Análise de Complexidade de Tempo do Método Merge Sort

Eduardo Costa de Paiva
eduardocspv@gmail.com
Frederico Franco Calhau
fredericoffc@gmail.com
Gabriel Augusto Marson
gabrielmarson@live.com

Faculdade de Computação
Universidade Federal de Uberlândia

18 de dezembro de 2015

Lista de Figuras

Lista de Tabelas

Lista de Listagens

Sumário

Capítulo 1

Introdução

Este documento foi feito com o intuito de exibir uma análise do algoritmo Merge Sort com relação a tempo. Além disso, será feita uma comparação da curva de tempo do que se espera do algoritmo, ou seja, $O(nlgn)$ com o caso prático.

1.1 Diretório

Dada a seguinte organização das pastas, utilizamos o arquivo testdriver.py, executando, uma função conveniente por vez. Para mais informações vá até ao apêndice.

OBS.: É necessário instalar o programa tree pelo terminal. Isso pode ser feito da seguinte maneira.

```
> sudo apt-get install tree
```

A seguir é mostrada a organização das pastas sendo que os diretórios significativas para o projeto são Codigos e Relatorio além do raíz:

```
tree --charset=ASCII -d
.
|-- Codigos
|   |-- Bubble
|   |   `-- __pycache__
|   |-- Bucket
|   |   `-- __pycache__
|   |-- Counting
|   |   `-- __pycache__
|   |-- Heap
|   |   `-- __pycache__
|   |-- Insertion
|   |   `-- __pycache__
|   |-- Merge
|   |   `-- __pycache__
|   |-- Quick
|   |   `-- __pycache__
|   `-- Selection
|       `-- __pycache__
|-- Other
```

```
|-- Plot
|-- __pycache__
`-- relatorio
    |-- imagens
    |   |-- Bubble
    |   |-- Bucket
    |   |-- Counting
    |   |-- Heap
    |   |-- Insertion
    |   |-- Merge
    |   |-- Quick
    |   |-- Radix
    |   `-- Selection
    |-- Relatorio_Bubble
    |-- Relatorio_Bucket
    |-- Relatorio_Counting
    |-- Relatorio_Heap
    |-- Relatorio_Insertion
    |-- Relatorio_Merge
    |-- Relatorio_Selection
    `-- Resultados
        |-- Bubble
        |-- Bucket
        |-- Counting
        |-- Heap
        |-- Insertion
        |-- Merge
        |-- Quick
        `-- Selection

47 directories
```

1.2 Códigos de programas

Seguem os códigos utilizados na análise de tempo do algoritmo Merge Sort.

1. MergeSort.py: Disponível na Listagem 1.1.

Listagem 1.1: MergeSort.py

```
1  import math
2  import sys
3  import gc
4
5  #sys.path.append('/home/gmarson/Git/AnaliseDeAlgoritmos/
       Trabalho_Final')
6  #from memoria import *
7
8  #@profile
9  def intercala(A,p,q,r):
10   #print("antes: ",  memory_usage_resource())
11   B = [0] *len(A)
12   for i in range(p,(q+1)):
13     B[i] = A[i]
14   for j in range(q+1,(r+1)):
```

```
15      B[r+q+1-j] = A[j]
16
17    i = p
18    j = r
19    #print("depois", memory_usage_resource())
20
21    for k in range (p,(r+1)):
22      if(B[i] <= B[j]):
23        A[k] = B[i]
24        i = i+1
25      else:
26        A[k] = B[j]
27        j = j-1
28
29  @profile
30  def merge(A):
31    mergeSort(A,0,len(A)-1)
32
33    #return A
34
35
36  #@profile
37  def mergeSort(A,esquerda,direita):
38    if(esquerda<direita):
39      meio = math.floor((esquerda+direita)/2)
40      mergeSort(A,esquerda,meio)
41      mergeSort(A,meio+1,direita)
42      intercala(A,esquerda,meio,direita)
43
44
45
46  #if __name__ == '__main__':
47  # gc.disable()
48  # A = [i for i in range(10000)]
49  # merge(A)
50  # print(merge(A))
51  # gc.enable()
52
53  #A = [i for i in range(10000)]
54  #merge(A)
55  #Para criar uma lista preenchida com 0's e que possua tamanho de A
        basta
56  #print(merge(A))
```

2. testeGeneric.py Disponível na Listagem 1.2

Listagem 1.2: testeGeneric.py

```
1  ##adicionei - Serve para importar arquivos em outro diretório
2  ###  A CADA NOVO MÉTODO MUDAR O IMPORT,  A CHAMADA DA FUNÇÃO E O SYS.
       PATH
3
4  import sys
5  sys.path.append('/home/gmarson/Git/AnaliseDeAlgoritmos/Trabalho_Final
       /Codigos/Merge')
6  sys.path.append('/home/gmarson/Git/AnaliseDeAlgoritmos/Trabalho_Final
       ')
7
8  from monitor import *
```

```
from memoria import *

from MergeSort import *
import argparse

parser = argparse.ArgumentParser()
parser.add_argument("n", type=int, help="número de elementos no vetor
    de teste")
args = parser.parse_args()

v = criavet(args.n)
merge(v)



## A EXECUÇÃO DESSE ARQUIVO EH ASSIM
## NA LINHA DE COMANDO VC MANDA O NOME DO ARQUIVO E O TAMANHO DO
   ELEMNTO DO vetor
##EXEMPLO testeBubble.py 10
##ele gera um vetor aleatório (criavet) e manda pro bubble_sort
```

3. monitor.py Disponível na Listagem 1.3

Listagem 1.3: monitor.py

```
# Para instalar o Python 3 no Ubuntu 14 ou 15
#
# sudo apt-get install python3 python3-numpy python3-matplotlib
   ipython3 python3-psutil
#

from math import *
import gc
import random
import numpy as np


from tempo import *

# Vetores de teste
def troca(m,v,n): ## seleciona o nível de embaralhamento do vetor
    m = trunc(m)
    mi = (n-m)//2
    mf = (n+m)//2
    for num in range(mi,mf):
        i = np.random.randint(mi,mf)
        j = np.random.randint(mi,mf)
        #print("i= ", i, " j= ", j)
        t = v[i]
        v[i] = v[j]
        v[j] = t
    return v


def criavet(n, grau=-0.5, inf=-1000, sup=1000):
    passo = (sup - inf)/n
    if grau < 0.0:
        v = np.arange(sup, inf, -passo)
        if grau <= -1.0:
```

```
34              return v
35          else:
36              return troca(-grau*n, v, n)
37      elif grau > 0.0:
38          v = np.arange(inf, sup, passo)
39          if grau >= 1.0:
40              return v
41          else:
42              return troca(grau*n, v, n)
43      else:
44          return np.random.randint(inf, sup, size=n)
45          #return [random.random() for i in range(n)] # for bucket sort
46
47
48
49  #print(criavet(20))
50
51  #Tipo                                    grau
52  #aleatorio                               0
53  #ordenado_crescente                      1
54  #ordenado_decrescente                    -1
55  #parcialmente_ordenado_crescente         0.5
56  #parcialmente_ordenado_decrescente       -0.5
57
58
59  def executa(fn, v):
60      gc.disable()
61      with Tempo(True) as tempo:
62          fn(v)
63      gc.enable()
```

4. testdriver.py Referenciado no apêndice A.

Capítulo 2

Gráficos

Seguem os Gráficos utilizadas no processo de análise do método Merge Sort:

1. Para um vetor aleatório
   (a) Complexidade de custo do método Merge Sort disponível na lista de imagens 2.1.

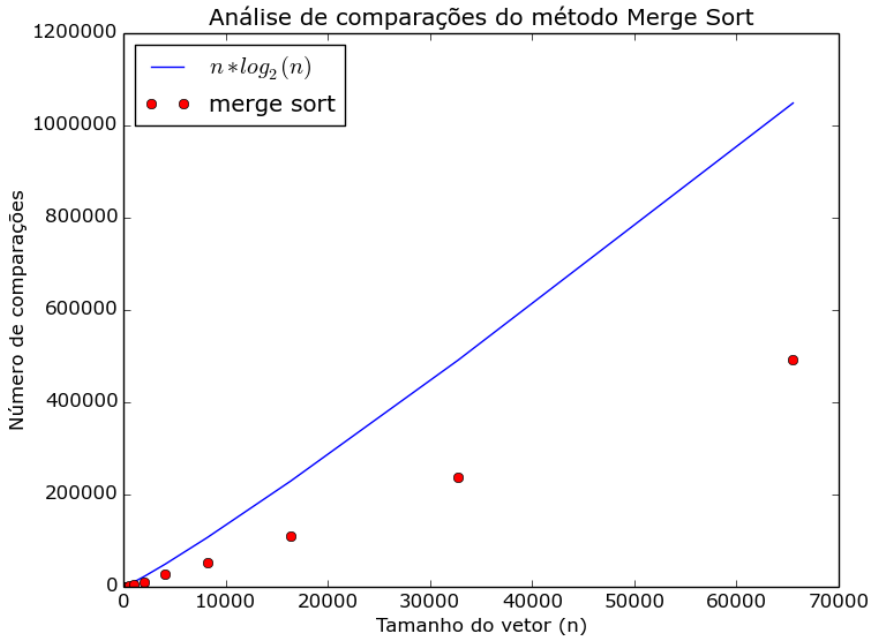

Figura 2.1: *Complexidade de custo do método Merge Sort (Vetor Aleatório)*

   (b) Complexidade de tempo do método Merge Sort disponível na lista de imagens 2.2.
   (c) Complexidade de tempo do método Merge Sort com mínimos quadrados disponível na lista de imagens 2.3.

2. Para um vetor ordenado crescente

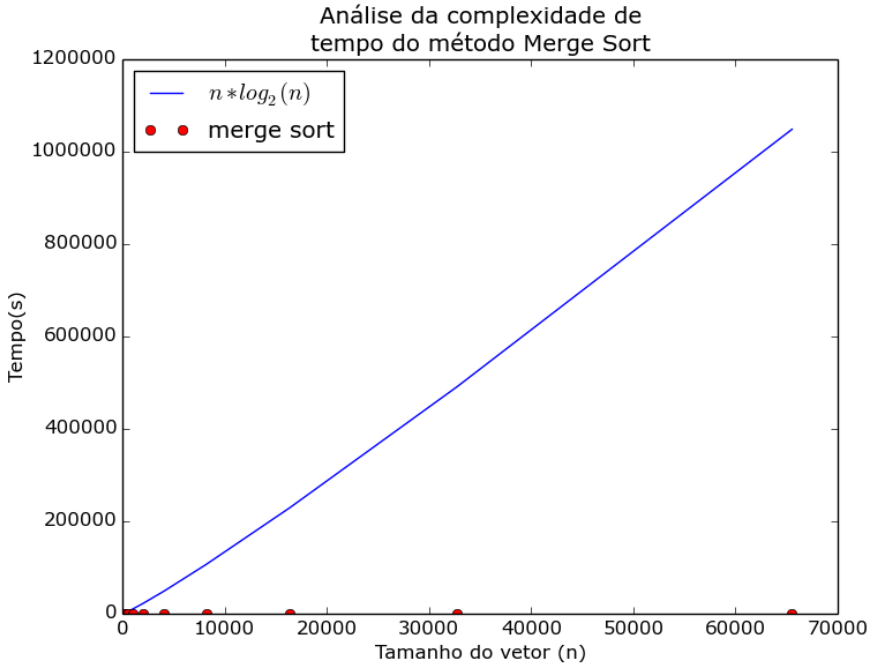

Figura 2.2: *Complexidade de tempo do método Merge Sort (Vetor Aleatório)*

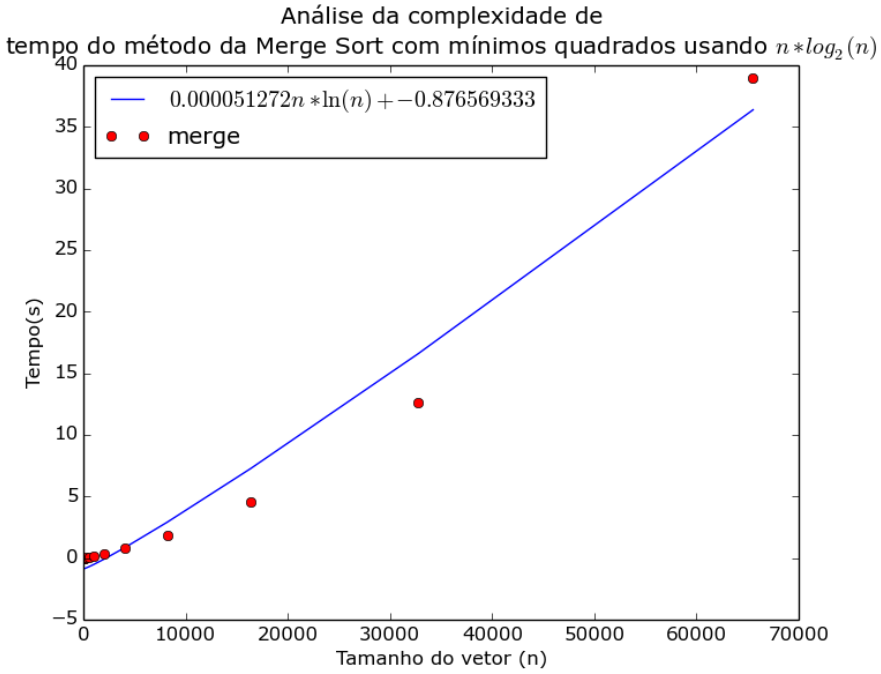

Figura 2.3: *Complexidade de tempo do método Merge Sort com mínimos quadrados (Vetor Aleatório)*

(a) Complexidade de custo do método Merge Sort disponível na lista de imagens 2.4.

(b) Complexidade de tempo do método Merge Sort disponível na lista de imagens 2.5.

(c) Complexidade de tempo do método Merge Sort com mínimos quadrados disponível na lista de imagens 2.6.

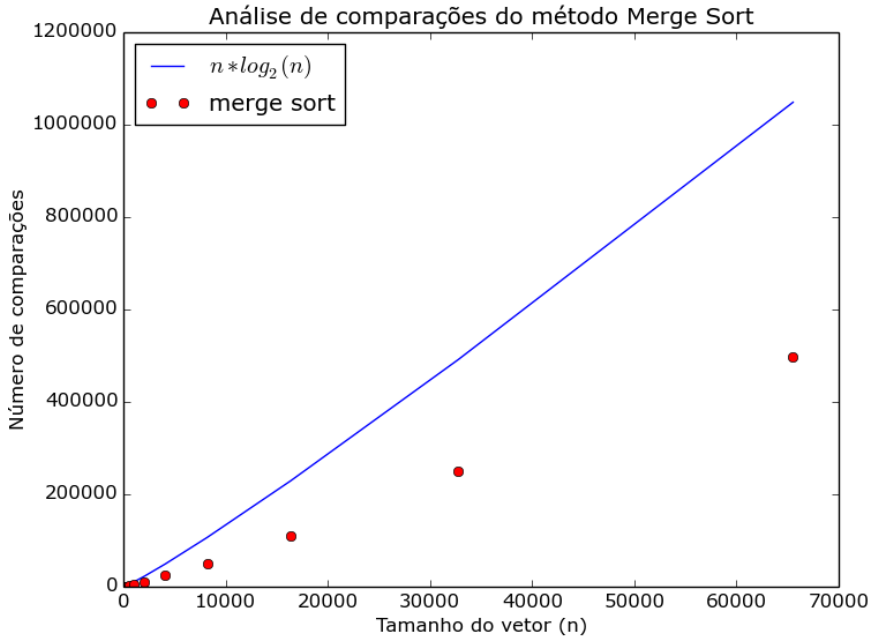

Figura 2.4: *Complexidade de custo do método Merge Sort (Vetor Ordenado Crescente)*

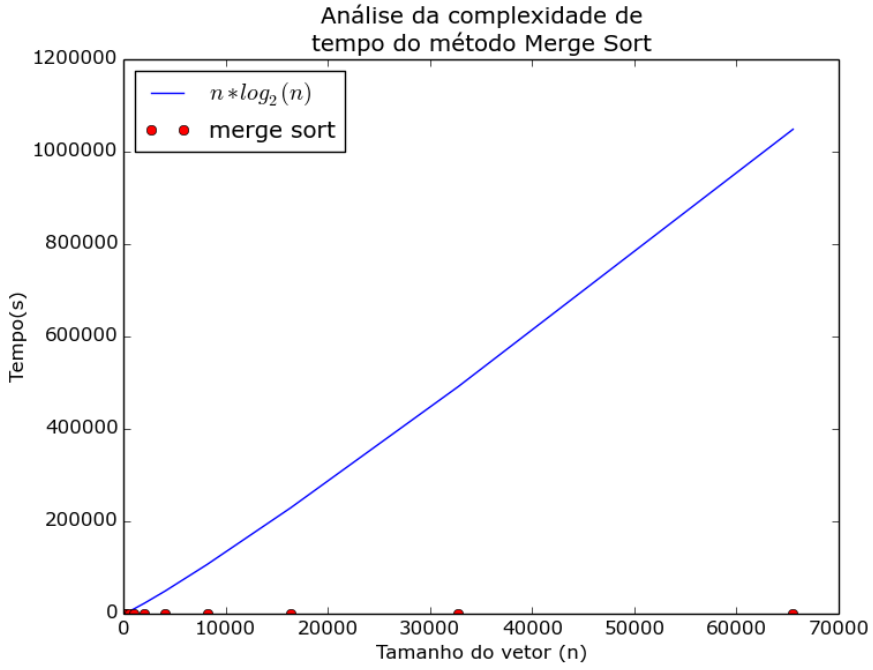

Figura 2.5: *Complexidade de tempo do método Merge Sort (Vetor Ordenado Crescente)*

3. Para um vetor ordenado decrescente

   (a) Complexidade de custo do método Merge Sort disponível na lista de imagens 2.7.

   (b) Complexidade de tempo do método Merge Sort disponível na lista de imagens 2.8.

   (c) Complexidade de tempo do método Merge Sort com mínimos quadrados disponível na lista de imagens 2.9.

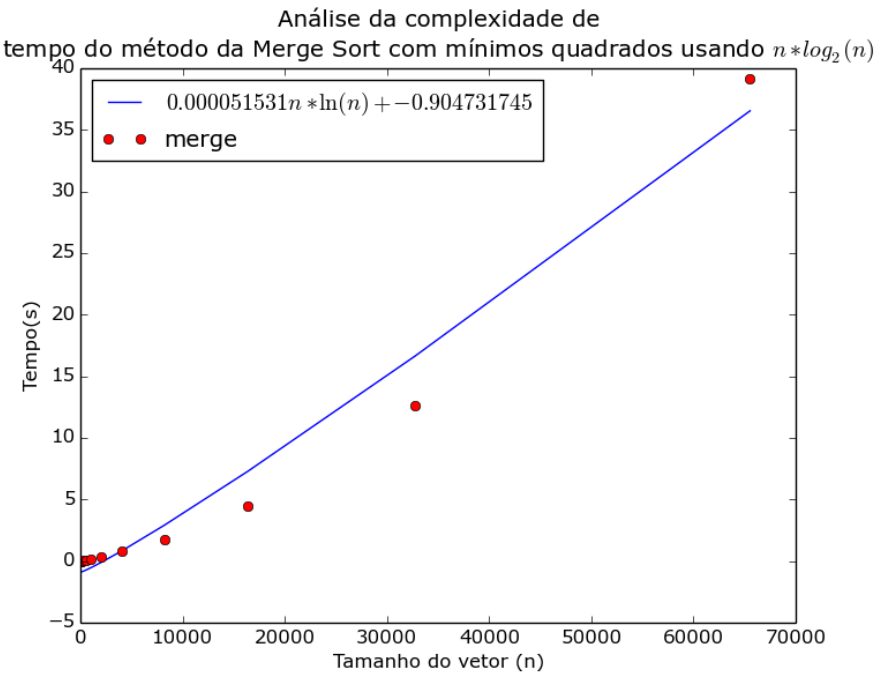

Figura 2.6: *Complexidade de tempo do método Merge Sort com mínimos quadrados (Vetor Ordenado Crescente)*

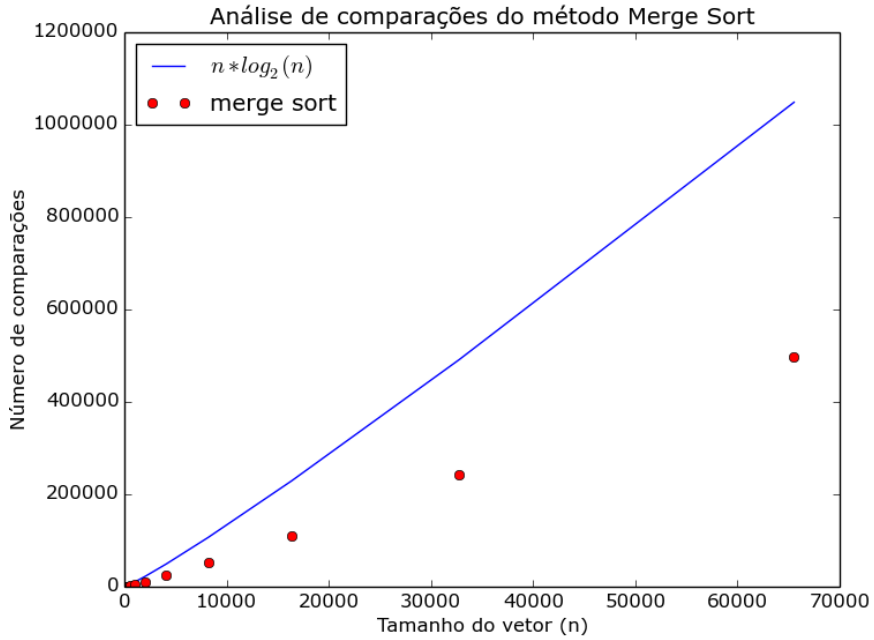

Figura 2.7: *Complexidade de custo do método Merge Sort (Vetor Ordenado Decrescente)*

4. Para um vetor parcialmente ordenado crescente

   (a) Complexidade de custo do método Merge Sort disponível na lista de imagens 2.10.

   (b) Complexidade de tempo do método Merge Sort disponível na lista de imagens 2.11.

   (c) Complexidade de tempo do método Merge Sort com mínimos quadrados dispo-

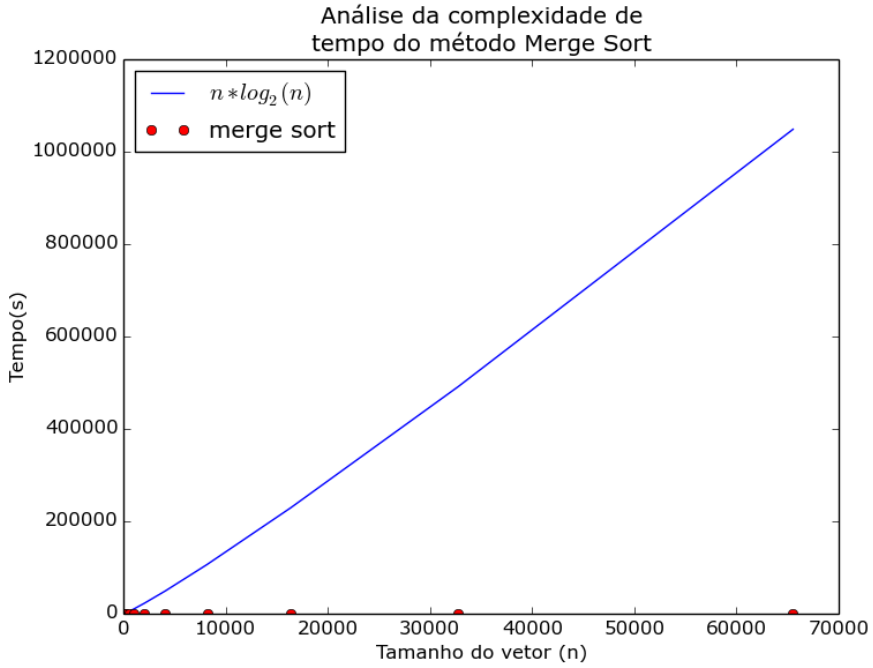

Figura 2.8: *Complexidade de tempo do método Merge Sort (Vetor Ordenado Decrescente)*

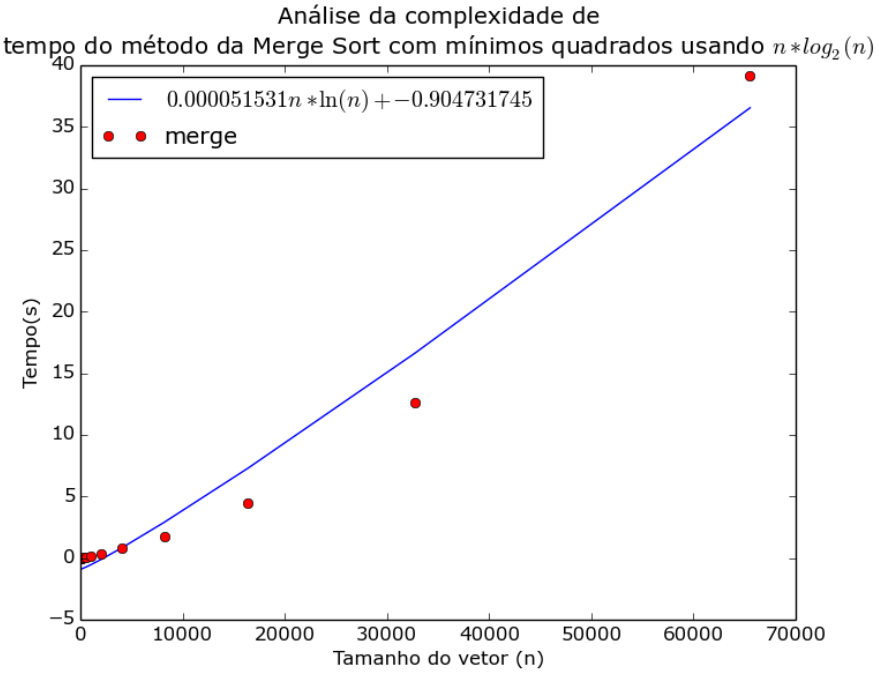

Figura 2.9: *Complexidade de tempo do método Merge Sort com mínimos quadrados (Vetor Ordenado Decrescente)*

nível na lista de imagens 2.12.

5. Para um vetor parcialmente ordenado decrescente

   (a) Complexidade de custo do método Merge Sort disponível na lista de imagens 2.13.

   (b) Complexidade de tempo do método Merge Sort disponível na lista de imagens

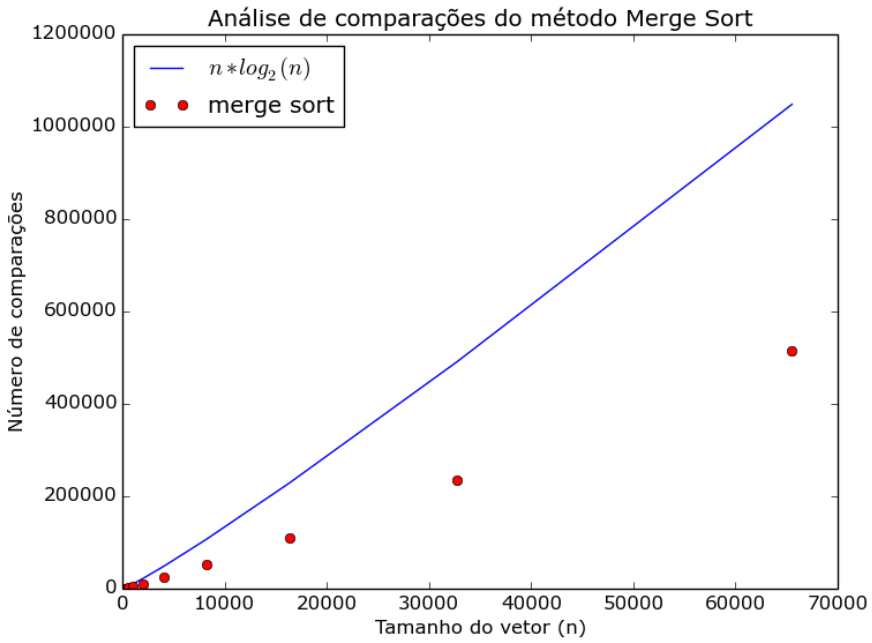

Figura 2.10: *Complexidade de custo do método Merge Sort (Vetor Parcialmente Ordenado Crescente)*

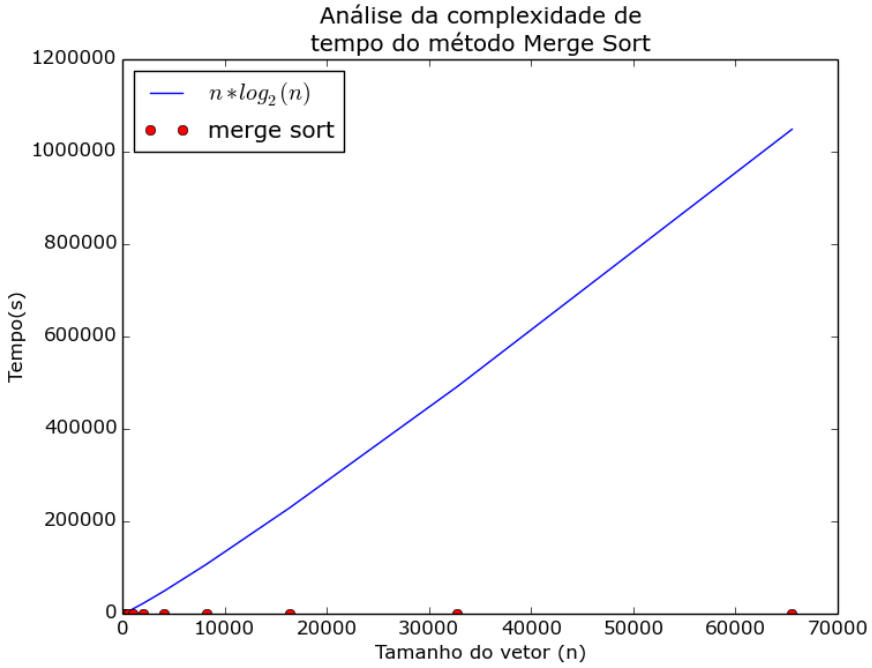

Figura 2.11: *Complexidade de tempo do método Merge Sort (Vetor Parcialmente Ordenado Crescente)*

2.14.

(c) Complexidade de tempo do método Merge Sort com mínimos quadrados disponível na lista de imagens 2.15.

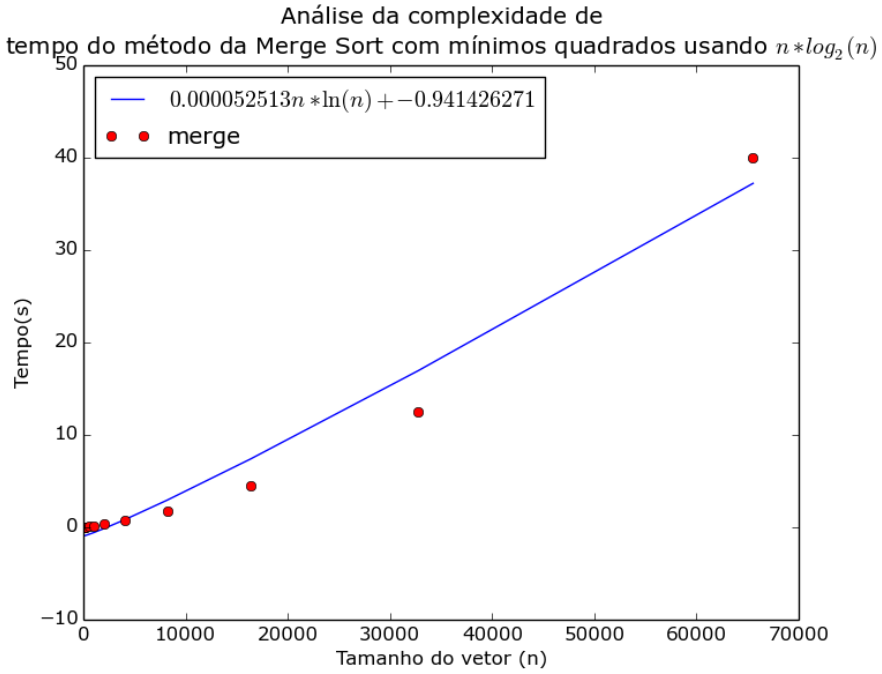

Figura 2.12: *Complexidade de tempo do método Merge Sort com mínimos quadrados (Vetor Parcialmente Ordenado Crescente)*

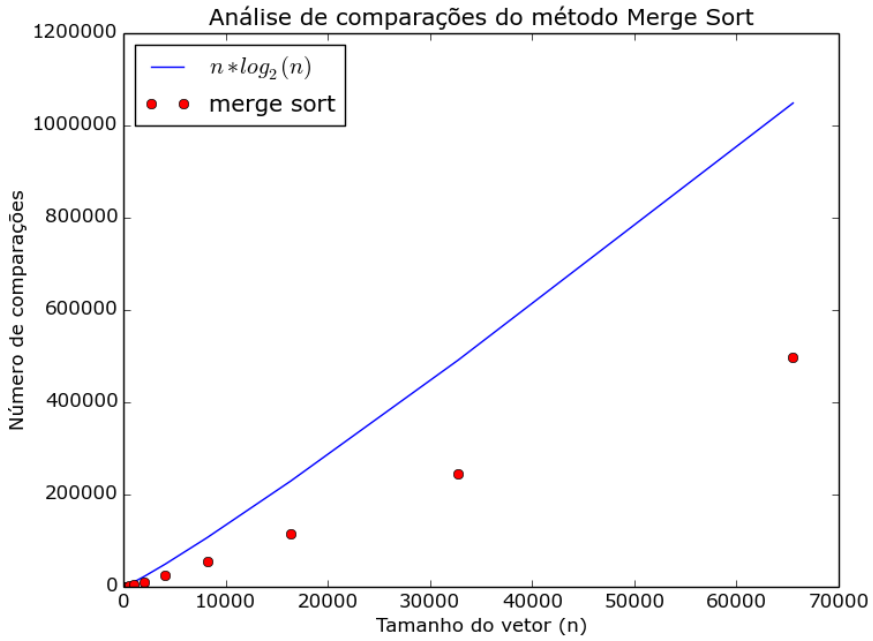

Figura 2.13: *Complexidade de custo do método Merge Sort (Vetor Parcialmente Ordenado Decrescente)*

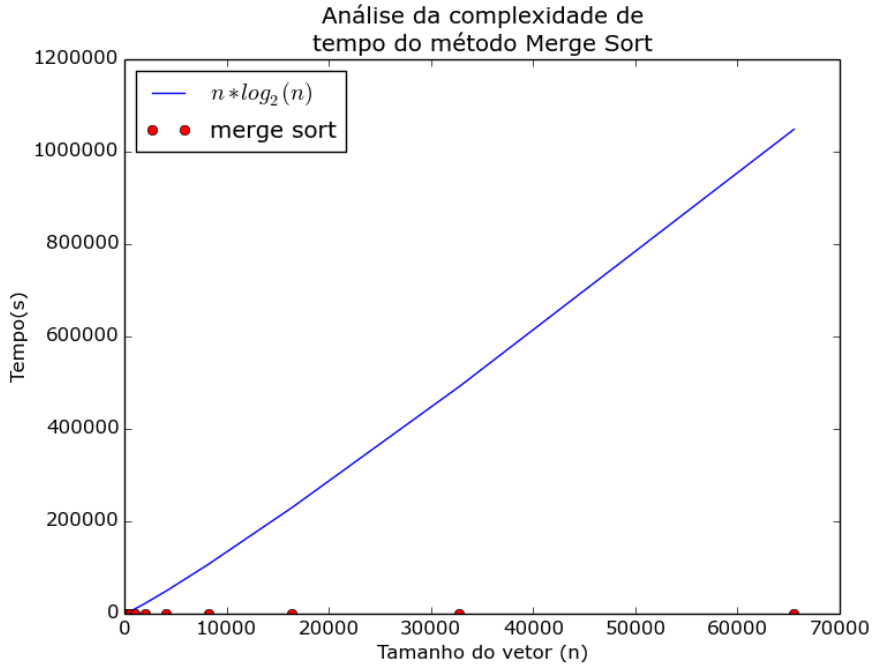

Figura 2.14: *Complexidade de tempo do método Merge Sort (Vetor Parcialmente Ordenado Decrescente)*

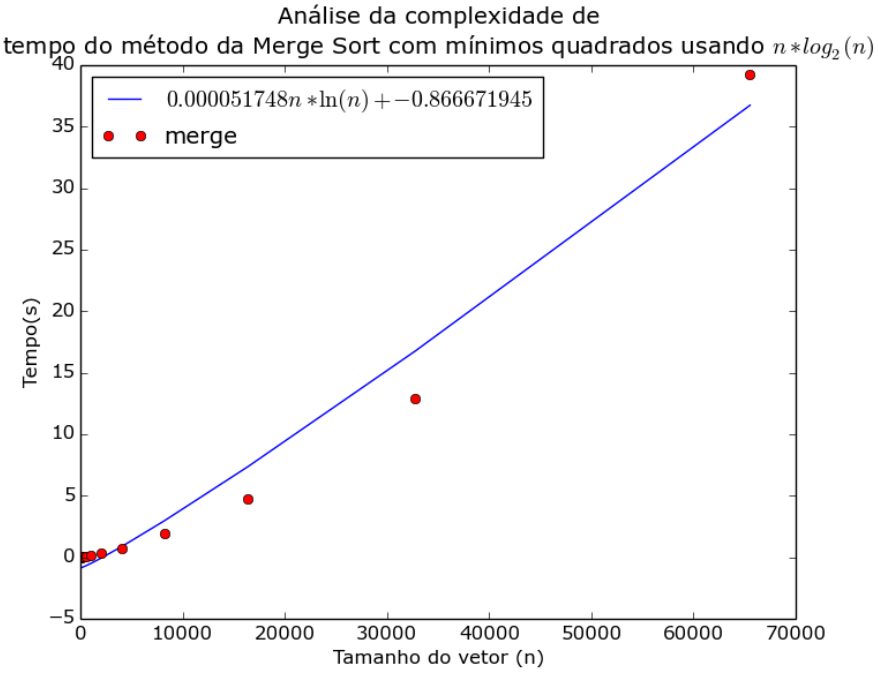

Figura 2.15: *Complexidade de tempo do método Merge Sort com mínimos quadrados (Vetor Parcialmente Ordenado Decrescente)*

Capítulo 3

Tabelas

Seguem as tabelas utilizadas para a análise do método Merge Sort.

Tabela 3.1: *Vetor Aleatorio*

Tamanho do Vetor	**Comparações**	**Tempo(s)**
32	95	0.002739
64	219	0.006053
128	475	0.013715
256	1203	0.033465
512	2334	0.067288
1024	5539	0.162378
2048	11118	0.331024
4096	26759	0.837335
8192	51853	1.817348
16384	110392	4.521457
32768	237663	12.608580
65536	491980	39.949010

Tabela 3.2: *Vetor Ordenado Crescente*

Tamanho do Vetor	Comparações	Tempo(s)
32	94	0.002563
64	239	0.006553
128	476	0.013587
256	1039	0.030059
512	2293	0.066084
1024	5084	0.147069
2048	11228	0.331113
4096	24201	0.766142
8192	51017	1.773804
16384	108998	4.481584
32768	251194	12.579250
65536	497903	39.168110

Tabela 3.3: *Vetor Ordenado Decrescente*

Tamanho do Vetor	Comparações	Tempo(s)
32	95	0.002664
64	231	0.006120
128	500	0.013601
256	1112	0.031692
512	2386	0.069012
1024	5132	0.149191
2048	10950	0.327610
4096	23895	0.753161
8192	52772	1.847495
16384	109071	4.501062
32768	242107	12.890980
65536	497543	39.163110

Tabela 3.4: *Vetor Parcialmente Ordenado Crescente*

Tamanho do Vetor	Comparações	Tempo(s)
32	92	0.002681
64	188	0.006145
128	485	0.013650
256	1086	0.031333
512	2314	0.067024
1024	5056	0.150097
2048	11063	0.329353
4096	24663	0.797015
8192	52275	1.806541
16384	110549	4.538953
32768	234740	12.476440
65536	514762	40.043390

Tabela 3.5: *Vetor Parcialmente Ordenado Decrescente*

Tamanho do Vetor	Comparações	Tempo(s)
32	87	0.002646
64	221	0.006141
128	597	0.016766
256	1058	0.030065
512	2316	0.066540
1024	5330	0.157471
2048	10864	0.329800
4096	23856	0.752486
8192	54738	1.895006
16384	114088	4.755551
32768	245564	12.874460
65536	496850	39.231150

Capítulo 4

Análise

O algoritmo Merge Sort é um algoritmo de divisão e conquista que utiliza a função intercala, $\theta(n)$, que tem como objetivo ordenar as subdivisões estabelecidades pelo merge. Como a cada momento são feitas $n/2$ divisões e isso prossegue até termos um elemento em cada subdivisão. Tal processo é feito $lg(n)$ vezes, gerando uma complexidade de $\theta(nlgn)$. Podemos observar que todas as curvas de todos os gráficos, exceto os de complexidade de tempo sem a interpolação dos mínimos quadrados(Gráficos 2.2,2.5,2.8,2.11,2.14), apresentaram uma correspondência forte com a curva da função $F(x) = x^2$, o que nos permite concluir que, dada a complexidade de tempo do algoritmo Merge Sort por $G(x)$ então $F(x) = c * G(x)$ sendo que c é uma constante maior que zero e $x > x_0$. Portanto, o Merge Sort é $\theta(nlgn)$.

Capítulo 5

Citações e referências bibliográficas

Apêndice A

Códigos extensos

A.1 testdriver.py

Listagem A.1: testdriver.py

```
# coding = utf-8
import subprocess
import numpy as np
import matplotlib.pyplot as plt
import sys , shutil


##PRA CADA NOVO METODO TEM QUE MUDAR
#Sys.path()

## PARA CADA VETOR NOVO OU NOVO METODO TEM QUE MUDAR
#Para o executa_teste a chamada das funções e o shutil.move()
#para os plots        a chamada das funções e o savefig

sys.path.append('/home/gmarson/Git/AnaliseDeAlgoritmos/Trabalho_Final/
    Codigos/Merge')  ## adicionei o código de ordenação
sys.path.append('/home/gmarson/Git/AnaliseDeAlgoritmos/Trabalho_Final/
    relatorio/Resultados/Merge') ## adicionei o resultado do executa_teste


def executa_teste(arqteste, arqsaida,  intervalo,nlin=21,nlin2=47,tempo
    =[3,28,39]):
    """Executa uma sequência de testes contidos em arqteste, com:
       arqsaida: nome do arquivo de saída, ex: tBolha.dat
       nlin: número da linha no arquivo gerado pelo line_profiler contendo
             os dados de interesse. Ex:
       intervalo: tamanhos dos vetores: Ex: 2 ** np.arange(5,10)
    """
    f = open(arqsaida,mode='w', encoding='utf-8')
    f.write('#      n    comparações     tempo(s)\n')

    for n in intervalo:
        cmd = ' '.join(["kernprof -l -v", "testeGeneric.py", str(n)])
        str_saida = subprocess.check_output(cmd, shell=True).decode('utf-8
            ')
        linhas = str_saida.split('\n')
```

```
33          #for i in linhas:
34          #    print(i)
35          #print (linhas)
36
37          #print(linhas[tempo[0]].split()[2])
38
39
40          tempo_total = float(linhas[tempo[0]].split()[2]) + float(linhas[
                tempo[1]].split()[2]) + float(linhas[tempo[2]].split()[2])
41          unidade_tempo = float(linhas[3].split()[2])
42          lcomp = int(linhas[nlin].split()[2]) + int(linhas[nlin2].split()
                [2])
43
44          #print(linhas[nlin].split()[2])
45
46          #print(linhas[nlin2].split()[2])
47          #print(linhas[nlin3].split()[2])
48
49          #print ("unidade tempo: ",unidade_tempo )
50          #print("lcomp: ",lcomp)
51          #print("tempo total",tempo_total)
52
53          #num_comps = int(lcomp[1])
54          str_res = '{:>8} {:>13}  {:13.6f}'.format(n, lcomp ,tempo_total)
55          print(str_res)
56          f.write(str_res + '\n')
57          lcomp = 0
58      f.close()
59      shutil.move("tMerge_vetor_parcialmente_ordenado_decrescente.dat", "/
            home/gmarson/Git/AnaliseDeAlgoritmos/Trabalho_Final/relatorio/
            Resultados/Merge/tMerge_vetor_parcialmente_ordenado_decrescente.dat
            ")
60
61  executa_teste("testeGeneric.py", "
        tMerge_vetor_parcialmente_ordenado_decrescente.dat", 2 ** np.arange
        (5,17))
62
63  def executa_teste_memoria(arqteste, arqsaida, nlin, intervalo):
64      """Executa uma sequência de testes contidos em arqteste, com:
65         arqsaida: nome do arquivo de saída, ex: tBolha.dat
66         nlin: número da linha no arquivo gerado pelo line_profiler contendo
67               os dados de interesse. Ex: 14
68         intervalo: tamanhos dos vetores: Ex: 2 ** np.arange(5,10)
69      """
70      f = open(arqsaida,mode='w', encoding='utf-8')
71      f.write('#       n    comparações       tempo(s)\n')
72
73      for n in intervalo:
74          cmd = ' '.join(["kernprof -l -v ", "testeGeneric.py", str(n)])
75
76          str_saida = subprocess.check_output(cmd, shell=True).decode('utf-8
                ')
77
78          linhas = str_saida.split('\n')
79          for i in linhas:
80              print(i)
81
82          print ("Linhas:",linhas[1])
83
```

```
84          unidade_tempo = float(linhas[1].split()[2])
85
86          str_res = '{:>8} {:>13} {:13.6f}'.format(n, n, n)
87          print(str_res)
88          f.write(str_res + '\n')
89      f.close()
90      #shutil.move("tMerge_memoria.dat", "/home/gmarson/Git/
            AnaliseDeAlgoritmos/Trabalho_Final/relatorio/Resultados/Merge/
            tMerge_memoria.dat")
91
92  #executa_teste_memoria("testeGeneric.py", "tMerge_memoria.dat", 14, 2 **
        np.arange(5,15))
93
94  def plota_teste1(arqsaida):
95      n, c, t = np.loadtxt(arqsaida, unpack=True)
96      #print("n: ",n,"\nc: ",c,"\nt: ",t)
97      #n eh o tamanho da entrada , c eh o tanto de comparações e t eh o
            tempo gasto
98      plt.plot(n, n * np.log2(n), label='$n * log_2(n)$')  ## custo esperado
             bubble Sort
99      plt.plot(n, c, 'ro', label='merge sort')
100     # Posiciona a legenda
101     plt.legend(loc='upper left')
102
103     # Posiciona o título
104     plt.title('Análise de comparações do método Merge Sort')
105
106     # Rotula os eixos
107     plt.xlabel('Tamanho do vetor (n)')
108     plt.ylabel('Número de comparações')
109
110     plt.savefig('relatorio/imagens/Merge/
            merge_plot_1_parcialmente_ordenado_decrescente.png')
111     plt.show()
112
113
114
115 def plota_teste2(arqsaida):
116     n, c, t = np.loadtxt(arqsaida, unpack=True)
117     plt.plot(n, n * np.log2(n), label='$n * log_2(n)$')
118     plt.plot(n, t, 'ro', label='merge sort')
119
120     # Posiciona a legenda
121     plt.legend(loc='upper left')
122
123     # Posiciona o título
124     plt.title('Análise da complexidade de \ntempo do método Merge Sort')
125
126     # Rotula os eixos
127     plt.xlabel('Tamanho do vetor (n)')
128     plt.ylabel('Tempo(s)')
129
130     plt.savefig('relatorio/imagens/Merge/
            merge_plot_2_parcialmente_ordenado_decrescente.png')
131     plt.show()
132
133
134
135
```

```
136 def plota_teste3(arqsaida):
137     n, c, t = np.loadtxt(arqsaida, unpack=True)
138 
139     # Calcula os coeficientes de um ajuste a um polinômio de grau 2 usando
140     # o método dos mínimos quadrados
141     coefs = np.polyfit(n, t, 2)
142     p = np.poly1d(coefs)
143 
144     plt.plot(n, p(n), label='$n^2$')
145     plt.plot(n, t, 'ro', label='merge sort')
146 
147     # Posiciona a legenda
148     plt.legend(loc='upper left')
149 
150     # Posiciona o título
151     plt.title('Análise da complexidade de \ntempo do método Merge Sort com
            mínimos quadrados')
152 
153     # Rotula os eixos
154     plt.xlabel('Tamanho do vetor (n)')
155     plt.ylabel('Tempo(s)')
156 
157     plt.savefig('relatorio/imagens/Merge/
           merge_plot_3_parcialmente_ordenado_decrescente.png')
158     plt.show()
159 
160 plota_teste1("/home/gmarson/Git/AnaliseDeAlgoritmos/Trabalho_Final/
       relatorio/Resultados/Merge/
       tMerge_vetor_parcialmente_ordenado_decrescente.dat")
161 plota_teste2("/home/gmarson/Git/AnaliseDeAlgoritmos/Trabalho_Final/
       relatorio/Resultados/Merge/
       tMerge_vetor_parcialmente_ordenado_decrescente.dat")
162 plota_teste3("/home/gmarson/Git/AnaliseDeAlgoritmos/Trabalho_Final/
       relatorio/Resultados/Merge/
       tMerge_vetor_parcialmente_ordenado_decrescente.dat")
163 
164 
165 def plota_teste4(arqsaida):
166     n, c, t = np.loadtxt(arqsaida, unpack=True)
167 
168     # Calcula os coeficientes de um ajuste a um polinômio de grau 2 usando
169     # o método dos mínimos quadrados
170     coefs = np.polyfit(n, c, 2)
171     p = np.poly1d(coefs)
172 
173     plt.plot(n, p(n), label='$n^2$')
174     plt.plot(n, c, 'ro', label='bubble sort')
175 
176     # Posiciona a legenda
177     plt.legend(loc='upper left')
178 
179     # Posiciona o título
180     plt.title('Análise da complexidade de \ntempo do método da bolha')
181 
182     # Rotula os eixos
183     plt.xlabel('Tamanho do vetor (n)')
184     plt.ylabel('Número de comparações')
185 
186     plt.savefig('bubble4.png')
```

```
187     plt.show()
188
189 def plota_teste5(arqsaida):
190     n, c, t = np.loadtxt(arqsaida, unpack=True)
191
192     # Calcula os coeficientes de um ajuste a um polinômio de grau 2 usando
193     # o método dos mínimos quadrados
194     coefs = np.polyfit(n, c, 2)
195     p = np.poly1d(coefs)
196
197     # set_yscale('log')
198     # set_yscale('log')
199     plt.semilogy(n, p(n), label='$n^2$')
200     plt.semilogy(n, c, 'ro', label='bubble sort')
201
202     # Posiciona a legenda
203     plt.legend(loc='upper left')
204
205     # Posiciona o título
206     plt.title('Análise da complexidade de \ntempo do método da bolha')
207
208     # Rotula os eixos
209     plt.xlabel('Tamanho do vetor (n)')
210     plt.ylabel('Número de comparações')
211
212     plt.savefig('bubble5.png')
213     plt.show()
```