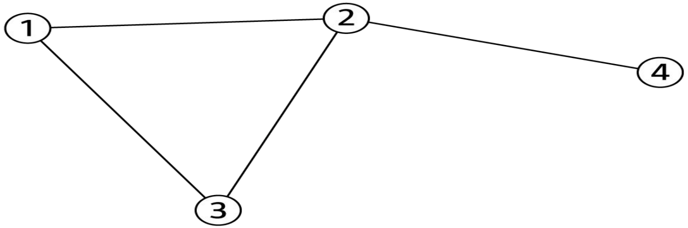
Nomenclatura: nó, ligação.

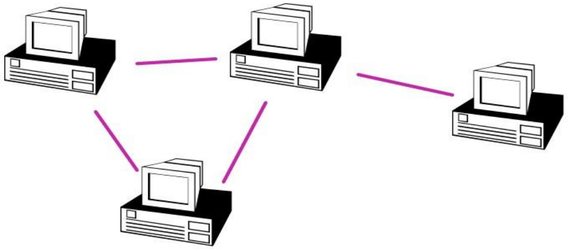
GRAFO – representação matemática

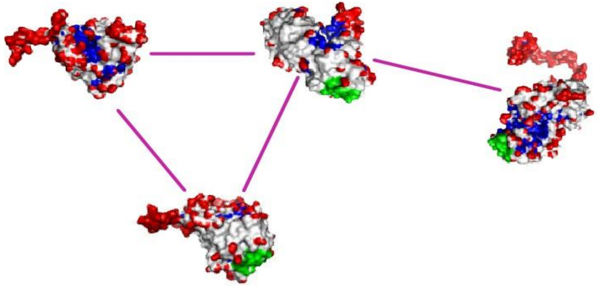
Grafo web

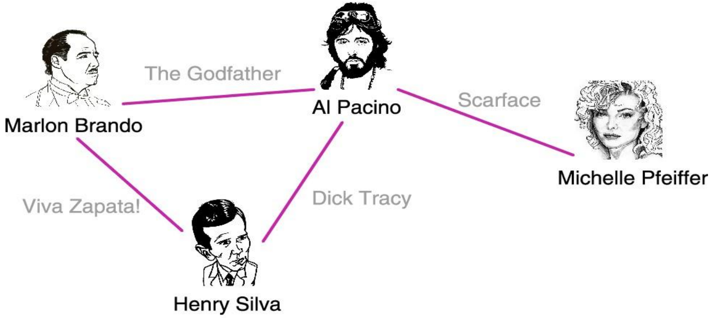
Grafo social

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

Number of nodes (N) and links (L), and the average degree for each network $\langle k \rangle$.

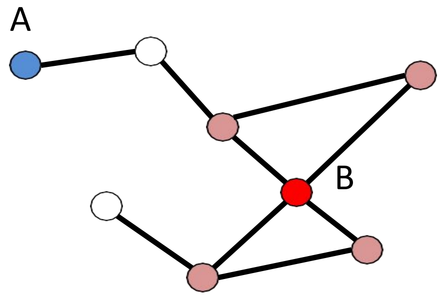
For directed networks the average degree shown is the average in- or out-degrees $\langle k \rangle = \langle k_{in} \rangle = \langle k_{out} \rangle$

GRAU, GRAU MÉDIO E DISTRIBUIÇÃO DE GRAU

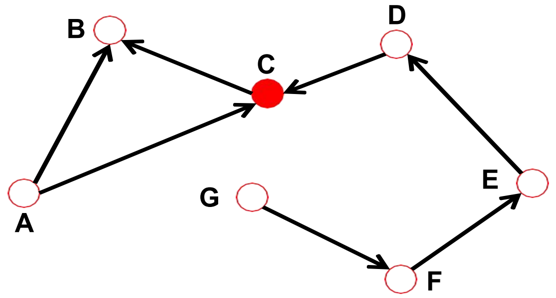
Grau e Grau Médio

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

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



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



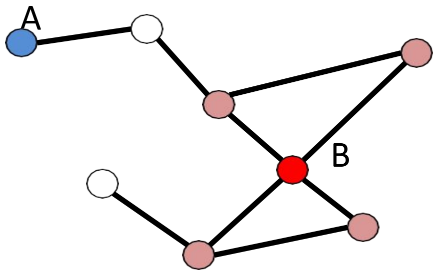
$$k_C^{in} = 2 \qquad k_C^{out} = 1 \qquad k_C = 3$$

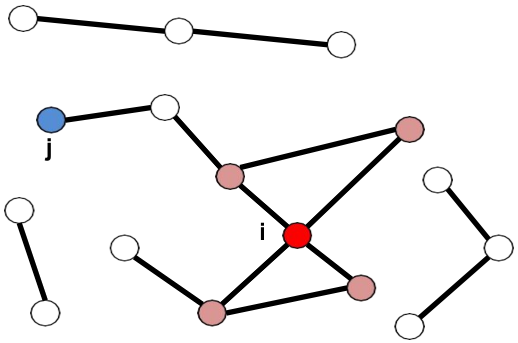
Grau e Grau Médio

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

- **Grau Médio <K>**:

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





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

Distribuição de Grau

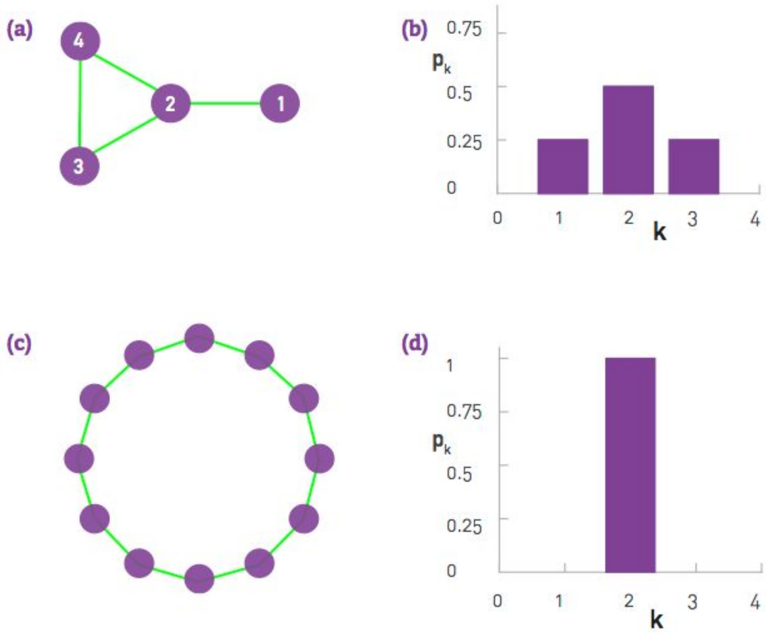
- **P(k):** probabilidade que um vértice escolhido aleatoriamente tenha grau **k.**

- **Distribuição empírica de grau**: frequência de vértices com grau k.

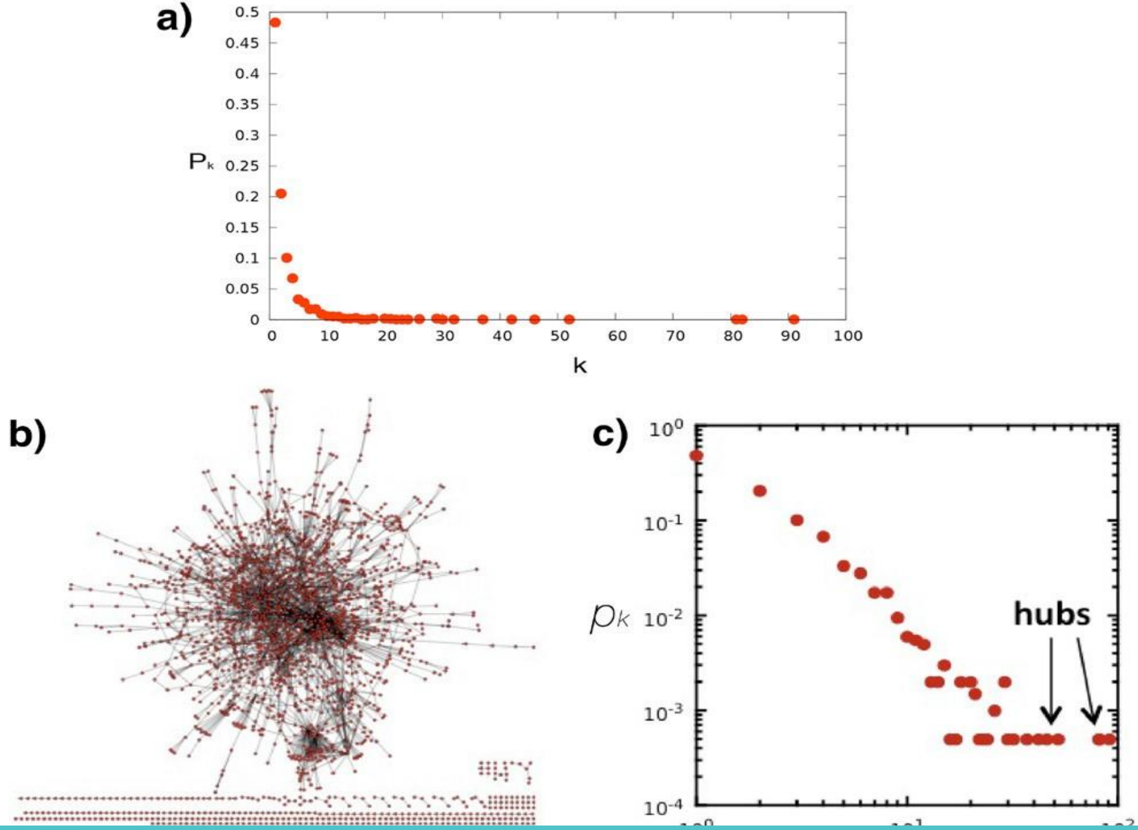
$$P(k_i = k) = P(k) = P_k = \frac{n_k}{\sum_k n_k} = \frac{n_k}{n}$$

k_i = grau de cada nó,

n_k = # vértices com grau k



Distribuição de Grau em redes PPI



MATRIZ DE ADJACÊNCIA

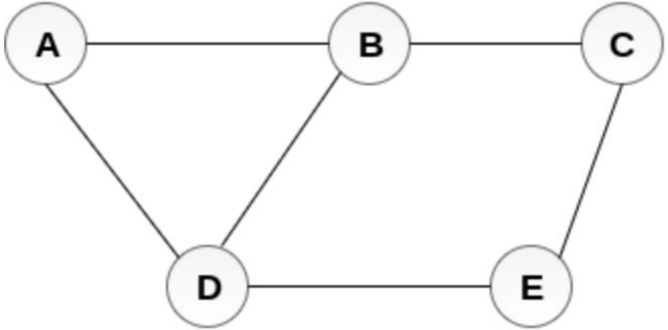
Grafos

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.

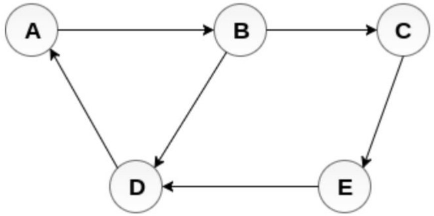


Undirected Graph

$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
\begin{bmatrix}
0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 0
\end{bmatrix}
\end{array}
$$

Adjacency Matrix

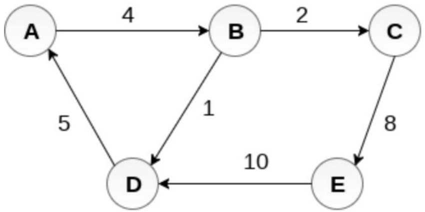
Grafos



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

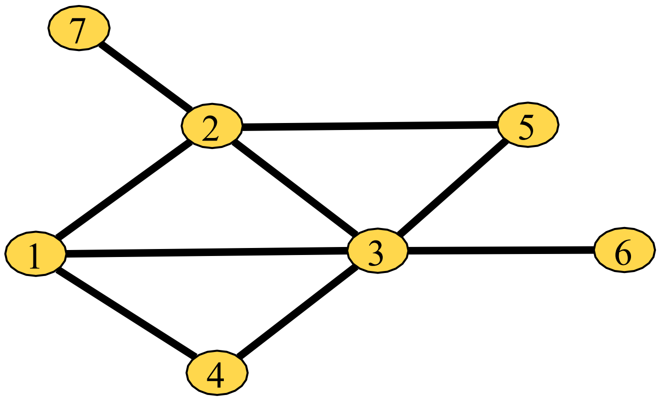
	A	B	C	D	E
A	0	4	0	0	0
B	0	0	2	1	0
C	0	0	0	0	8
D	5	0	0	0	0
E	0	0	0	10	0

Adjacency Matrix

Representação de Grafos

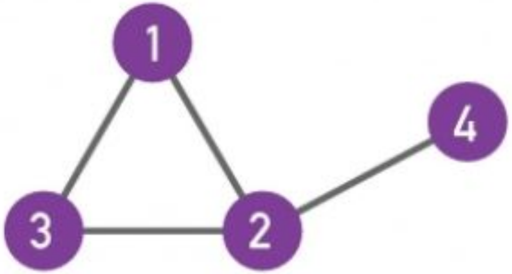
Matriz de adjacência $(n \times n)$

- $a_{ij} = 1$, 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)



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

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

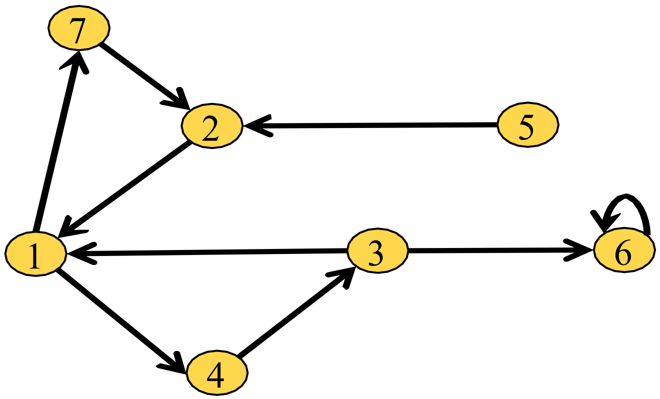
$$A_{ij} = A_{ji} \qquad A_{ii} = 0$$

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

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

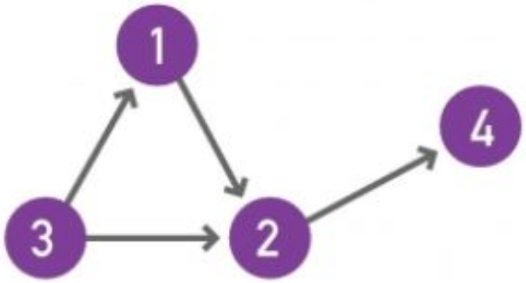
Representação de Grafos

- 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)



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

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

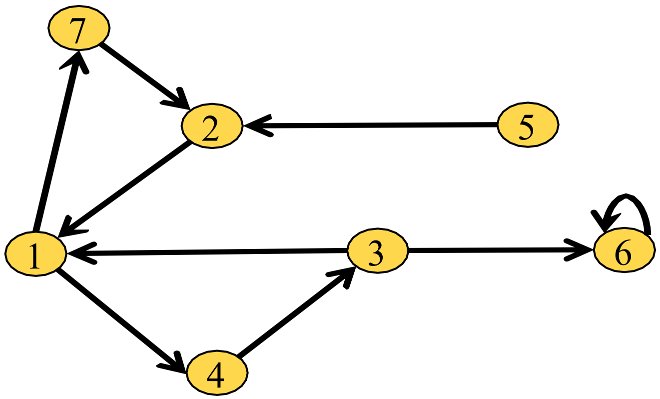
$$A_{ij} \neq A_{ji} \qquad A_{ii} = 0$$

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

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

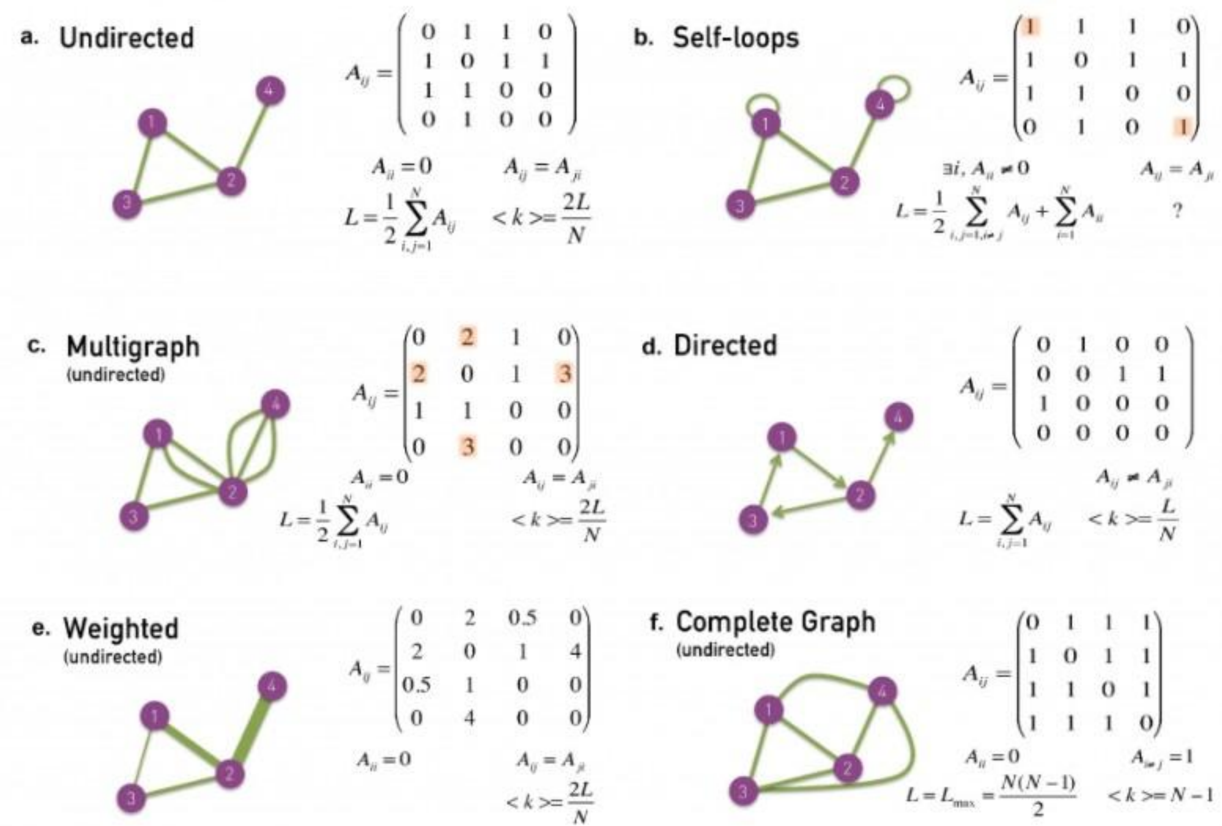
Representação de Grafos

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



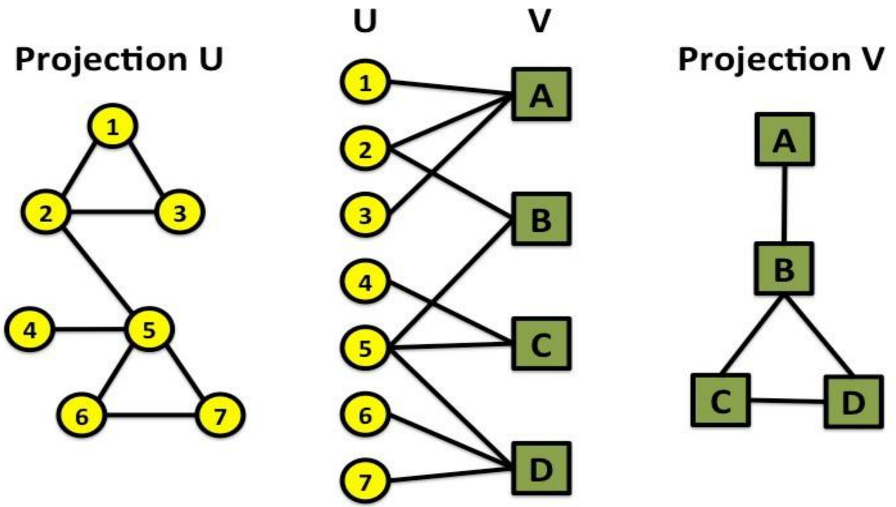
1: 4, 7
2: 1
3: 1, 6
4: 3
5: 2
6: 6
7: 2

Resumo: Tipos de Redes



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.

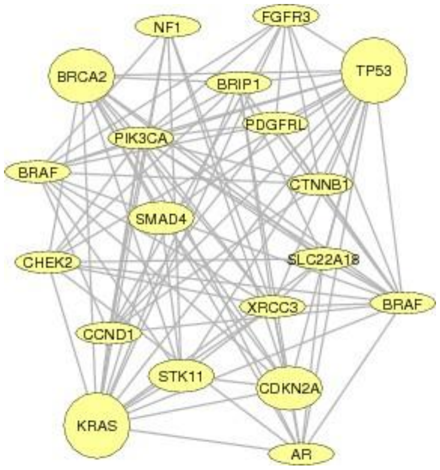


Rede de atores de Hollywood

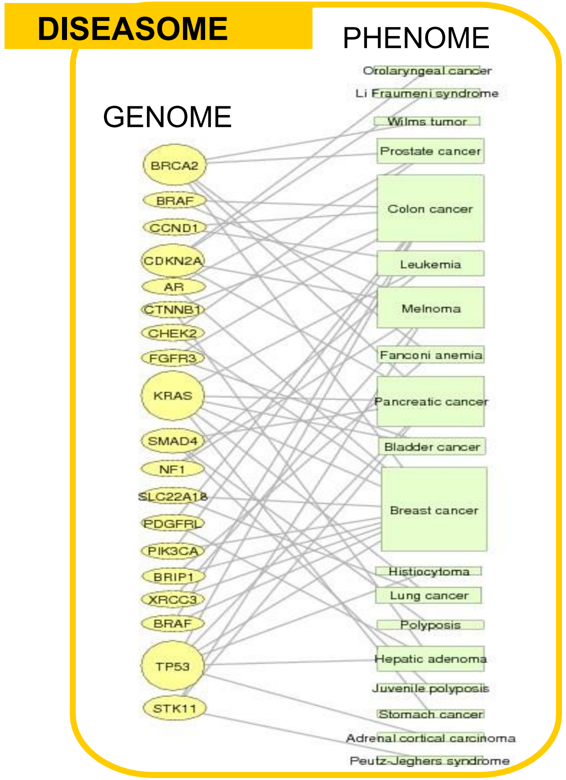
Redes de colaborações

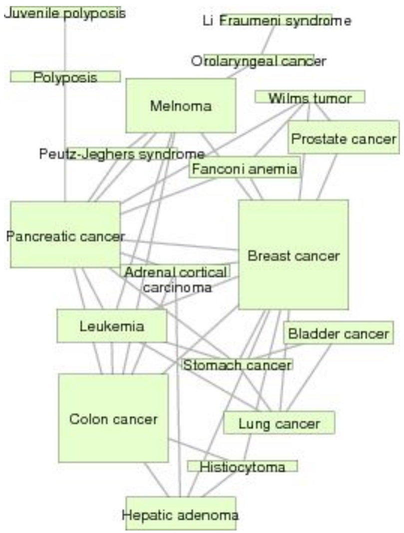
Rede de doenças (diseasome)

Grafos Bipartidos (Bígrafo)



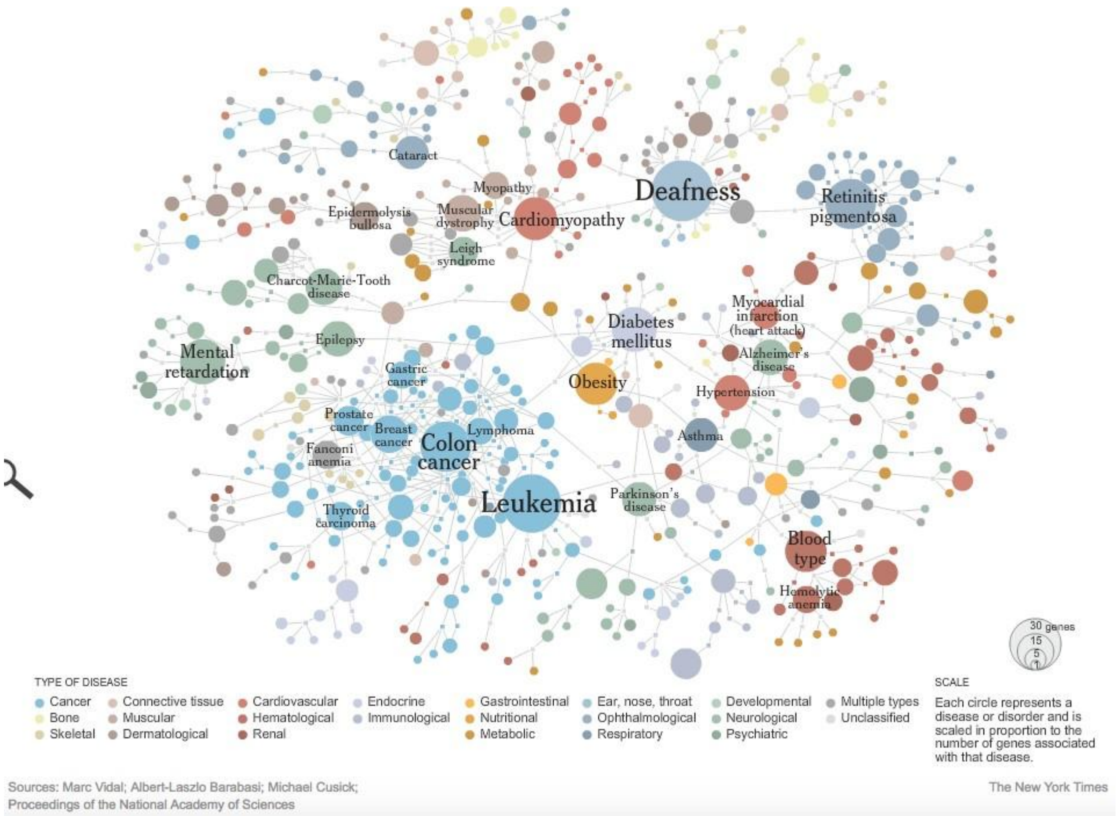
Gene network





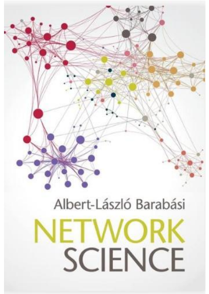
Disease network

Human Diseasome Network

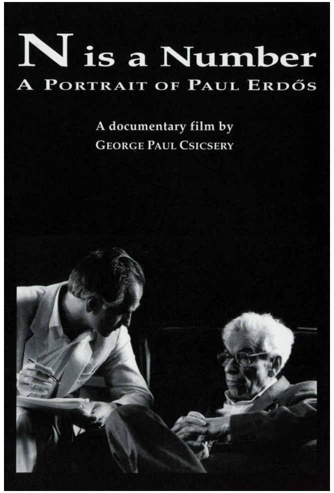


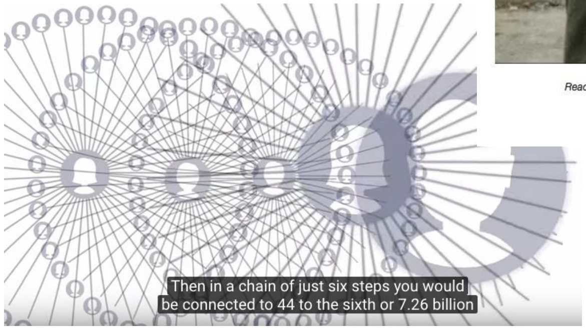
Para Saber Mais...

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



www.oakland.edu/enp/compute/





www.youtube.com/watch?v=TcxZSmzPw8k

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

www.oracleofbacon.org

www.youtube.com/watch?v=BQ7UDWn_uws