CIÊNCIA DE DADOS (BIG DATA PROCESSING AND ANALYTICS)

RECUPERAÇÃO DA INFORMAÇÃO NA WEB E EM REDES SOCIAIS



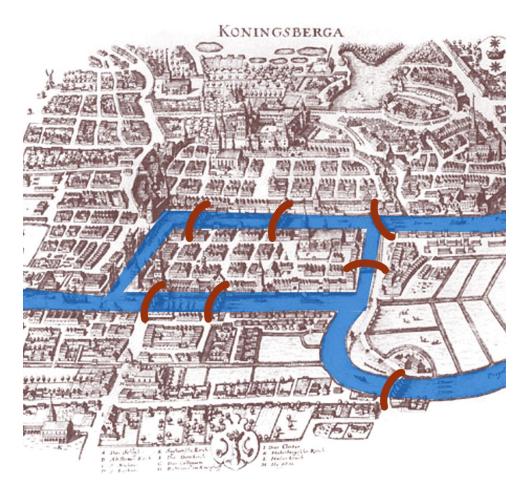


TRILHA 7 REDES COMPLEXAS INTRODUÇÃO À TEORIA DOS GRAFOS

A origem

 Problema das Sete pontes de Conisberga (atual Kaliningrado).

 Para resolvê-lo, Leonhard Paul Euler formulou a Teoria dos Grafos.



Fonte: Wikipedia, com arte do professor.

Teor

en tica que estuda a relação

5

• Grassa form dos por **nós e arestas**.

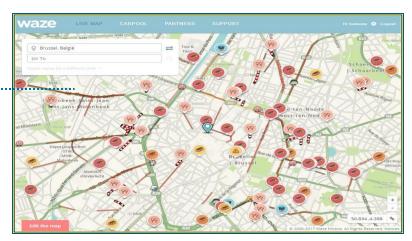
2 7

 Pe nite uma visio abstrata de um problema e é útil em siste nas complexos.

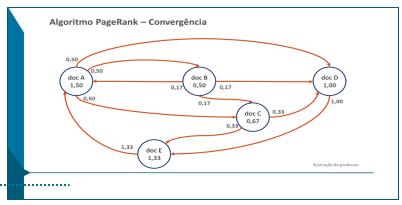
3

Por que estudar Grafos?

- Caminho mais curto.
- Caixeiro viajante.
- Inspeção de rota.
- Redes sociais.
- Disseminação de vírus.
- Propagação de fake news.
- Ranqueamento de páginas.



Fonte: Waze



Fonte: slide da Trilha 1 – Algoritmo pageRank.

Tipos de Grafos Grafos os • Twitter Não direcionados • Facebook Fonte: Elabo Fonte: Elaborada pelo autor.

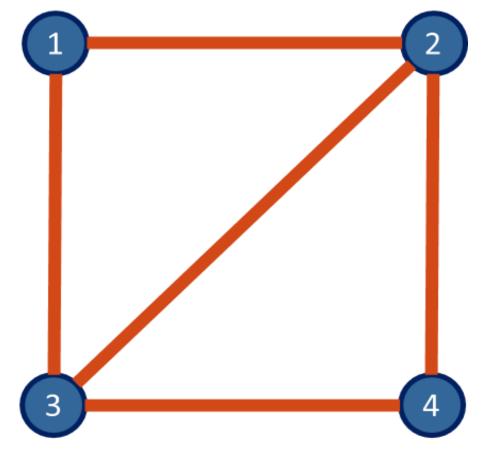
Representação Algébrica

Matriz de Adjacências

 $egin{bmatrix} 0 & 1 & 1 & 0 \ 1 & 0 & 1 & 1 \ 1 & 1 & 0 & 1 \ 0 & 1 & 1 & 0 \end{bmatrix}$

1 onde tem ligação • 0 onde não tem

Nota: em dígrafos, preenche-se apenas a linha da conexão.



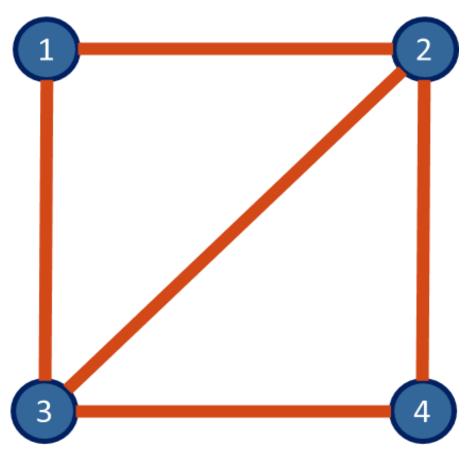
Grau (ou valência) dos nós

Dois nós são considerados vizinhos

 (adjacentes) se há uma aresta os conectando.

 O grau de um nó é medido pela quantidade de arestas que o mesmo possuí.

- Graus da figura ao lado
 - Nó 1: 2
 - Nó 2: 3
 - Nó 3: 3
 - Nó 4: 2



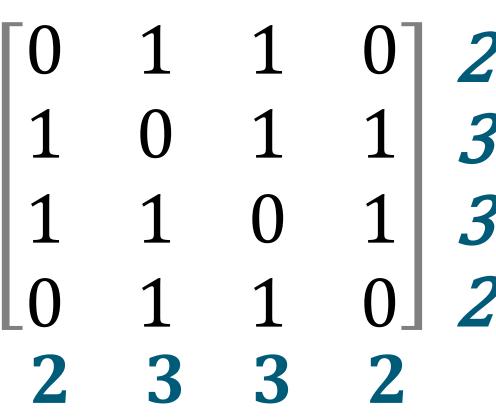
Grau (ou valência) dos nós

Dois nós são considerados vizinhos

 (adjacentes) se há uma aresta os conectando.

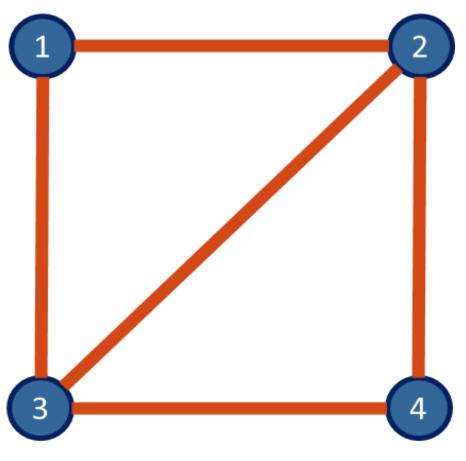
• O grau de um nó é medido pela quantidade de arestas que o mesmo possuí.

- Graus da figura ao lado
 - Nó 1: 2
 - Nó 2: 3
 - Nó 3: 3
 - Nó 4: 2



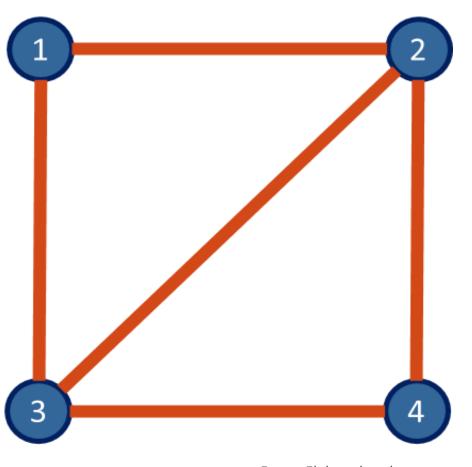
(Comprimento de um) caminho

- Medido pela quantidade de arestas mínimas necessária para conectar dois pontos de um grafo conexo.
- Exemplo: comprimento de um caminho do nó 1 ao nó 4 tem distancia 2:
 - 1 para 2 para 4; ou ainda,
 - 1 para 3 para 4.
- Em alguns casos, pode ser útil colocar peso nas arestas (por exemplo, distância entre cidades).



Diâmetro

- O diâmetro de um grafo é a maior distância entre dois pontos pelo menor caminho.
- Exemplo: para a rede ao lado, temos as seguintes distancias:
 - C(1,2): 1
 C(2,3): 1
 C(2,4): 1
 C(3,4): 1
- Neste caso, o diâmetro é igual comprimento do nó 1 e 4.



Centralidades

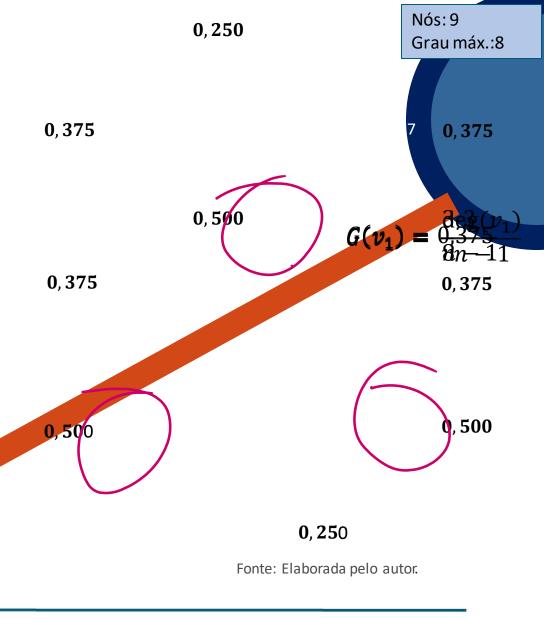
- Medida relativa que busca encontrar os nós mais importantes da rede.
- Pode ser mensurado de diversas formas, sendo as mais comuns:
 - Grau.
 - Proximidade.
 - Intermediação.

Centralidades: grau

- Primeira medida de centralidade aplicada a grafos.
- Simples e intuitivo.
- Definido como a quantidade de arestas incidentes cada nó possui dividido pelo valor máximo teórico (n-1):

$$C(v_1) = \frac{\deg(v_1)}{n-1}$$

 Divisão é importante como mecanismo de normalização em comparação de grafos de tamanho distintos.



Centralidades: proximidade

- Útil na medida da dispersão de informação pelos nós, como no caso de contágio de vírus ou dispersão de noticias nas redes sociais.
- Ordena os nós baseado na menor distância entre eles.

$$C(v_1) = \frac{n-1}{\sum_n d(v_n, v_1)}$$



- $d(v_n, v_1) = \{2: 1, 3: 2, 4: 1, 5: 2, 6: 1, 7: 2, 8: 2, 9: 3\}$
- $\sum_{n} d(v_n, v_1)$: 14
- $C(v_1) = 8 / 14 = 0.57$



0,35

0,57

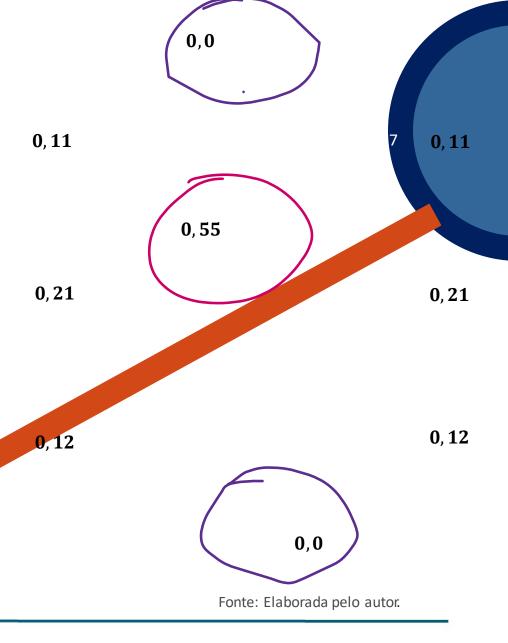
0,50

0,36

0,57

Centralidades: intermediação

- Identifica os nós mais importantes no fluxo da informação.
- Definida como a quantidade de vezes em que um nó i foi utilizado no menor caminho entre dois nós te z.
- Requer alto poder computacional e, muitas vezes, é estimado por aproximação e amostragem.



NetworkX

- NetworkX é um pacote Python para a criação, manipulação e estudo da estrutura, dinâmica e funções de redes complexas.
- Diversos algoritmos existentes.
- Utilizado em diversos casos de estudo de redes complexas encontrados na literatura.



Fonte: https://networkx.org/>.

NetworkX - Carregando o grafo

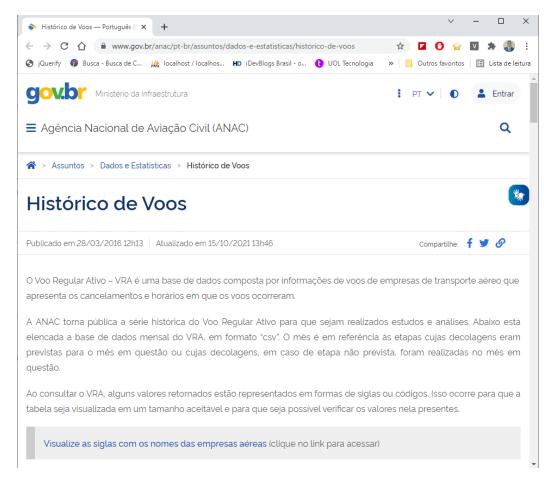
```
1. import networkx as nx
2. import matplotlib.pyplot as plt
3.
4. #inicia o grafo
5. G3 = nx.Graph()
6.
7. #cria um dicionário com as posições [opcional, apenas para visualizão]
8. pos = \{\}
9. pos[1] = (1000, 1500); pos[2] = (1000, 500); pos[3] = (500, 0);
10. pos[4] = (0, 500); pos[5] = (0, 1500); pos[6] = (500, 2000);
11. pos[7] = (1000, 2500); pos[8] = (0, 2500); pos[9] = (500, 3500)
12.
13. #Adiciona as arestas pela referência aos nós.
14. G3.add edge(1, 2);G3.add edge(1, 4);G3.add edge(1, 6)
15. G3.add edge(2, 5);G3.add edge(2, 3);G3.add edge(2, 4)
16. G3.add_edge(3, 4);G3.add_edge(4, 5);G3.add_edge(5, 6)
17. G3.add_edge(6, 7);G3.add_edge(6, 8);G3.add_edge(7, 8)
18. G3.add edge(7, 9);G3.add edge(8, 9)
19.
20. #desenha o grafo
21. plt.figure(figsize = (14,9))
22. nx.draw networkx(G3, pos)
```

NetworkX – Todos os cálculos dessa trilha

```
1. #cálculo dos nós
print("Graus dos nós", nx.degree(G3))
3.
   #Calculo de distancia
5. print("\n\n----\n\n")
6. for i in range(1, 10):
     print(f"A menor distancia entre os nós {i} será")
   for j in range(1, 10):
9.
      print(f" nó {j}:{nx.shortest path length(G3, i, j)}")
10.
11. '''Para imprimir todos de uma vez use os comandos abaixo'''
12. # print("\n\nDistancia entre todos os nos")
13. # print(list(nx.shortest path length(G3)))
14.
15. #Calculo de centralidade
16. print("\n\n----\n\n")
17. print("Centralidade por grau:", nx.degree centrality(G3))
18. print("Centralidade por proximidade:", nx.closeness centrality(G3))
19. print("Centralidade por intermediação:", nx.betweenness centrality(G3))
```

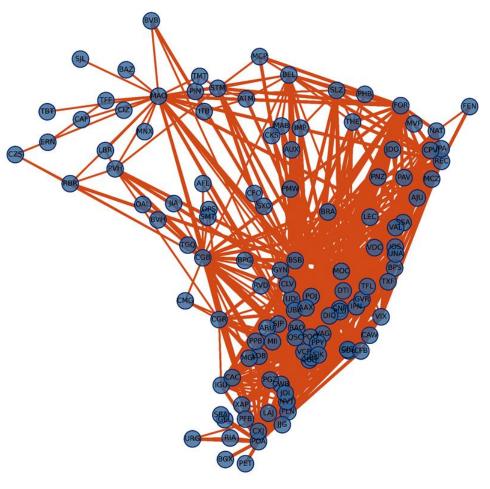
Caso de uso – Tarefa

- A Agencia Nacional de Aviação Civil (ANAC) disponibiliza todos os voos realizados no país.
- Um voo entre duas cidades pode ser visto como uma aresta entre dois aeroportos.
- A partir dos dados da ANAC, é
 possível elaborar um grafo com os
 voos brasileiros e responder a
 algumas perguntas.



Caso de uso – Tarefa

- Monte o grafo a partir da base compilada pelo professor para o ano de 2018 e responda às perguntas:
 - Qual é o grau médio, ou seja, o número de outros aeroportos conectados, dos aeroportos brasileiros com voos em 2018?
 - Qual aeroporto brasileiro voava, em 2018, para mais cidades distintas?
 - Quais são os dois aeroportos que precisavam, em 2018, de mais escalas para se interligarem?



Grafo elaborado pelo professor com dados do VRA 2018 da ANAC.



