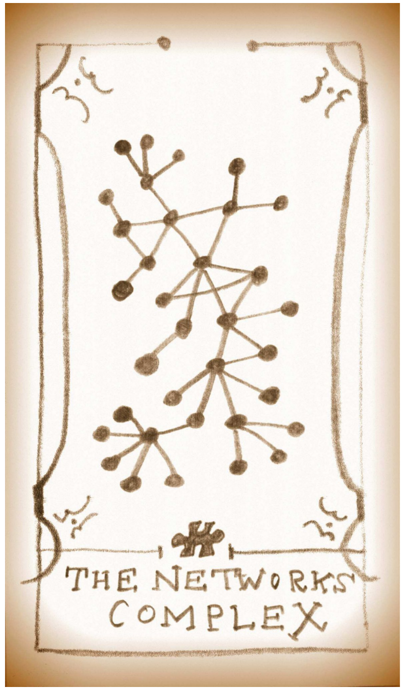
Análise de Redes

Uma breve Introdução

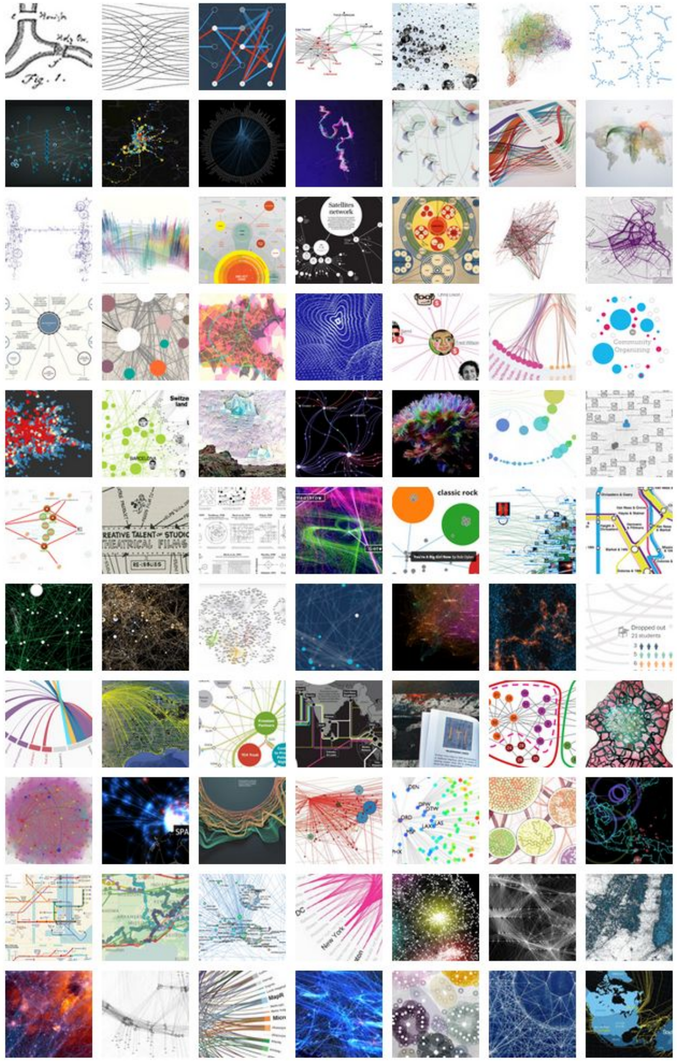
Prof. Patrick Terrematte

Análise de Redes

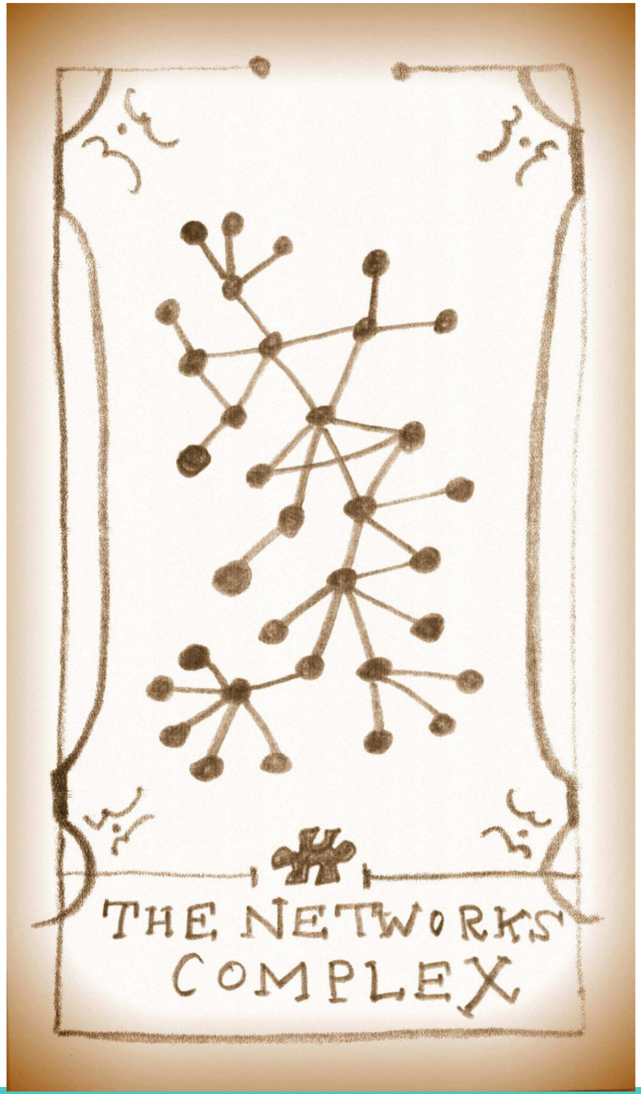
- Elementos de redes:
  - definições básicas, densidade, esparcidade, subredes, graus, e representações.
- Small worlds
  - Assortatividade, Caminhos, Distâncias, Componentes de conexões, Coeficientes de clustering.
- Hubs
  - Distribuições de centralidade, Decomposição, Betweennes, Eigenvector Centrality.
- Aplicações
  - Estudo de caso da Wikipedia
  - Estudo de caso do Twitter

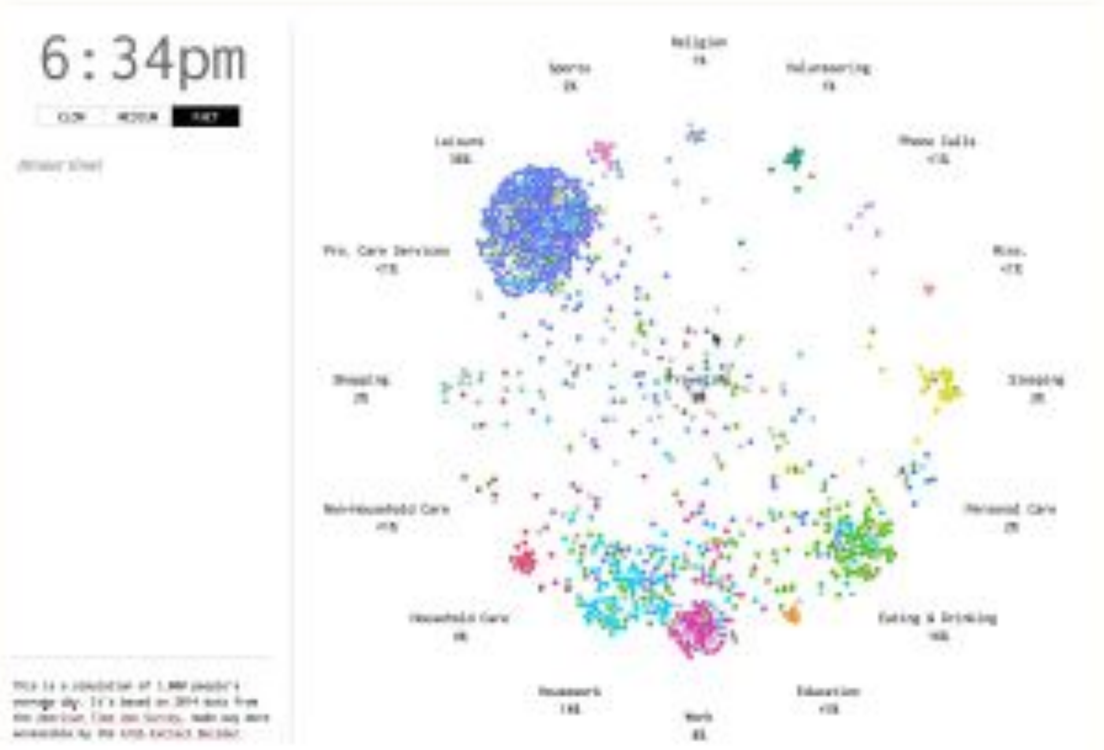


Datasets clássicos

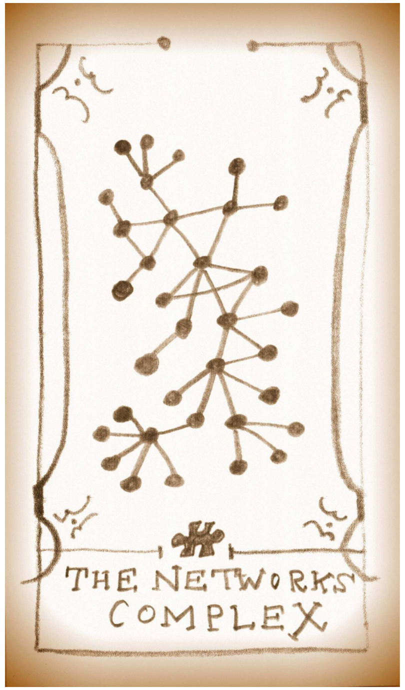


http://www.visualcomplexity.com/vc/



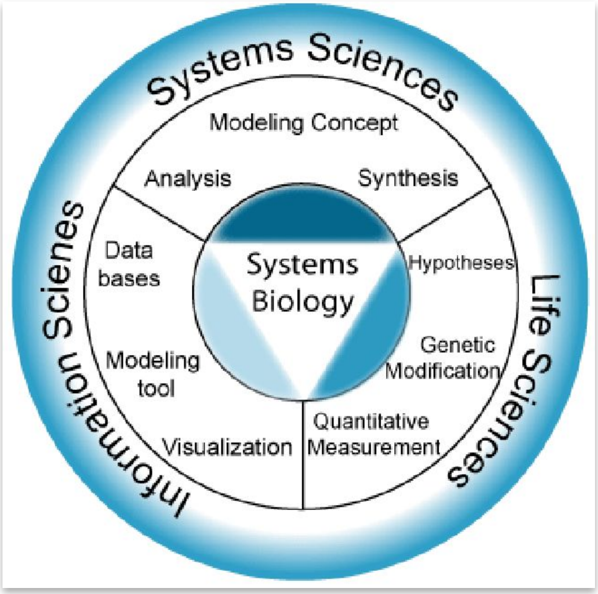


http://flowingdata.com/2015/12/15/a-day-in-the-life-of-americans

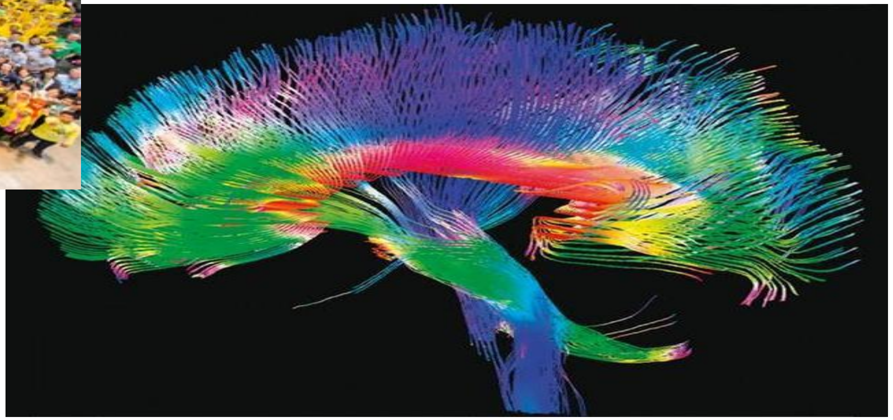


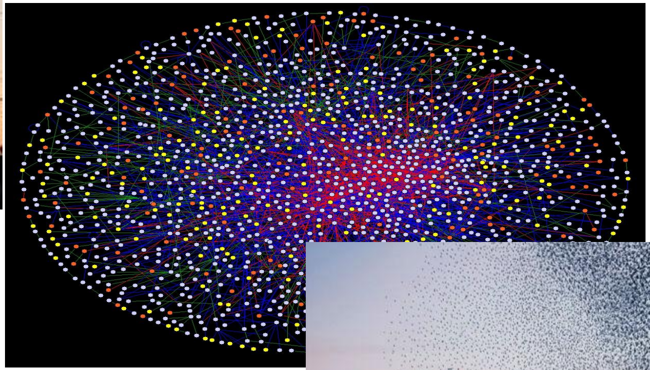
Biologia de Sistemas como aplicação de Teoria dos Grafos

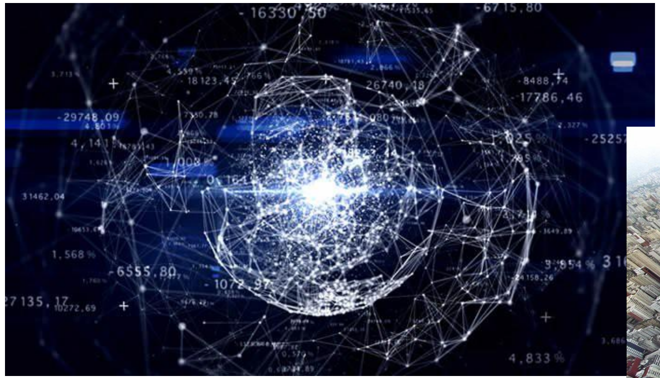
- Grande quantidade de **dados experimentais**.
- Proposição de **modelos matemáticos** que explicam aspectos significativos dos dados.
- **Simulações computacionais** e análises numéricas.
- Avaliação da qualidade do modelo por comparação dos resultados com dados experimentais.



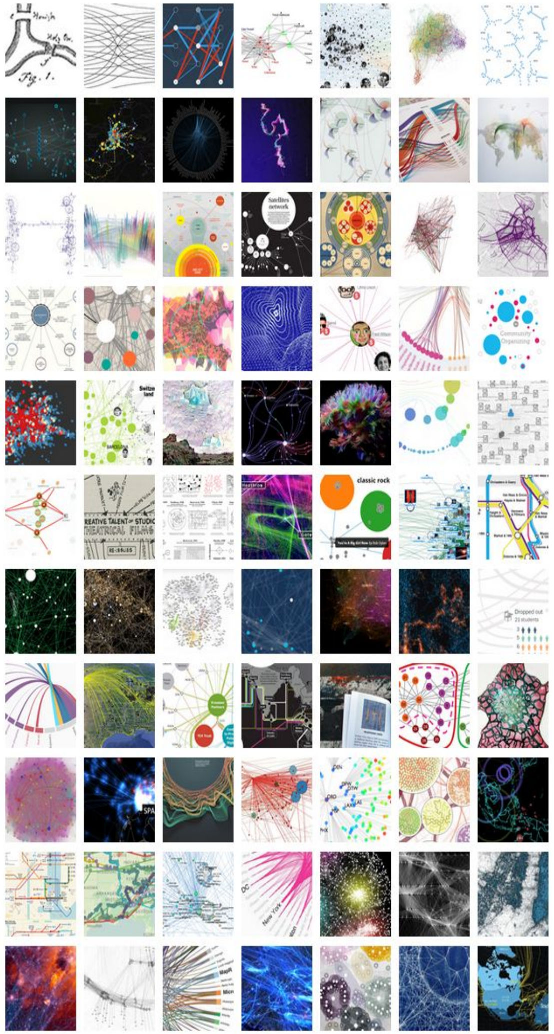
Sistemas complexos como aplicação de Teoria dos Grafos







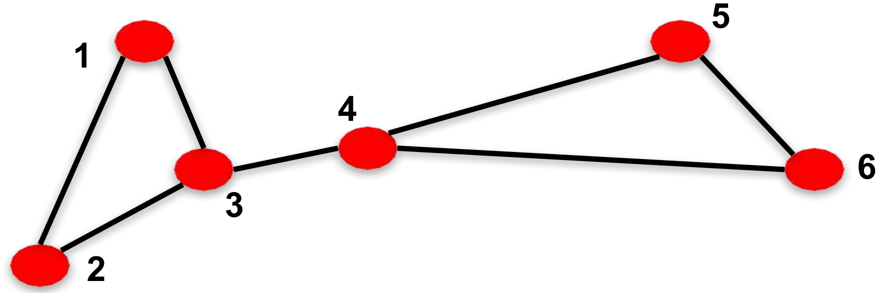
Teoria de Grafos



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

Grafos

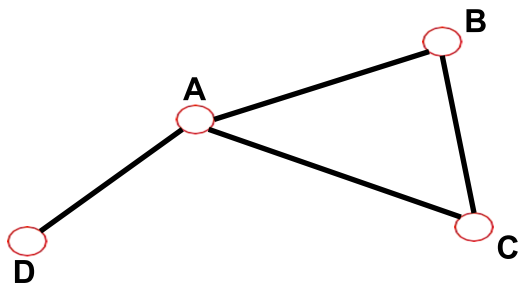
Conjunto composto pelo **par ordenado** $\boldsymbol{G = (N\ ,\ L\ )}$

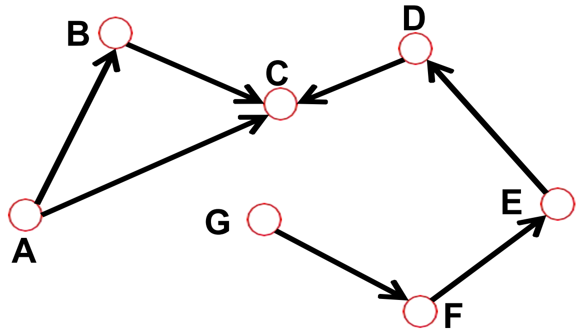


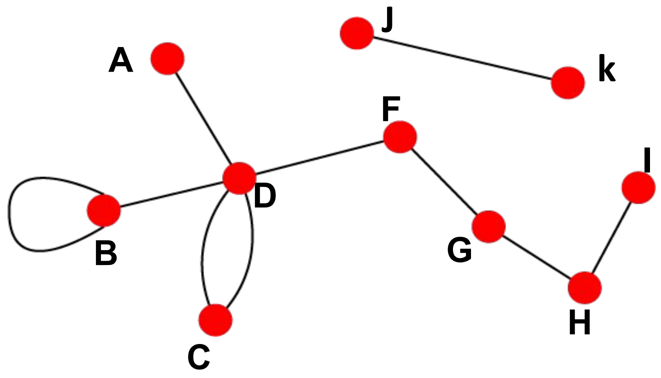
- **Ordem**: # vértices n(G) = 6
- **Tamanho**: # arestas l(G) = 7

- 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

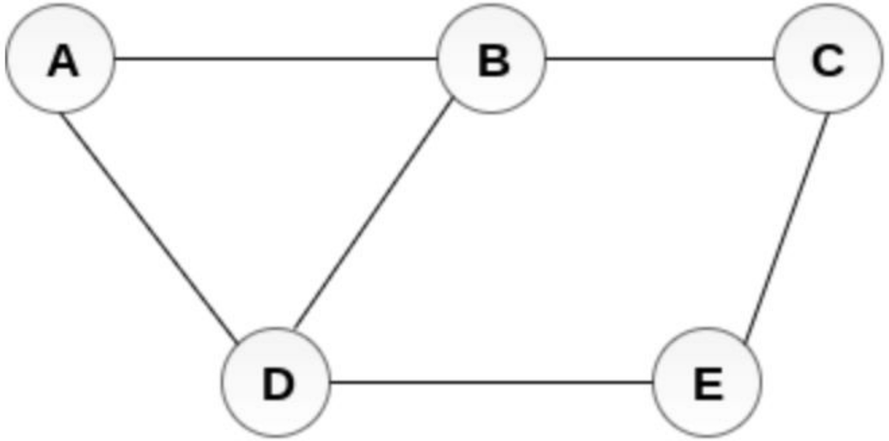
Grafos

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.

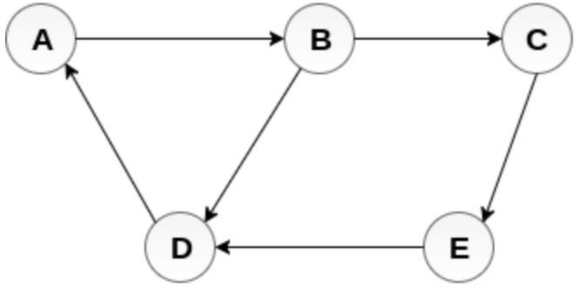


Undirected Graph

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

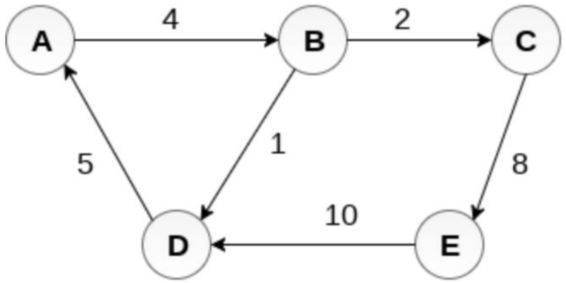
Adjacency Matrix

Grafos



	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

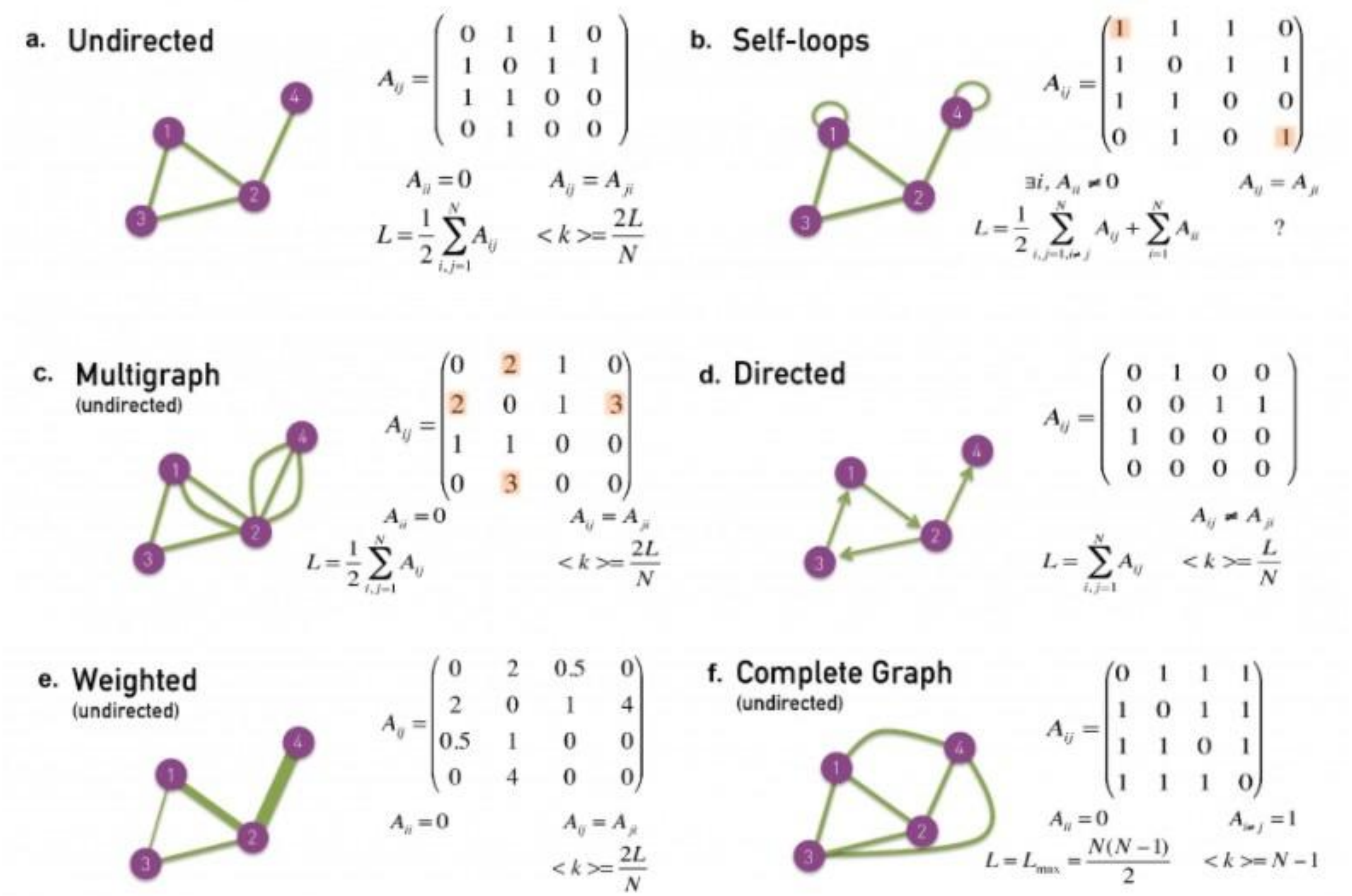
Directed Graph **Adjacency Matrix**



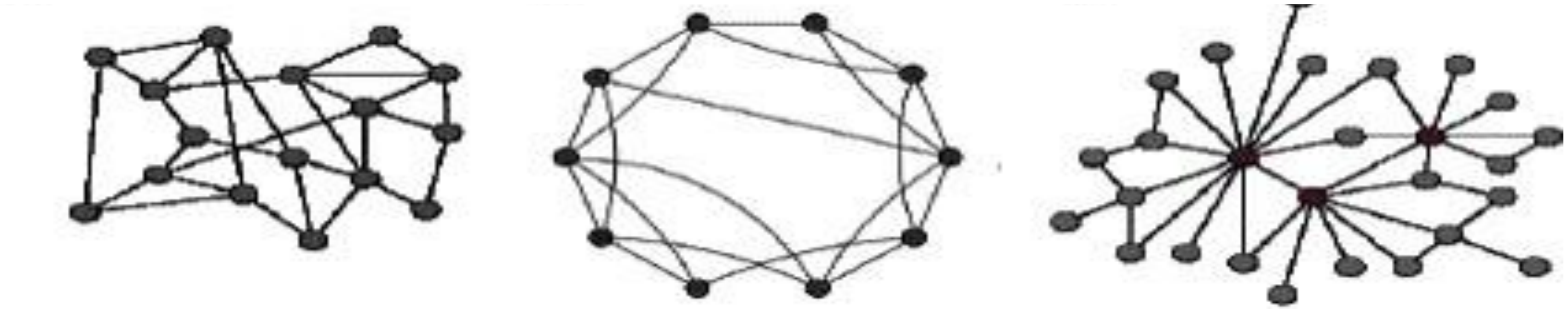
	A	B	C	D	E
A	0	4	0	0	0
B	0	0	2	1	0
C	0	0	0	0	8
D	5	0	0	0	0
E	0	0	0	10	0

Weighted Directed Graph **Adjacency Matrix**

Resumo: Tipos de Redes

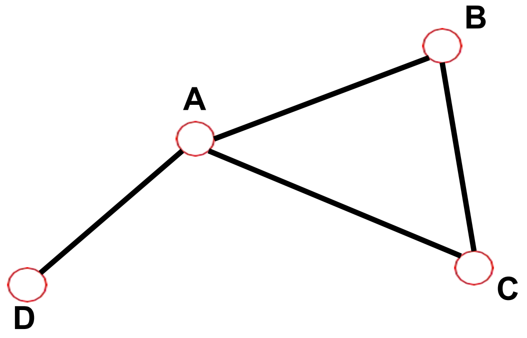


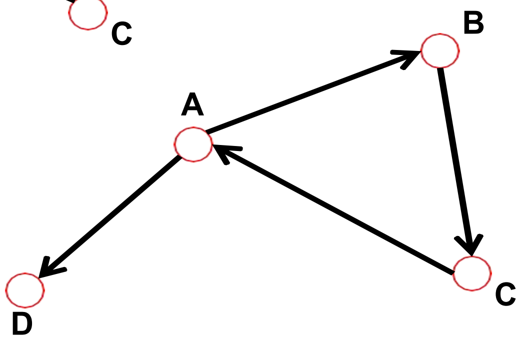
Propriedades de Grafos



Caminho

- **Caminho**: sequência de vértices consecutivos conectados por arestas <**s**, u, v, … , **t**>.
- Em um grafo direcionado, o caminho segue o sentido da aresta. AB ≠ BA.
- **Distância** (*caminho mínimo, caminho geodésico*): o menor caminho entre dois vértices.





Grafo não-direcionado

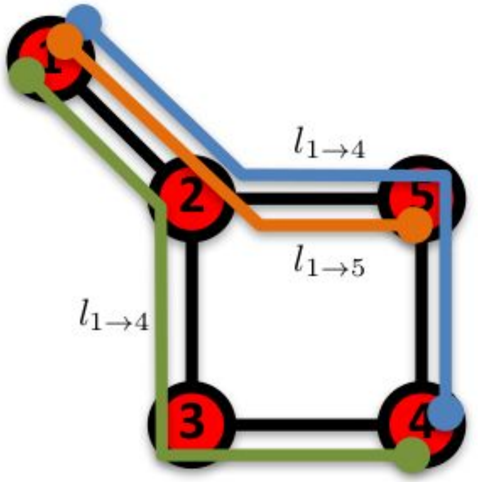
<B, C, A, D> é caminho de comprimento **3**.

<B, A, D, C> não é caminho.

Grafo direcionado

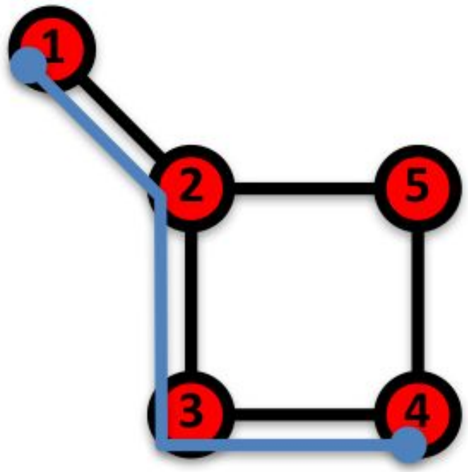
<A, B, C> é caminho.
<A, C, B> não é caminho.

Caminhos



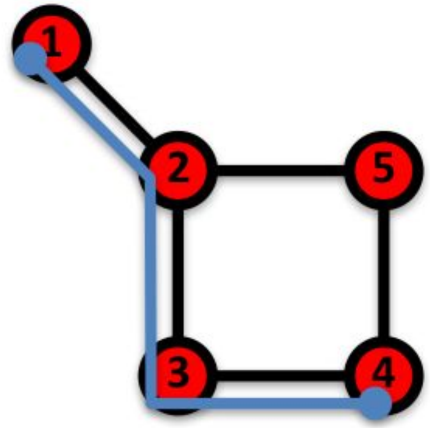
$$l_{1\to4} = 3 \quad l_{1\to5} = 2$$

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



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

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



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

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

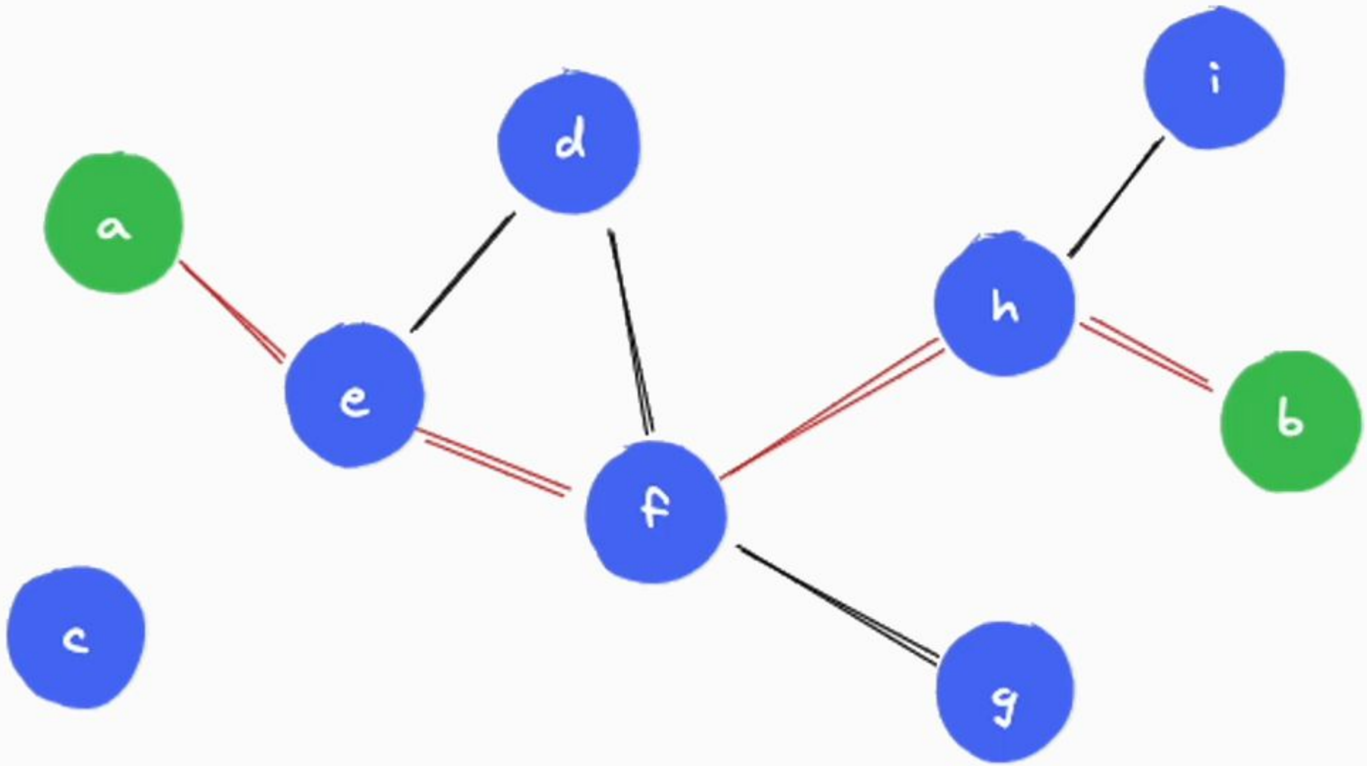
Diametro da rede

$$l_{max} = \max_{ij} \ l_{ij}$$

(src, dest)
(a,b)
a - e - f - h - b
(a,c)
-
(a,d)
a - e - d
(a,e)
a - e
....

(b,c)
-
(b,d)
b - h - f - d
(b,e)
b - h - f - e
(b,f)
b - h - f
....

....



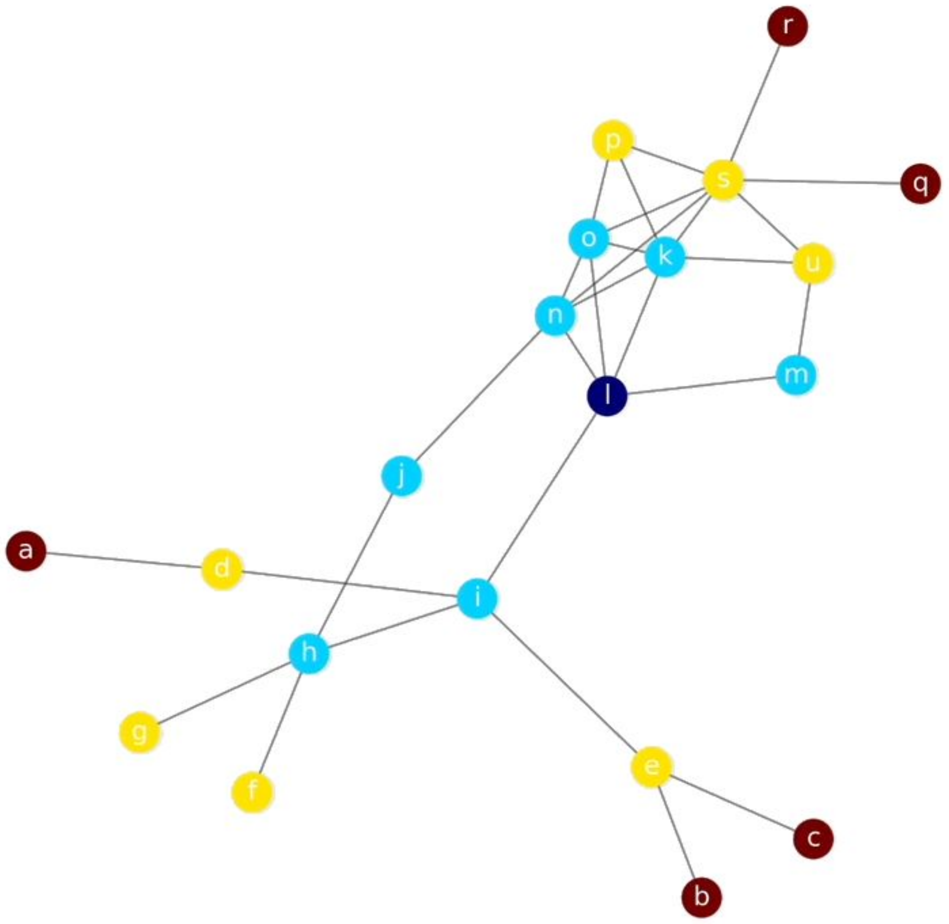
Diameter

The diameter of a network is the maximum eccentricity.

```
nx.eccentricity(g)
{'a': 6, 'b': 6, 'c': 6, 'd': 5,
 'e': 5, 'f': 5, 'g': 5, 'h': 4,
 'i': 4, 'j': 4, 'k': 4, 'l': 3,
 'm': 4, 'n': 4, 'o': 4, 'p': 5,
 'q': 6, 'r': 6, 's': 5, 'u': 5}

nx.diameter(g)
6

[k for k,v in nx.eccentricity(g).items()
if v == nx.diameter(g)]
['a', 'b', 'c', 'q', 'r']
```



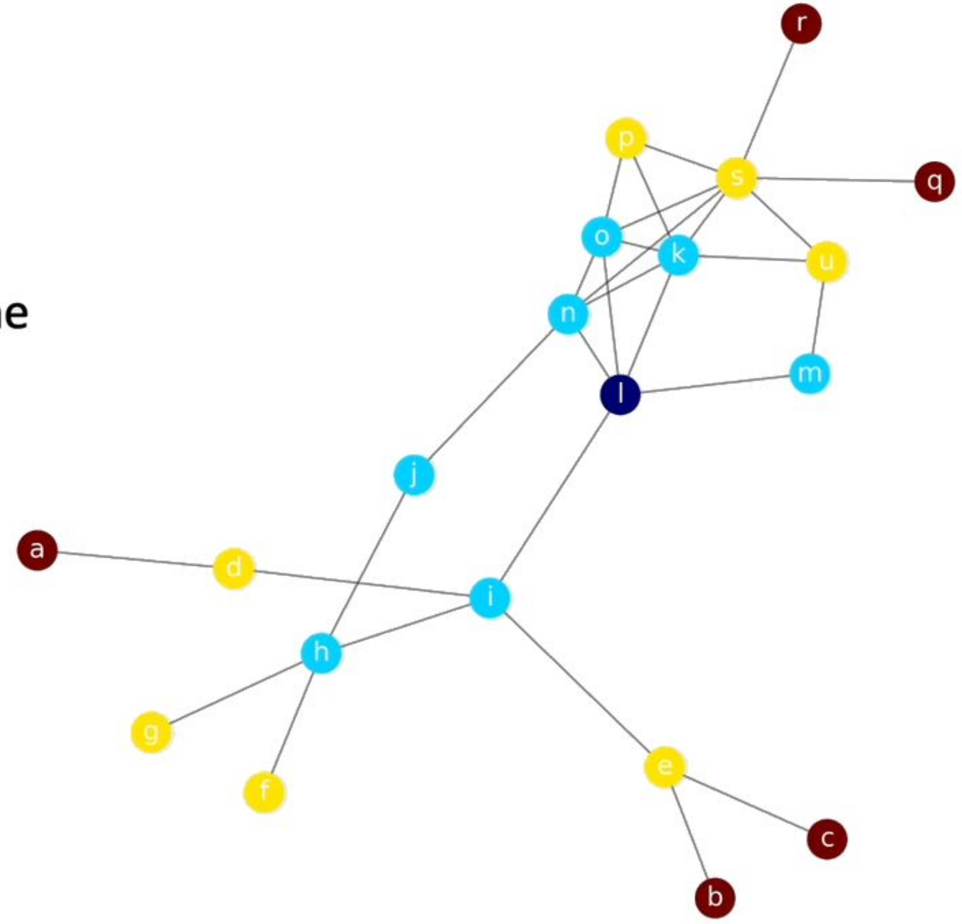
Periphery

The periphery of a network is a set of all nodes whose eccentricity equals the diameter.

```
nx.eccentricity(g)
{'a': 6, 'b': 6, 'c': 6, 'd': 5,
 'e': 5, 'f': 5, 'g': 5, 'h': 4,
 'i': 4, 'j': 4, 'k': 4, 'l': 3,
 'm': 4, 'n': 4, 'o': 4, 'p': 5,
 'q': 6, 'r': 6, 's': 5, 'u': 5}

nx.diameter(g)
6

nx.periphery(g)
['a', 'b', 'c', 'q', 'r']
```

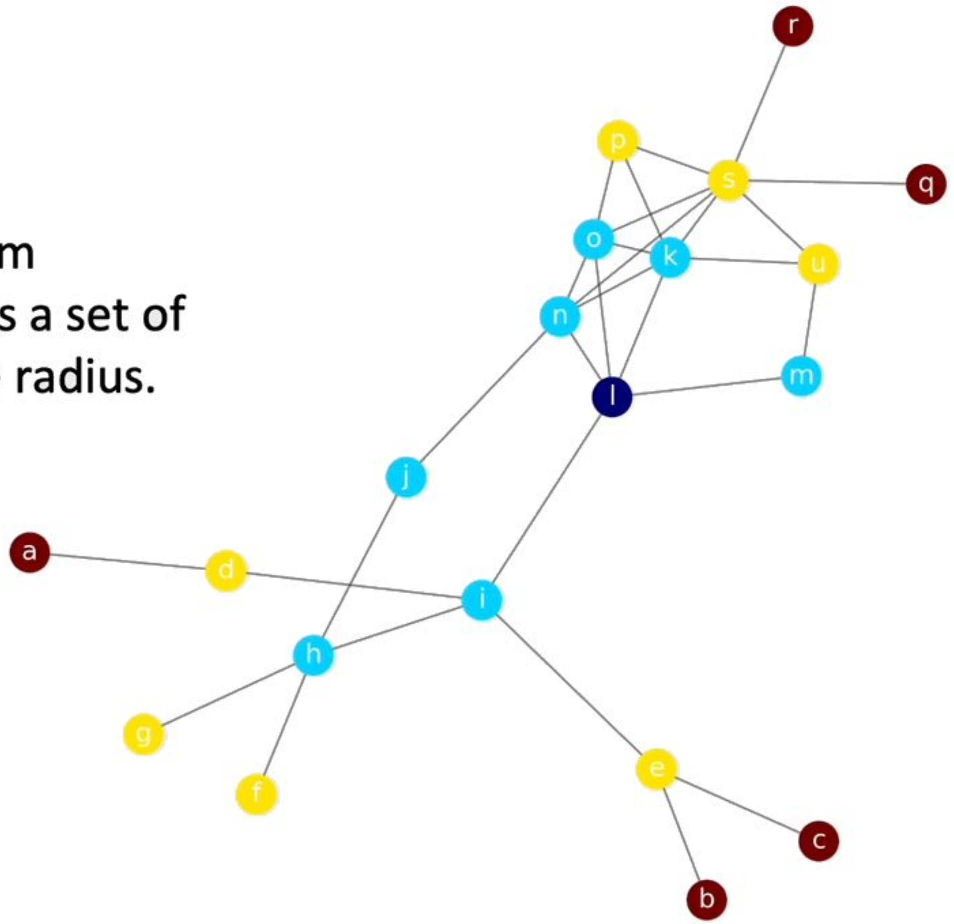


Radius & Center

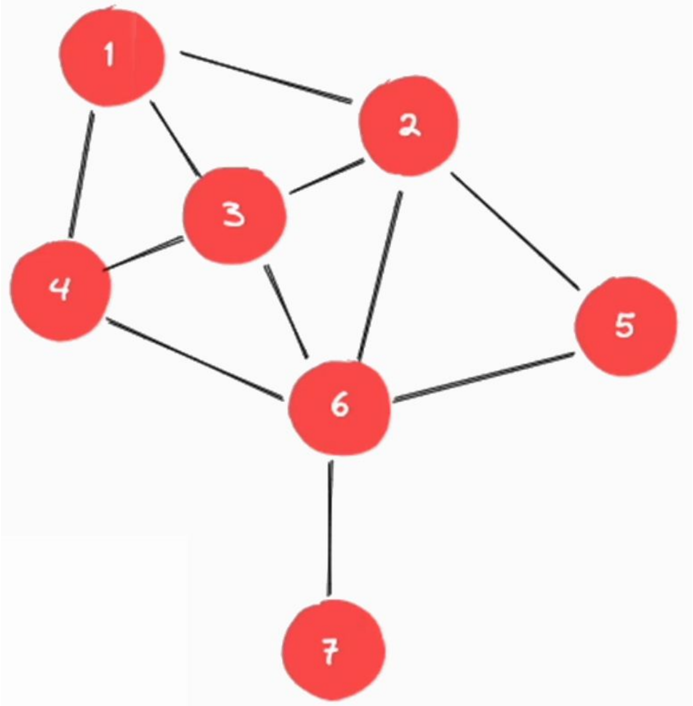
The radius of a network is the minimum eccentricity. The center of a network is a set of all nodes whose eccentricity equal the radius.

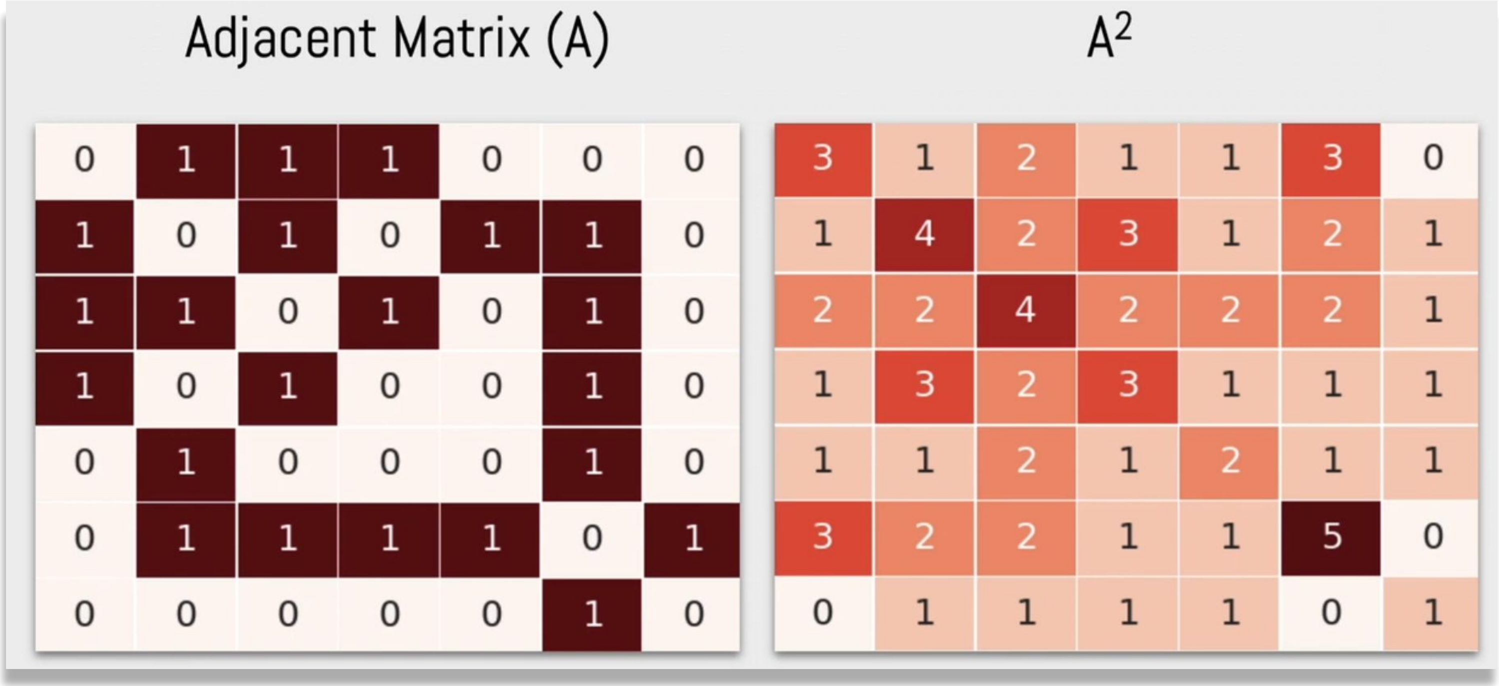
```
nx.eccentricity(g)
{'a': 6, 'b': 6, 'c': 6, 'd': 5,
 'e': 5, 'f': 5, 'g': 5, 'h': 4,
 'i': 4, 'j': 4, 'k': 4, 'l': 3,
 'm': 4, 'n': 4, 'o': 4, 'p': 5,
 'q': 6, 'r': 6, 's': 5, 'u': 5}

nx.radius(g)
3
[k for k,v in nx.eccentricity(g).items()
if v == nx.radius(g)]
['l']
nx.center(g)
['l']
```



Walk





Adjacent Matrix (A)

0	1	1	1	0	0	0
1	0	1	0	1	1	0
1	1	0	1	0	1	0
1	0	1	0	0	1	0
0	1	0	0	0	1	0
0	1	1	1	1	0	1
0	0	0	0	0	1	0

A^2

3	1	2	1	1	3	0
1	4	2	3	1	2	1
2	2	4	2	2	2	1
1	3	2	3	1	1	1
1	1	2	1	2	1	1
3	2	2	1	1	5	0
0	1	1	1	1	0	1

Os caminhos de comprimento 2 da rede.
A diagonal representa os graus de cada nó.

Grafos Completos

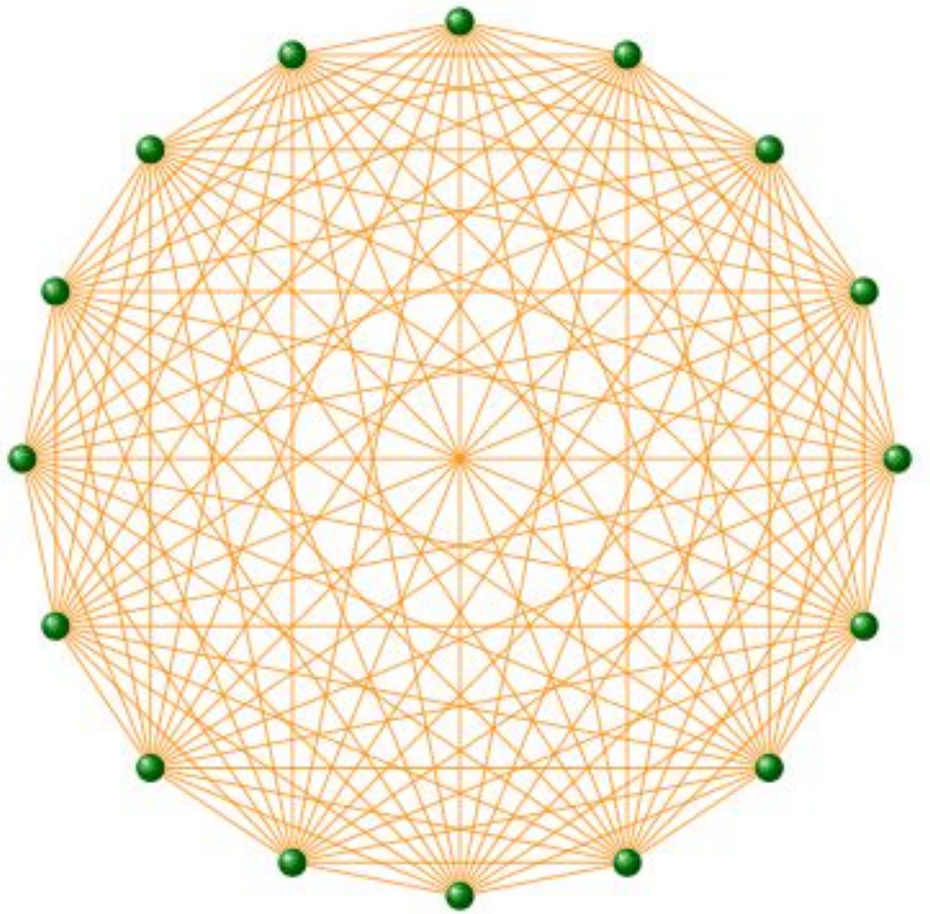
- Grafo com tamanho $L = L_{max}$ e grau médio $\langle k \rangle = N\text{-}1$.

- 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 esparço: $L \sim N$.
  - Grafo denso: $L \sim N^2$.



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.

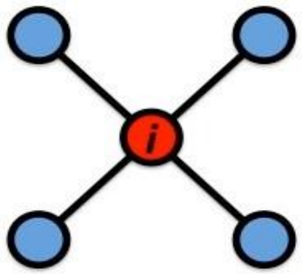
$$C_i = \frac{E_i}{\binom{d_i}{2}}$$

de arestas entre os vizinhos de i

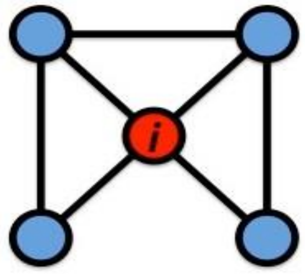
grau do vértice i

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

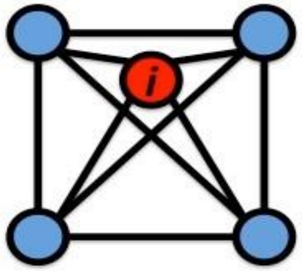
- CC não expressa uma propriedade do vértice e sim dos seus vizinhos!



CC = 0/12 = 0

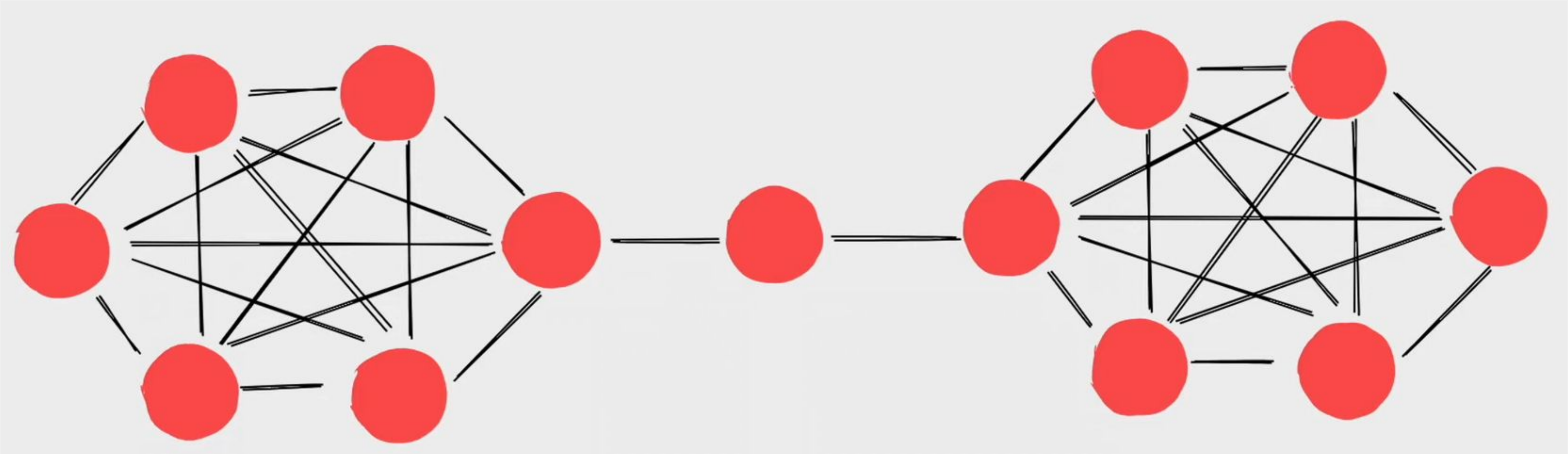


CC = (2*3)/12 = 0,5



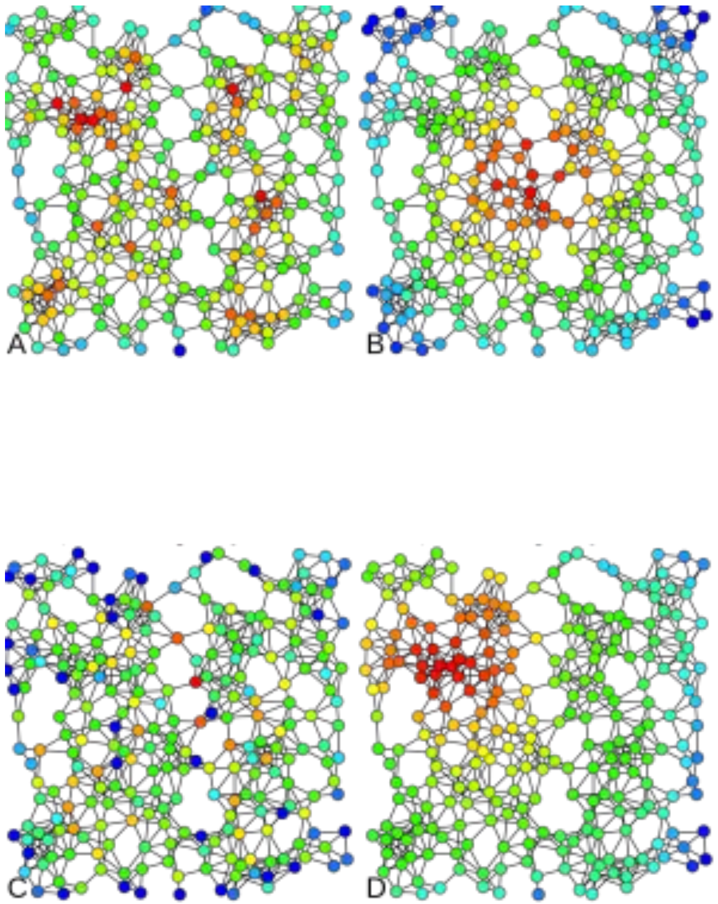
CC = 12/12 = 1

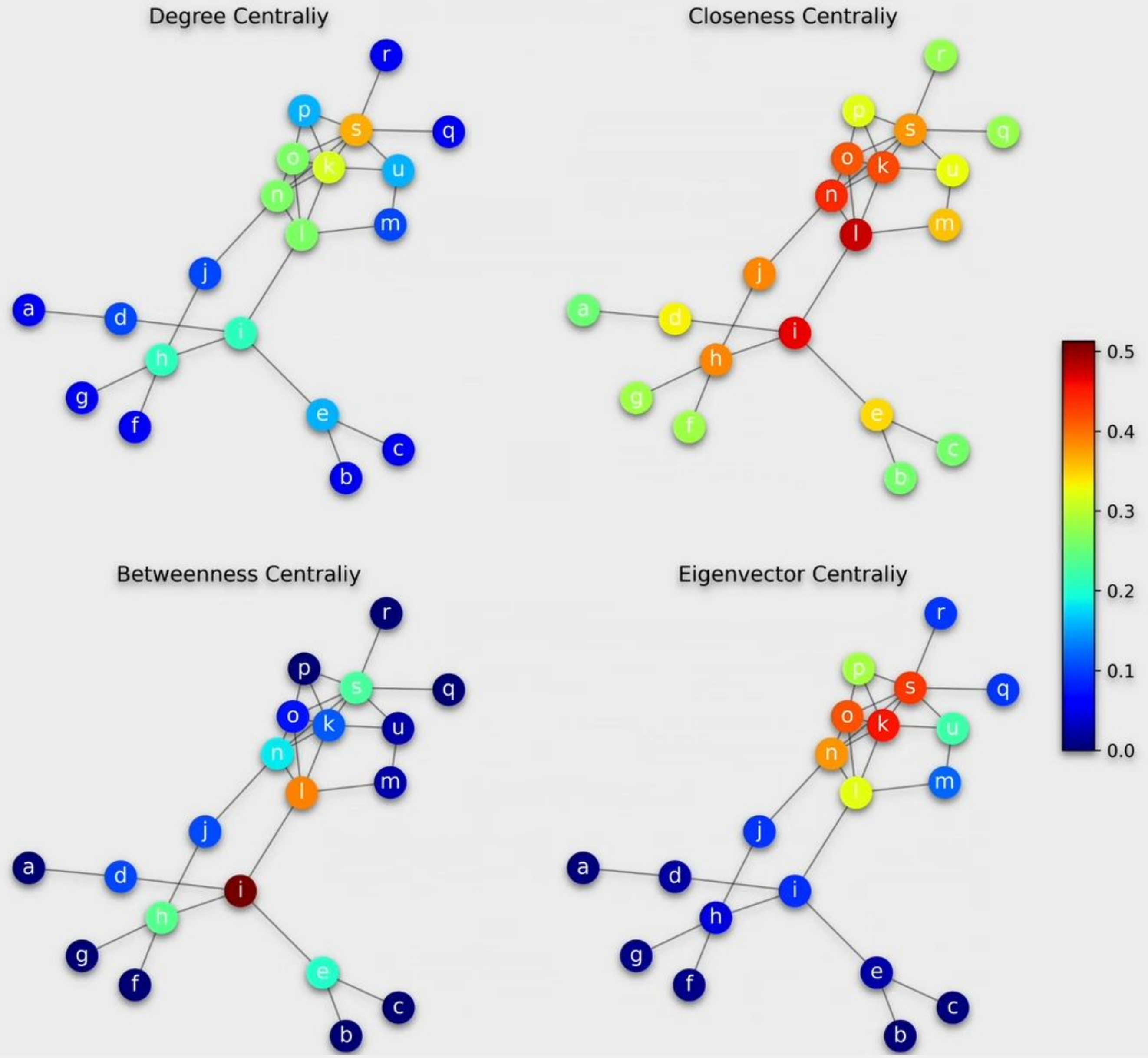
Como medir a importância de um nó?



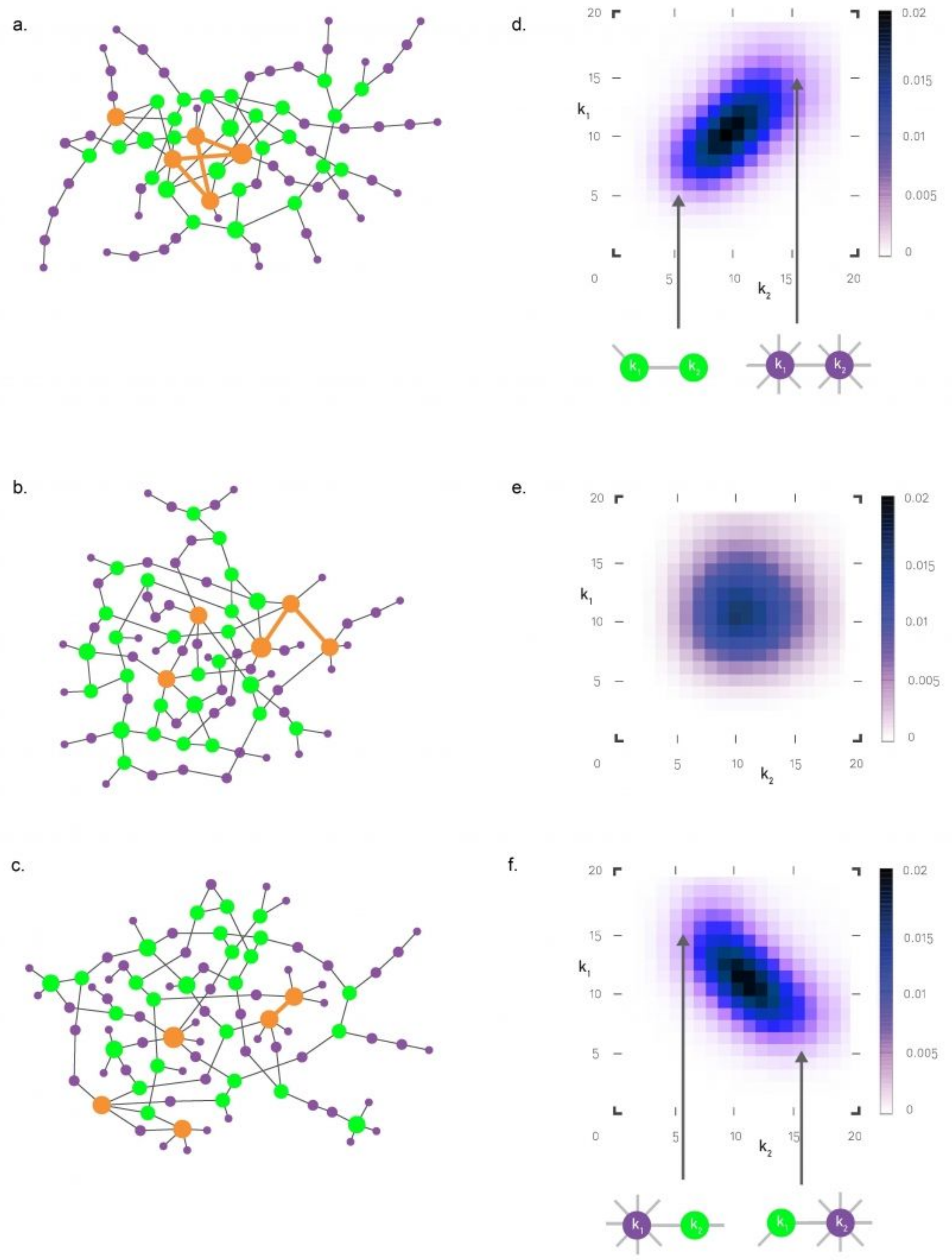
Medidas de Centralidade

- 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.




Degree Centraliy
Closeness Centraliy
Betweenness Centraliy
Eigenvector Centraliy
0.5
0.4
0.3
0.2
0.1
0.0

- Assortative Network
- Neutral Network
- Disassortative Network



Here networks with same degree distribution.

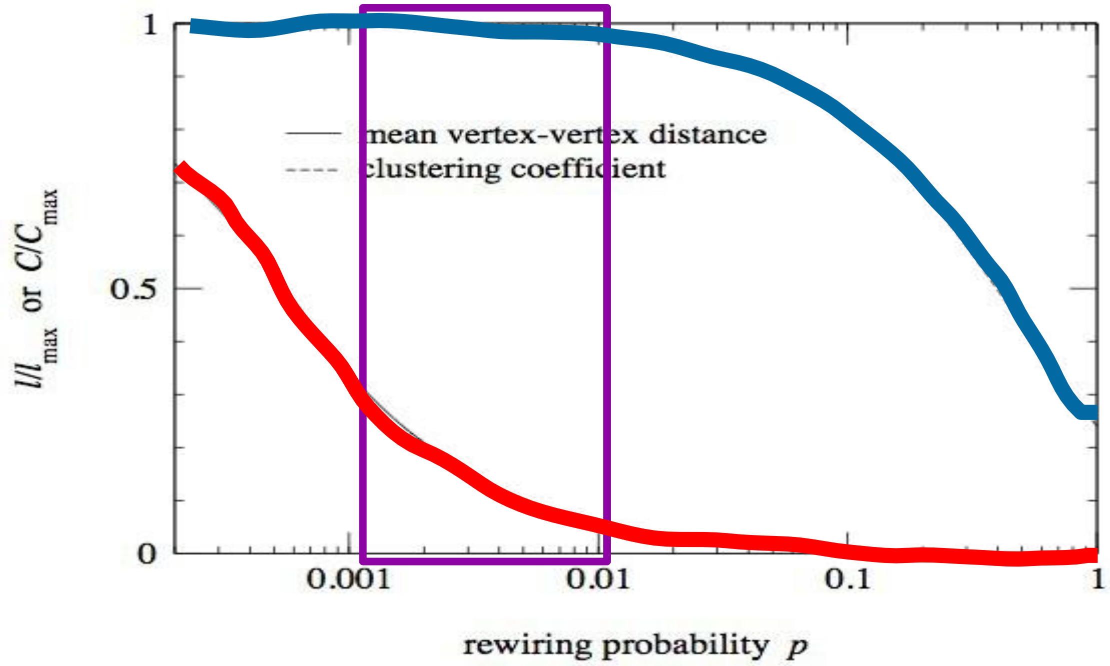
In orange, five highest degree nodes.

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*)

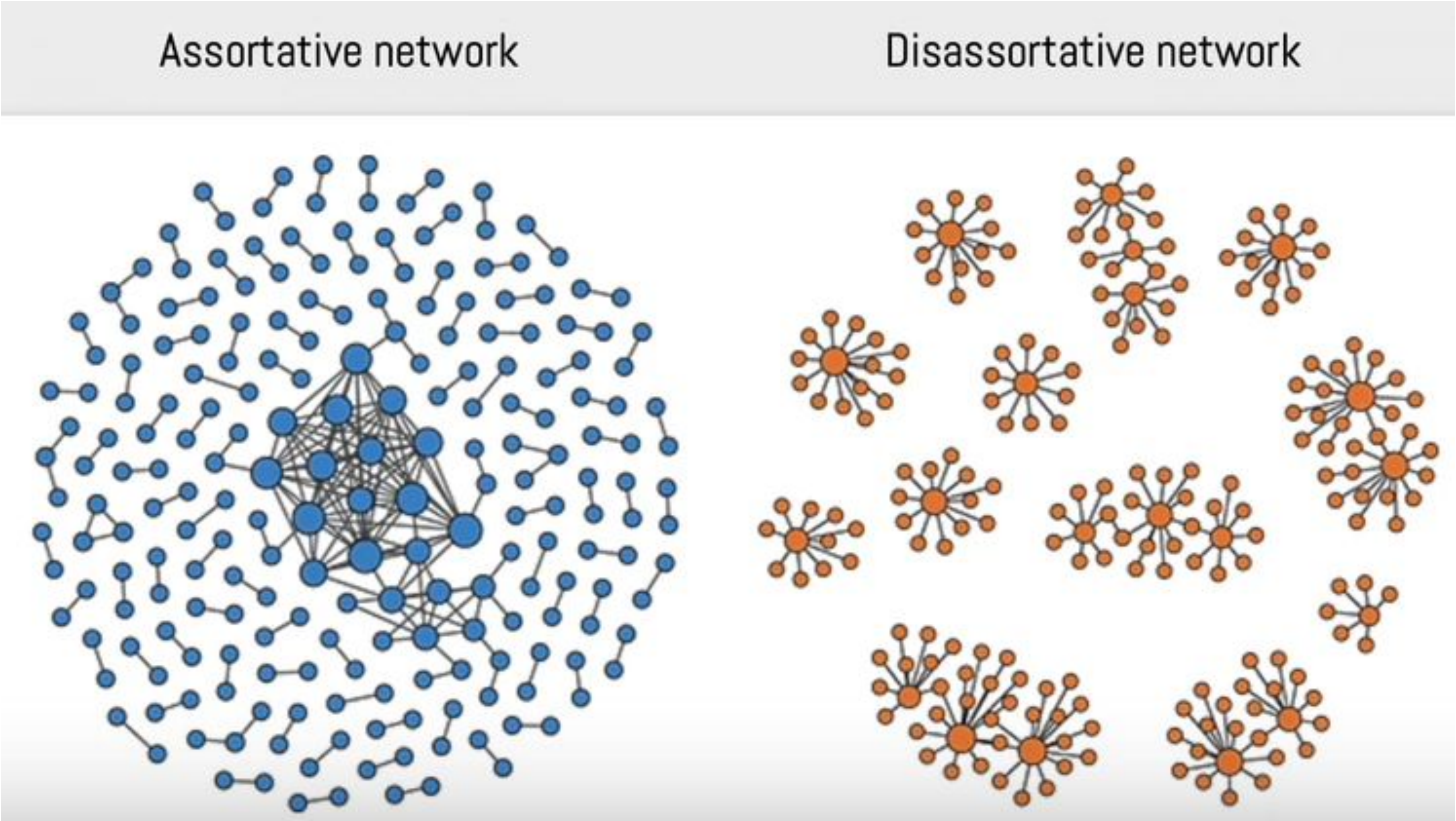


Baixo P => Distância média <L> pequena e coeficiente de clusterização <C> alto

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
C. elegans	2.65	2.25	0.28	0.05

D J Watts & S H Strogatz. *Collective Dynamics of "Smallworld" Networks*. Nature, 393, 1998



Degree assortativity / Degree correlation

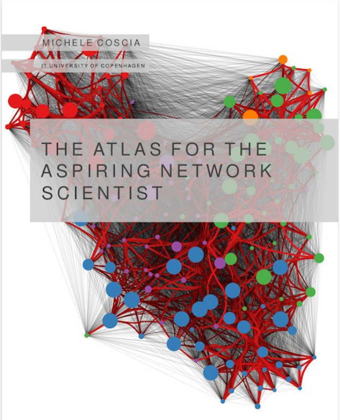
Referências

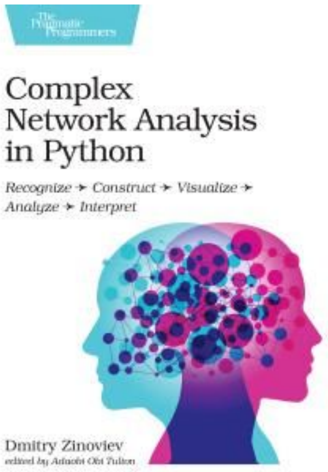
- Network Science by Albert-László Barabási
  http://networksciencebook.com/
- Network Analysis Course - Prof. Ivanovitch (DCA/UFRN) YouTube Playlist

metrópole DIGITAL

Referências

- The Atlas for the Aspiring Network Scientist, 2021
  - https://www.networkatlas.eu/index.htm
- Complex Network Analysis in Python, 2018
  - https://pragprog.com/titles/dzcnapy/complex-network-analysis-in-python/





Referências

- Caldarelli, Guido; Chessa, Alessandro. **Data science and complex networks - Real cases studies with Python**, 2016.

- Filippo Menczer; Santo Fortunato; Clayton A. Davis. **A First Course in Network Science**, 2020.