



## Universidade Federal de Pernambuco Centro de Informática - CIn

Disciplina: Algoritmos e Estruturas de Dados Docente: Sérgio Ricardo de Melo Queiroz Discente: Leonardo Bezerra de Oliveira

# Relatório do projeto

#### Contexto do problema

Diante da minha escolha do Algoritmo de Prim para realização desse projeto, tentei buscar uma base de dados que fizesse sentido para mim tendo em vista a utilização desse algoritmo, que é baseado na busca pelo menor caminho através das arestas e vértices. Pensando bastante nesse assunto, acabei tendo a ideia de utilizar a distância entre países, estados, cidades, por ser algo mais palpável, utilizaria as distâncias como os pesos e os nomes das cidades como os vértices para criar o grafo baseado nesse algoritmo.

Pesquisando bases de dados tive dificuldade de encontrar esse tipo de base em sites como o kaggle, que usei de exemplo por ser um bastante conhecido, então resolvi procurar no Google mesmo e encontrei a base de dados fornecida pela PUCRS que indica a distância entre algumas cidades e capitais.

A base de dados utilizada conta com a distância rodoviária entre algumas das principais cidades brasileiras, como é o caso por exemplo da distância entre Anápolis, a cidade do estado de Goiás e a cidade de Campinas em São Paulo.

## Implementação

## Algoritmo utilizado.

O algoritmo escolhido foi o algoritmo de Prim. A princípio tinha a ideia de utilizar o algoritmo de Dijkstra, mas já não haviam vagas disponíveis para utilização desse na planilha de indicação. Julgando os processos do algoritmo de Bellman-Ford, pensei em optar por esse, mas resolvi pesquisar mais sobre os outros algoritmos disponíveis e mudei de ideia e acabei optando por realizar o projeto com o algoritmo de Prim, pois parecia interessante e para mim fazia mais sentido.

## Desenvolvimento.

Primeiramente resolvi criar um planejamento para que fossem bem divididas as atividades que ia realizar de acordo com os dias e a minha rotina. Optei por estar fazendo o trabalho individualmente para que pudesse aprender o máximo possível, tendo em vista que não tive tanto contato com esse tipo de algoritmo e estava com dificuldades em relação às listas e criar os códigos em geral, achei que seria uma oportunidade de me desenvolver, tendo em vista também o uso de bibliotecas e a organização possível de dados que esse algoritmo geraria, que traria bons aprendizados para mim.

Mesmo criando esse planejamento, tive algumas dificuldades, até mesmo para seguir o planejamento, por conta de motivos pessoais acabei tendo que reduzir meu escopo e realizar mais atividades nos dias que estivesse livre, para adiantar o processo, tendo em vista que em alguns dos dias não consegui uma grande ou alguma evolução do processo, realizar o projeto sozinho também se torna mais complexo.

Para iniciar o processo criei primeiramente o algoritmo de prim básico, para que pudesse ter uma ideia do que iria fazer a seguir. Após ter encontrado a base de dados que me agradou e estava dentro do esperado em relação a quantidade de vértices e arestas, comecei na tentativa de interligar essa base ao algoritmo para que o grafo esperado fosse gerado, tive bastante dificuldade nesse sentido.

Como já tinha utilizado outras vezes o pandas e achava uma ferramenta mais simples de entender, decidi utilizá-lo, então fiz inicialmente a ligação com o link online mesmo, pensando eu que daria certo, mas não funcionou.

```
import pandas as pd
url = "https://www.inf.pucrs.br/~danielc/peng1a/outros/distancias.xls"
df = pd.read_excel(url)
vertices = list(df.columns)
graph = {vertex: {} for vertex in vertices}

como não deu certo, alterei a forma de chamada da base:

import pandas as pd
file_path = '/content/distancias.xls'
df = pd.read_excel(file_path)

e por fim:

import pandas as pd
file_path = 'distancias.xls'
df = pd.read_excel(file_path, header=None, skiprows=2)
```

A implementação se inicia com a chamada das bibliotecas utilizadas, pandas, networkx e matplotlib.

import, início, utilização do pandas para ligar o código com a base de dados.

```
import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt

file_path = 'distancias.xls'

file_path = 'distancias.xls'

wertices = df.iloc[:, 0].tolist()

graph = {vertex: {} for vertex in vertices}
```

Preenchimento do grafo, alterações

```
12
       for i, row in enumerate(df.iterrows(), start=1):
          vertex1 = vertices[i - 1]
 13
          for j, distance in enumerate(row[1].tolist(), start=1):
 14
              if pd.notna(distance) and isinstance(distance, (int, float)) and distance > 0:
 15
                  vertex2 = vertices[j - 1]
 16
                  graph[vertex1][vertex2] = distance
 17
                  graph[vertex2][vertex1] = distance
 18
 19
       for vertex, neighbors in graph.items():
 20
          print(f"{vertex} -> {neighbors}")
 21
          número
                                  vértices
print
                         de
                                                        arestar
                                                                      grafo
                                                е
                                                                                 geral
  24
         num_vertices = len(graph)
  25
         num arestas = sum(len(neighbors) for neighbors in graph.values())
  26
         print(f"Número de vértices no Grafo: {num_vertices}")
         print(f"Número de arestas no Grafo: {num_arestas}")
  27
algoritmo
                                            de
                                                                                  prim
  29
         def prim(graph):
  30
              start_vertex = vertices[0]
  31
              visited = {vertex: False for vertex in vertices}
              distances = {vertex: float('inf') for vertex in vertices}
  32
  33
              parent = {vertex: None for vertex in vertices}
  34
              distances[start vertex] = 0
  35
  36
              for in range(len(vertices)):
  37
  38
                  current_vertex = find_min_vertex(distances, visited)
  39
                  visited[current vertex] = True
  40
  41
                  for neighbor, weight in graph[current_vertex].items():
  42
                      if not visited[neighbor] and weight < distances[neighbor]:</pre>
  43
                           distances[neighbor] = weight
                           parent[neighbor] = current_vertex
  44
  45
              mst = {vertex: {} for vertex in vertices}
  46
              for vertex, par in parent.items():
  47
                  if par is not None:
  48
                      mst[vertex][par] = distances[vertex]
  49
                      mst[par][vertex] = distances[vertex]
  50
  51
  52
              return mst
```

```
gerar
                                       árvore
                                                                                mínima:
 55
         def find_min_vertex(distances, visited):
 56
             min_vertex = None
 57
             min distance = float('inf')
 58
 59
             for vertex in distances:
                 if not visited[vertex] and distances[vertex] < min_distance:</pre>
 60
 61
                      min_vertex = vertex
                      min_distance = distances[vertex]
 62
 63
 64
             return min_vertex
 65
Criar grafo no NetworkX, printar árvore mínima e distância, print da quantidade de
                                                              árvore
                                                                                mínima:
vértices
                              arestas
                                                 da
 66
        G = nx.Graph()
 67
        for vertex, neighbors in graph.items():
           for neighbor, weight in neighbors.items():
 68
               G.add_edge(vertex, neighbor, weight=weight)
 69
 70
 71
        minimum_spanning_tree = prim(graph)
 72
 73
 74
        for vertex, neighbors in minimum_spanning_tree.items():
 75
           for neighbor, weight in neighbors.items():
 76
               print(f"{vertex} - {neighbor}: {weight}")
 77
  78
        num vertices = len(minimum spanning tree)
 79
        num_arestas = sum(len(neighbors) for neighbors in minimum_spanning_tree.values()) // 2
        print(f"Número de vértices na Árvore Geradora Mínima: {num_vertices}")
 80
        print(f"Número de arestas na Árvore Geradora Mínima: {num_arestas}")
 81
 22
Criação do grafo final do NetworkX, salvando o grafo como arquivo:
          G = nx.Graph()
  83
          for vertex, neighbors in graph.items():
  84
              for neighbor, weight in neighbors.items():
  85
                  G.add_edge(vertex, neighbor, weight=weight)
  86
  87
          pos = nx.spring_layout(G)
  88
          nx.draw(G, pos, with labels=True, font weight='bold')
  89
          edge_labels = {(u, v): d['weight'] for u, v, d in G.edges(data=True)}
  90
          nx.draw_networkx_edge_labels(G, pos, edge_labels=edge_labels)
  91
  92
          plt.show()
```

No começo não sabia bem como dividir para que as colunas e linhas que constavam os nomes das cidades não entrassem como valores das distâncias, para que somente a partir de uma linha fosse contabilizado isso. Pesquisando mais sobre a biblioteca utilizada e usando de seus facilitadores, depois de incontáveis tentativas, consegui um resultado mediano, ainda havia erros, mas ao menos as cidades estavam sendo "printadas", o que me deu alguma esperança.

| ₽ | 0<br>1<br>2<br>3<br>4 | QUADRO DE DISTÂNCIAS RODOVIÁRIAS ENTRE AS PRINCIPAIS CIDADES BRASILEIRAS \ NAN AMERICANA ANÁPOLIS ARACAJÚ ARAÇATUBA |             |               |              |                          |                            |        |  |  |
|---|-----------------------|---|-------------|---------------|--------------|--------------------------|----------------------------|--------|--|--|
|   |                       | Unnamed: 1 U  | nnamed: 2   | Unnamed:      | 3 Unnamed: 4 | 1 Unnamed: 5             | Unnamed: 6                 | \      |  |  |
|   | 0                     | ARACAJÚ   |             | BELO HORIZONT | E BOA VISTA  | A BRASÍLIA               | CAMPINAS                   | •      |  |  |
|   | 1                     | 2216  | 2808        | 59            | 5 4633       | 887                      | 42                         |        |  |  |
|   | 2                     | 1885  | 1964        | 87            | 0 412        | 3 160                    | 882                        |        |  |  |
|   | 3                     | NaN   | 2079        | 157           | '8 600¢      | 1652                     | 2182                       |        |  |  |
|   | 4                     | 2330  | 2662        | 85            | 4487         | 7 854                    | 456                        |        |  |  |
|   | 0                     | Unnamed:<br>CAMPO GRAND   |             | ed: 8 Unnamed |              | named: 28<br>IO BRANCO R | Unnamed: 2<br>IO DE JANEIR | - '    |  |  |
|   | 1                     | 97  | 9           | 1084          |              | 3479                     | 54                         | 5      |  |  |
|   | 2                     | 98  | 2           | 1767          | 981          | 2971                     | 129                        | 6      |  |  |
|   | 3                     | 276   | 5           | 3169 2        | 775          | 4763                     | 185                        | 5      |  |  |
|   | 4                     | 49  | 0           | 1187          | .343         | 3333                     | 94                         | 2      |  |  |
|   |                       | Unnamed: 30   | Unnamed: 31 | . Unnamed:    | 32 Unnamed   | : 33 Unnamed             | : 34 Unnamed               | : 35 \ |  |  |
|   | 0                     | SALVADOR  | SANTOS      | S.J.DOS CAN   | IPOS SÃO L   | LUÍS SÃO P               | AULO TERE                  | SINA   |  |  |
|   | 1                     | 1971  | 205         |               | 208          | 2845                     | 133                        | 2664   |  |  |
|   | 2                     | 1679  | 1045        | 1             | .052         | 2005                     | 973                        | 1937   |  |  |
|   | 3                     | 356   | 2249        | 2             | 1086         | L578                     | 2187                       | 1142   |  |  |
|   | 4                     | 2130  | 604         |               | 595          | 2699                     | 532                        | 2631   |  |  |

Nesse momento percebi que havia alguns problemas em relação aos dados que não constavam na tabela (distâncias vazias) da base de dados. Observando a imagem acima também pude perceber o Unnamed que não fazia sentido com o resultado esperado, que na primeira linha do print consta a primeira linha do nome da tabela, além de "NaN" como nome do índice 0, o que não deveria acontecer, por esse ser na tabela somente uma imagem.

Então fiz alterações no código, para que ele excluísse as linhas que indicavam nomes na tabela da contagem das distâncias e onde o espaço estivesse vazio não fosse contabilizado a distância. Além de alterações no final do código, então foi possível visualizar a ligação entre os vértices e as distâncias(pesos das arestas).

Para a criação de interface utilizei as bibliotecas NetworkX e Matplotlib, pesquisando consegui entender melhor a funcionalidade e pude gerar o grafo através do próprio código. Para ter outra visualização do grafo coloquei ele no aplicativo Gephi e obtive uma melhor imagem, trazida na conclusão.

#### Bibliotecas utilizadas.

Pandas, <a href="https://pandas.pydata.org/">https://pandas.pydata.org/">https://pandas.pydata.org/</a>;
NetworkX <a href="https://networkx.org/">https://networkx.org/</a>>
Matplotlib <a href="https://matplotlib.org/">https://matplotlib.org/</a>>

## Apps utilizados.

Gephi para gerar interface mais visualizável. <a href="https://gephi.org/">https://gephi.org/</a>>

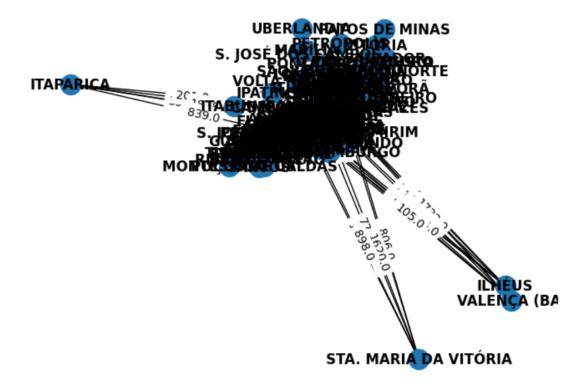
#### Conclusão

Através desse projeto pude perceber como é importante ter um planejamento de desenvolvimento, tive algumas dificuldades porque não consegui seguir bem o planejamento que tinha feito, por conta de alguns contratempos. Foi interessante o projeto também para aprender um pouco mais sobre o pandas, além do uso do VSCode e do github em si.

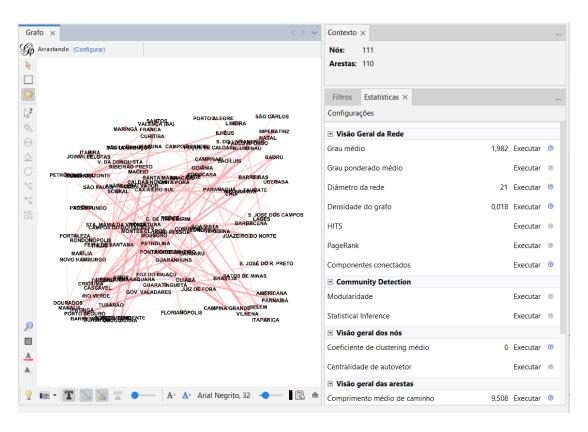
O algoritmo de prim é interessante para fazer esse tipo de filtragem, analisando os pesos e na questão escolhida faz muito sentido, tendo em vista que se pode escolher, por exemplo, que cidades passar em uma viagem para economizar tempo ou a depender do que foi objetificado pelo viajante.

Percebi durante esse projeto também que um dos meios mais interessantes para me auxiliar foi o Youtube. Assisti vídeos e algo interessante foi como o vídeo mais curto que assisti foi o que me trouxe mais informação sobre o algoritmo e me fez entender a sua funcionalidade. O vídeo citado é do canal no youtube "Papa Thalyson", intitulado Algoritmo Prim em 3 minutos - Ideia e exemplos, além desse outros me trouxeram muito auxílio em relação ao projeto, para entender as funcionalidades das bibliotecas, por exemplo.

A questão da interface gráfica também parecia complicada para mim, mas utilizei o Gephi e as bibliotecas network e matplotlib.pyplot para que fosse gerado o grafo mostrado a seguir:



Utilizando o aplicativo próprio do Gephi, com o arquivo gerado através das bibliotecas citadas, pude gerar outra imagem do grafo, que ao meu ver é mais simples e tem uma visualização melhor em relação à anterior.



A funcionalidade do programa é bem clara, do meu ponto de vista, no início fiz o import do pandas, a biblioteca que utilizei para fazer a ligação do banco de dados com o

código. Além disso, utilizei essa biblioteca para que fosse possível organizar os dados que seriam usados da tabela, por exemplo a partir de qual linha/coluna deveria ser contabilizado os números(por a tabela contar com nomes das cidades por escrito na vertical e na horizontal) e para organizar caso não houvessem valores de distância, pois algumas entre cidades não constavam na tabela.

Os outros imports tem relação com a geração final do grafo, utilizei o NetworkX e a Matplotlib, utilizando das informações trazidas no próprio site sobre essas ferramentas e as que encontrei em algumas das referências citadas.

Criei no início a lista de vértices, o grafo em si como dicionário e o seu preenchimento. Coloquei para serem imprimidas distâncias no grafo e coloquei também linhas para imprimir o número total de vértices e arestas do grafo, antes da geração mínima.

Em seguida está o código relacionado ao algoritmo de prim, mas em relação às ligações totais do grafo, em seguida fiz um caminho para a geração da árvore mínima, buscando o caminho com menor peso entre os vértices, fazendo comparações.

Em seguida constam códigos relacionados à imagem que seria gerada desse grafo, a geração da árvore mínima em si e os valores referentes às quantidades de vértices e arestas desse caminho mínimo.

## Prints de parte do output:

AMERICAMA -> ('AMERICAMA': 2216.0, 'ARACAJŪ': 2888.0, 'ARACAJŪBA': 595.0, 'ARACATUBA': 2330.0, 'ARARAQUARA': 2188.0, 'BARBETOS': 2426.0, 'BARBACEMA': 797.0, 'BARBETOS': 2886.0, 'BARBACEMA': 1755.0, 'ARACATUBA': 2300.0, 'ARACAJŪ-> ('AMERICAMA': 2888.0, 'AMACAJŪ-S': 1885.0, 'ARACAJŪ': 297.0, 'ARACATUBA': 2330.0, 'ARACAJŪ-> ('AMERICAMA': 2888.0, 'AMACAJŪ-S': 2888.0, 'ARACAJŪ-S': 2888.0,

FRRICA > (AMERICAMA': 133.9, 'AMAPOLIS': 2041.0, 'ARACAJU': 209.0, 'ARACAJURA': 469.0, 'ARARAQUARA': 4432.0, 'BAGE': 676.0, 'BARRETOS': 204.0, 'BA

Número de vértices no Grafo: 111 Número de arestas no Grafo: 6551

AMERICANA - BARRETOS: 42.0

ANÁPOLIS - MACEIÓ: 294.0

ANÁPOLIS - FEIRA DE SANTANA: 322.0

ANÁPOLIS - PAULO AFONSO: 307.0

ARACAJÚ - IMPERATRIZ: 609.0

ARACAJÚ - TUCURUÍ: 350.0

ARAÇATUBA - BARBACENA: 169.0

ARAÇATUBA - ARARAQUARA: 606.0

ARAÇATUBA - GOV. VALADARES: 324.0

ARAÇATUBA - IPATINGA: 217.0

ARACATUBA - ITABIRA: 99.0

ARAÇATUBA - MONTES CLAROS: 418.0

ARARAOUARA - ARACATUBA: 606.0

ARARAQUARA - MANAUS: 785.0

BAGÉ - GOIÂNIA: 209.0

BAGÉ - STA. MARIA DA VITÓRIA: 597.0

BARRETOS - AMERICANA: 42.0

BARRETOS - LIMEIRA: 65.0

BARRETOS - POÇOS DE CALDAS: 150.0

BARRETOS - RIBEIRÃO PRETO: 238.0

BARRETOS - SÃO PAULO: 99.0

BARRETOS - SOROCABA: 85.0

BARBACENA - ARAÇATUBA: 169.0

BARBACENA - DOURADOS: 224.0

BARBACENA - CORUMBÁ: 403.0

BARBACENA - PONTA PORÃ: 335.0

BARREIRAS - LAGES: 226.0

BARREIRAS - NOVO HAMBURGO: 94.0

BARREIRAS - PASSO FUNDO: 226.0

SANTOS - FRANCA: 72.0

SÃO CARLOS - CUIABÁ: 116.0

S. JOSÉ DO R. PRETO - CUIABÁ: 189.0

S. JOSÉ DOS CAMPOS - FRANCA: 97.0

SÃO LUÍS - GUARANHUNS: 446.0

SÃO LOURENÇO - CAMPINA GRANDE: 208.0

SÃO LOURENÇO - FORTALEZA: 182.0

SÃO PAULO - FLORIANÓPOLIS: 72.0

SÃO PAULO - FORTALEZA: 97.0

SÃO PAULO - BARRETOS: 99.0

STA. MARIA DA VITÓRIA - BAGÉ: 597.0

SINOP - BAURU: 488.0

SOBRAL - BOA VISTA: 238.0

SOROCABA - FRANCA: 87.0

SOROCABA - BARRETOS: 85.0

TAUBATÉ - FORTALEZA: 43.0

TERESINA - FOZ DO IGUAÇÚ: 446.0

TUBARÃO - BLUMENAU: 144.0

TUCURUÍ - ARACAJÚ: 350.0

UBERABA - GOIÂNIA: 113.0

UBERABA - CUIABÁ: 178.0

UBERLANDIA - CUIABÁ: 281.0

URUGUAIANA - CAXIAS DO SUL: 604.0

VALENCA (BA) - BELO HORIZONTE: 194.0

VALENÇA (BA) - FEIRA DE SANTANA: 105.0

VILHENA - CORUMBÁ: 704.0

VITÓRIA - CAMPINA GRANDE: 519.0

V. DA CONQUISTA - BELO HORIZONTE: 411.0

VOLTA REDONDA - DOURADOS: 112.0

Número de vértices na Árvore Geradora Mínima: 111 Número de arestas na Árvore Geradora Mínima: 110

As características da tabela me trouxeram algumas dificuldades, tendo em vista essas especificidades da posição que deveria começar o grafo, por exemplo, tendo que especificar linha e coluna, além da contabilização dos pesos (as distâncias entre as cidades, as arestas), o início mesmo do grafo e as ligações entre as cidades(vértices).

Mesmo com todos os problemas que tive durante o processo consegui fazer essa ligação entre as cidades e organizar as distâncias em si, até para ter uma visualização melhor e pensei na possibilidade que esse resultado poderia trazer caso fosse fazer algum tipo de análise de dados, o que faz sentido tendo em vista o uso das funcionalidades da biblioteca pandas dentro dessa área.

#### Referências

Guia Básico de Git - Como fazer commit no GitHub. Blog Coder 2021 , disponível em: <a href="https://blog.cod3r.com.br/quia-basico-de-git-como-fazer-commit-no-github/">https://blog.cod3r.com.br/quia-basico-de-git-como-fazer-commit-no-github/</a>

Introdução Pandas Python. Hashtag Treinamentos. 2021, disponível em: <a href="https://www.hashtagtreinamentos.com/introducao-pandas-python">https://www.hashtagtreinamentos.com/introducao-pandas-python</a>>

Pandas o que é e para quê serve, Alura. 2021, dispoível em:<<u>https://www.alura.com.br/artigos/pandas-o-que-e-para-que-serve-como-instalar</u>>

Gephi, guia básico, disponível em:<<a href="https://gephi.org/users/">https://gephi.org/users/</a>>

O que é commit e como usar commits semânticos, Geek Hunter, disponível em: <a href="https://blog.geekhunter.com.br/o-que-e-commit-e-como-usar-commits-semanticos/">https://blog.geekhunter.com.br/o-que-e-commit-e-como-usar-commits-semanticos/</a>

Algoritmo de Prim em 3 minutos - Ideia e exemplos, Papa Thalyson, 2021. disponível em: <a href="https://www.youtube.com/watch?v="xxgMkLSNpg&ab channel=PapaThalyson">https://www.youtube.com/watch?v="xxgMkLSNpg&ab channel=PapaaThalyson">https://www.youtube.com/watch?v="xxgMkLSNpg&ab channel=PapaaThalyson">https://www.yout

Algoritmo de Prim - Teoria dos grafos (Exemplos práticos), Industrial 21, 2021. disponível em: <a href="https://www.youtube.com/watch?v=Pgf-AL1vYb4&ab\_channel=INDUSTRIAL21">https://www.youtube.com/watch?v=Pgf-AL1vYb4&ab\_channel=INDUSTRIAL21</a>

Manipulação de grafos em Python com NetworkX, Professor Alexandre Nevada, disponível em:

<a href="https://www.youtube.com/watch?v=PfT8">https://www.youtube.com/watch?v=PfT8</a> 2sKReo&ab channel=Prof.AlexandreLevada>

Como sair do ZERO em Gráficos no Python [Matplotlib], Hashtag Programação, 2021, disponível em: <a href="https://www.youtube.com/watch?v=FDU-D8ddTU4&ab\_channel=HashtagPrograma%C3%A7%C3%A3o">https://www.youtube.com/watch?v=FDU-D8ddTU4&ab\_channel=HashtagPrograma%C3%A7%C3%A3o</a>>

Python - Introdução a Biblioteca Matplotlib e Numpy pythonaltas, 2022, disponível em: <a href="https://www.youtube.com/watch?v=QLACZgfUAhY&ab\_channel=Pythonaltas">https://www.youtube.com/watch?v=QLACZgfUAhY&ab\_channel=Pythonaltas</a>>

Algoritmo de Prim - Árvores Geradoras Mínimas - Algoritmos em Grafos, aulas de computação, 2021, disponível em: <a href="https://www.youtube.com/watch?v=vXdtCJmkoEU&ab\_channel=AulasdeComputa%C3%A7%C3%A3o">https://www.youtube.com/watch?v=vXdtCJmkoEU&ab\_channel=AulasdeComputa%C3%A7%C3%A3o</a>>

Gephi e análise baseada em grafos, André Santachè, 2015, disponível em: <a href="https://www.youtube.com/watch?v=4tAOA7Skqtw&t=204s&ab\_channel=Andr%C3%A9Santanch%C3%A8">https://www.youtube.com/watch?v=4tAOA7Skqtw&t=204s&ab\_channel=Andr%C3%A9Santanch%C3%A8></a>

Grafos com Python - Matriz de Adjacências - Aula 15 de Grafos, Professor Douglas Maioli, disponível em: <a href="https://www.youtube.com/watch?v=lzG9l">https://www.youtube.com/watch?v=lzG9l</a> 7GaZM&ab channel=ProfessorDouglasMaioli>