Quicksort Estrutura de Dados — QXD0010



Prof. Atílio Gomes Luiz gomes.atilio@ufc.br

Universidade Federal do Ceará

 2° semestre/2022





Vimos três algoritmos de ordenação $O(n^2)$:

- bubblesort
- insertionsort
- selectionsort



Vimos três algoritmos de ordenação $O(n^2)$:

- bubblesort
- insertionsort
- selectionsort

Nessa aula veremos um algoritmo de ordenação $O(n\log n)$ no caso médio e $O(n^2)$ no pior caso



Vimos três algoritmos de ordenação $O(n^2)$:

- bubblesort
- insertionsort
- selectionsort

Nessa aula veremos um algoritmo de ordenação $O(n\log n)$ no caso médio e $O(n^2)$ no pior caso

Ele é baseado em uma técnica de projeto de algoritmos chamada Divisão e Conquista ou Dividir para Conquistar



Algoritmo proposto por C. A. R. Hoare em 1960.



- É o algoritmo de ordenação mais rápido que se conhece para uma ampla variedade de situações.
- Apesar disso, possui complexidade $O(n^2)$ no pior caso.
- Provavelmente é o mais utilizado (ou pelo menos deveria ser).



Divisão e Conquista



Observação:

- A recursão parte do princípio que é mais fácil resolver problemas menores.
- Para certos problemas, podemos dividi-los em duas ou mais partes.



Observação:

- A recursão parte do princípio que é mais fácil resolver problemas menores.
- Para certos problemas, podemos dividi-los em duas ou mais partes.

Etapas do paradigma de Divisão e Conquista:

- o Dividir: Quebramos o problema em vários subproblemas menores.
 - ex: quebramos um vetor a ser ordenado em dois.



Observação:

- A recursão parte do princípio que é mais fácil resolver problemas menores.
- Para certos problemas, podemos dividi-los em duas ou mais partes.

Etapas do paradigma de Divisão e Conquista:

- o Dividir: Quebramos o problema em vários subproblemas menores.
 - ex: quebramos um vetor a ser ordenado em dois.
- Conquistar: Os subproblemas são resolvidos recursivamente. Se eles forem pequenos o bastante, eles são resolvidos usando o próprio algoritmo que está sendo definido.
 - ex: um subvetor com um único elemento já está ordenado.



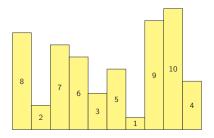
Observação:

- A recursão parte do princípio que é mais fácil resolver problemas menores.
- Para certos problemas, podemos dividi-los em duas ou mais partes.

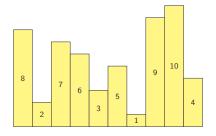
Etapas do paradigma de Divisão e Conquista:

- o Dividir: Quebramos o problema em vários subproblemas menores.
 - ex: quebramos um vetor a ser ordenado em dois.
- Conquistar: Os subproblemas são resolvidos recursivamente. Se eles forem pequenos o bastante, eles são resolvidos usando o próprio algoritmo que está sendo definido.
 - ex: um subvetor com um único elemento já está ordenado.
- Combinar: Combinamos a solução dos problemas menores a fim de obter a solução para o problema maior.
 - ex: intercalamos os dois vetores ordenados ou, se todos os elementos de um vetor forem menores que o do outro, apenas concatenamos os vetores.



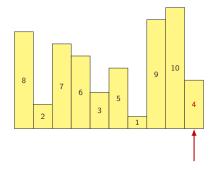






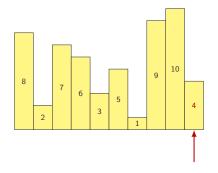
• Escolhemos um pivô (ex: 4)





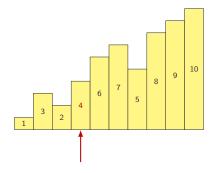
• Escolhemos um pivô (ex: 4)





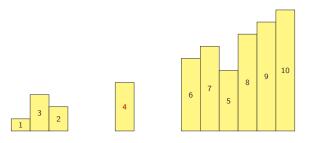
- Escolhemos um pivô (ex: 4)
- Colocamos os elementos menores que o pivô à esquerda dele
- e os elementos maiores que o pivô à direita





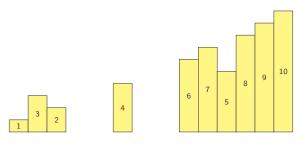
- Escolhemos um pivô (ex: 4)
- Colocamos os elementos menores que o pivô à esquerda dele
- e os elementos maiores que o pivô à direita





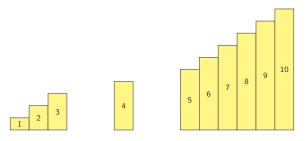
- Escolhemos um pivô (ex: 4)
- Colocamos os elementos menores que o pivô à esquerda dele
- e os elementos maiores que o pivô à direita
- Com isso, o pivô já está na posição correta





- Escolhemos um pivô (ex: 4)
- Colocamos os elementos menores que o pivô à esquerda dele
- e os elementos maiores que o pivô à direita
- Com isso, o pivô já está na posição correta
- O lado esquerdo e o direito podem ser ordenados independentemente





- Escolhemos um pivô (ex: 4)
- Colocamos os elementos menores que o pivô à esquerda dele
- e os elementos maiores que o pivô à direita
- Com isso, o pivô já está na posição correta
- O lado esquerdo e o direito podem ser ordenados independentemente

Divisão e Conquista no Quicksort



Quicksort aplica o paradigma de Divisão e Conquista:

• Dividir: rearranja o vetor A[p..r] em dois subvetores (possivelmente vazios) A[p..q-1] e A[q+1..r] tal que

$$A[p..q-1] \le A[q] < A[q+1..r].$$

O índice q é calculado como parte deste procedimento de separação.

Divisão e Conquista no Quicksort



Quicksort aplica o paradigma de Divisão e Conquista:

• Dividir: rearranja o vetor A[p..r] em dois subvetores (possivelmente vazios) A[p..q-1] e A[q+1..r] tal que

$$A[p..q-1] \le A[q] < A[q+1..r].$$

O índice q é calculado como parte deste procedimento de separação.

• Conquistar: Ordena os subvetores A[p..q-1] e A[q+1..r] por meio de chamadas recursivas ao quicksort.

Divisão e Conquista no Quicksort



Quicksort aplica o paradigma de Divisão e Conquista:

• Dividir: rearranja o vetor A[p..r] em dois subvetores (possivelmente vazios) A[p..q-1] e A[q+1..r] tal que

$$A[p..q-1] \le A[q] < A[q+1..r].$$

O índice q é calculado como parte deste procedimento de separação.

- Conquistar: Ordena os subvetores A[p..q-1] e A[q+1..r] por meio de chamadas recursivas ao quicksort.
- Combinar: Os subvetores já estão ordenados, não há o que fazer: o vetor A[p..r] encontra-se ordenado.



Um problema subjacente: Particionar um vetor





O núcleo do algoritmo Quicksort é o seguinte problema da partição:

ullet rearranjar um vetor $A[p \dots r]$ de modo que

$$A[p \dots j-1] \le A[j] < A[j+1 \dots r]$$

para algum j tal que $p \leq j < r$.



O núcleo do algoritmo Quicksort é o seguinte problema da partição:

ullet rearranjar um vetor $A[p \dots r]$ de modo que

$$A[p \dots j-1] \le A[j] < A[j+1 \dots r]$$

para algum j tal que $p \le j < r$.

• Exemplo: aqui, A[j] é o pivô.

777 222 111 777 999 444 555 666 555 888

j 666 | 222 | 111 | 777 | 555 | 444 | 555 | 777 | 999 | 888



• O ponto de partida para a solução deste problema é a escolha de um pivô, digamos p.



- O ponto de partida para a solução deste problema é a escolha de um pivô, digamos p.
- Os elementos do vetor que forem maiores que p serão considerados grandes e os demais serão considerados pequenos.



- O ponto de partida para a solução deste problema é a escolha de um pivô, digamos p.
- Os elementos do vetor que forem maiores que p serão considerados grandes e os demais serão considerados pequenos.
- É importante escolher p de tal modo que as duas partes do vetor rearranjado sejam estritamente menores que o vetor todo.



- O ponto de partida para a solução deste problema é a escolha de um pivô, digamos p.
- Os elementos do vetor que forem maiores que p serão considerados grandes e os demais serão considerados pequenos.
- É importante escolher p de tal modo que as duas partes do vetor rearranjado sejam estritamente menores que o vetor todo.
- A dificuldade está em resolver o problema da partição de maneira rápida sem usar muito espaço de trabalho.

O algoritmo da partição



```
1 /* Recebe um vetor A[1..r] com l <= r.
   * Rearranja os elementos do vetor e devolve
   * j em l..r tal que A[1..j-1] \le A[j] \le A[j+1..r].
   */
  int partition (int A[], int 1, int r) {
       int pivo = A[r];
6
       int i = 1:
      for (int k = 1; k < r; k++) {
8
           if (A[k] <= pivo) {</pre>
               int aux = A[k]:
10
               A[k] = A[i];
11
12
               A[j] = aux;
13
               j++;
14
15
      A[r] = A[i];
16
      A[i] = pivo;
17
18
      return j;
19 }
```

Corretude do algoritmo da separação



- No início de cada iteração do laço **for** valem os seguintes invariantes:
- (a) A[l...r] é uma permutação do vetor original,
- (b) $A[l...j-1] \le \text{pivo } < A[j...k-1],$
- (c) A[r] = p
- (d) $l \leq j \leq k \leq r$.



Início de uma iteração do laço for

Corretude do algoritmo da separação



- No início de cada iteração do laço **for** valem os seguintes invariantes:
 - (a) A[l...r] é uma permutação do vetor original,
 - (b) $A[l...j-1] \le \text{pivo } < A[j...k-1],$
 - (c) A[r] = p
- (d) $l \leq j \leq k \leq r$.



Início de uma iteração do laço for



Última iteração do laço for.

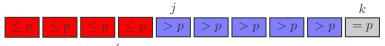
Corretude do algoritmo da separação



- No início de cada iteração do laço **for** valem os seguintes invariantes:
- (a) A[l...r] é uma permutação do vetor original,
- (b) $A[l...j-1] \le \text{pivo } < A[j...k-1],$
- (c) A[r] = p
- (d) $l \leq j \leq k \leq r$.



Início de uma iteração do laço for



Última iteração do laço for.



Passagem pela última linha da função separa.





```
int partition (int A[], int 1, int r) {
       int pivo = A[r];
       int j = 1;
       for (int k = 1; k < r; k++) {
           if (A[k] <= pivo) {</pre>
                int aux = A[k];
6
               A[k] = A[j];
               A[i] = aux;
9
                j++;
10
11
      A[r] = A[j];
12
      A[j] = pivo;
13
      return j:
14
15 }
```

 O consumo de tempo da função partition é proporcional ao número de iterações.





```
int partition (int A[], int 1, int r) {
       int pivo = A[r];
       int j = 1;
       for (int k = 1; k < r; k++) {
           if (A[k] <= pivo) {
                int aux = A[k];
6
               A[k] = A[j];
               A[i] = aux;
9
               j++:
10
11
      A[r] = A[i]:
12
      A[j] = pivo;
13
14
      return j;
15 }
```

- O consumo de tempo da função partition é proporcional ao número de iteracões.
- Como o número de iterações é r-l+1, podemos dizer que o consumo de tempo é proporcional ao número de elementos do vetor.

Quicksort



```
1 /**
2  * Esta funcao rearranja o vetor A[p..r],
3  * com p <= r+1, de modo que ele fique
4  * em ordem crescente.
5  */
6  void quicksort (int A[], int 1, int r) {
7    if (1 < r) {
8        int j = partition(A, 1, r);
9        quicksort(A, 1, j-1);
10        quicksort(A, j+1, r);
11   }
12 }</pre>
```

O desempenho do Quicksort



- O tempo de execução do quicksort depende do particionamento ser balanceado ou não ser balanceado.
- Se o particionamento é balanceado, o algoritmo é executado em tempo $O(n \lg n)$.
- Contudo, se o particionamento é não balanceado, ele pode ser executado assintoticamente tão lento quanto a ordenação por inserção.

Pior caso do Quicksort



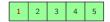
- O comportamento do pior caso para o quicksort ocorre quando a rotina de separação produz um subproblema com n-1 elementos e um com 0 elementos.
 - Isso acontece, por exemplo, se o vetor já estiver ordenado ou quase ordenado

 Vamos considerar que esse particionamento n\u00e3o balanceado surja em cada chamada recursiva.

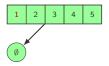


1	2	3	4	5
		-		-

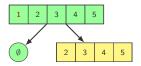




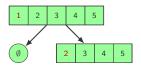




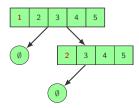




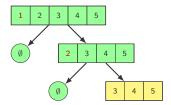




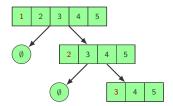




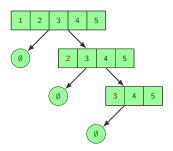




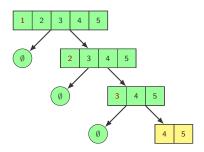




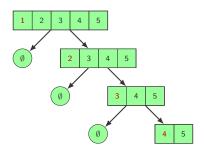




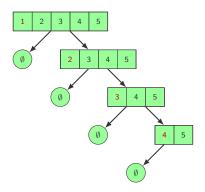




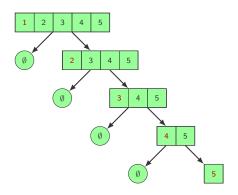




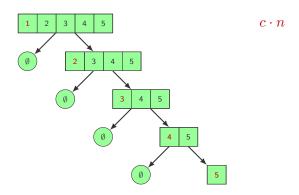




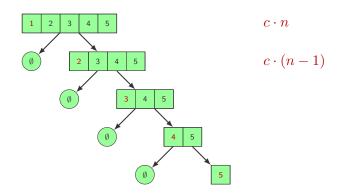




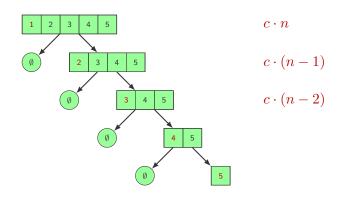




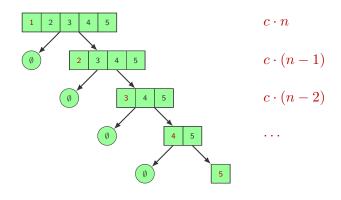




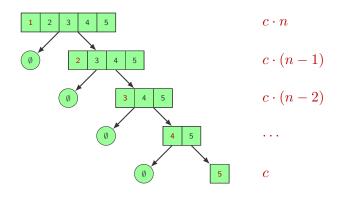




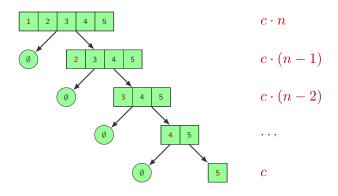




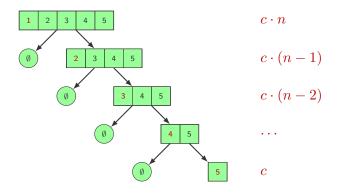






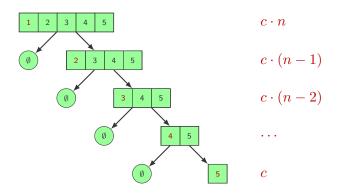






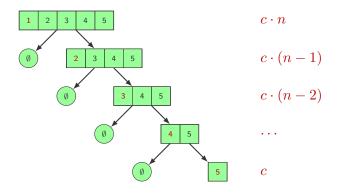
$$c \cdot n + c \cdot (n-1) + \cdots + c$$





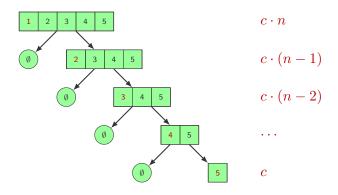
$$c \cdot n + c \cdot (n-1) + \dots + c = c \sum_{j=1}^{n} j$$





$$c \cdot n + c \cdot (n-1) + \dots + c = c \sum_{j=1}^{n} j = c \frac{n(n+1)}{2}$$





$$c \cdot n + c \cdot (n-1) + \dots + c = c \sum_{j=1}^{n} j = c \frac{n(n+1)}{2} = O(n^2)$$



• Na divisão mais equitativa possível, a função partition produz dois subproblemas, cada um de tamanho não maior que n/2.



- Na divisão mais equitativa possível, a função partition produz dois subproblemas, cada um de tamanho não maior que n/2.
- Nesse caso, teríamos a seguinte recorrência para o tempo de execução:

$$T(n) = 2T(n/2) + O(n).$$



- Na divisão mais equitativa possível, a função partition produz dois subproblemas, cada um de tamanho não maior que n/2.
- Nesse caso, teríamos a seguinte recorrência para o tempo de execução:

$$T(n) = 2T(n/2) + O(n).$$

• Resolvendo a recorrência, temos que $T(n) = O(n \log n)$.

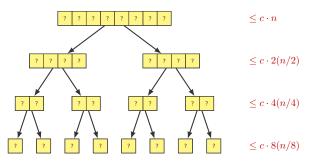


- Na divisão mais equitativa possível, a função partition produz dois subproblemas, cada um de tamanho não maior que n/2.
- Nesse caso, teríamos a seguinte recorrência para o tempo de execução:

$$T(n) = 2T(n/2) + O(n).$$

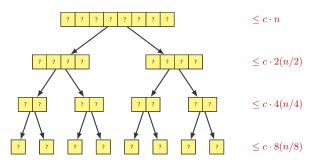
- Resolvendo a recorrência, temos que $T(n) = O(n \log n)$.
- Balanceando igualmente os dois lados da partição em todo nível da recursão, obtemos um algoritmo assintoticamente mais rápido.





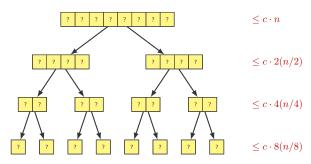
• No nível k gastamos tempo $\leq c \cdot n$





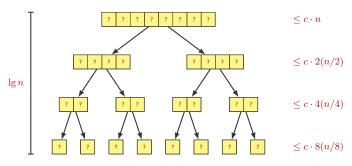
- No nível k gastamos tempo $\leq c \cdot n$
- Quantos níveis temos?





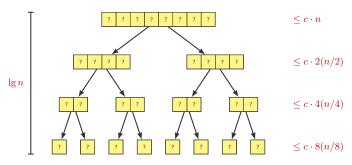
- No nível k gastamos tempo $\leq c \cdot n$
- Quantos níveis temos?
 - \circ Dividimos n por 2 até que fique menor ou igual a 1





- No nível k gastamos tempo $\leq c \cdot n$
- Quantos níveis temos?
 - \circ Dividimos n por 2 até que fique menor ou igual a 1
 - \circ Ou seja, $l = \lg n$





- No nível k gastamos tempo $\leq c \cdot n$
- Quantos níveis temos?
 - \circ Dividimos n por 2 até que fique menor ou igual a 1
 - \circ Ou seja, $l = \lg n$
- Tempo total: $c n \lg n = O(n \lg n)$



Altura da pilha de execução



Altura da pilha de execução do Quicksort



• Na versão básica do Quicksort, o código cuida imediatamente do subvetor A[l..j-1] e trata A[j+1..r] somente depois que A[l..j-1] já está ordenado.



- Na versão básica do Quicksort, o código cuida imediatamente do subvetor A[l..j-1] e trata A[j+1..r] somente depois que A[l..j-1] já está ordenado.
- Dependendo do valor de j nas sucessivas invocações da função, a pilha de execução pode crescer muito. Atingindo altura igual ao número de elementos do vetor.
 - \circ Uso de memória adicional no pior caso: O(n)



- Na versão básica do Quicksort, o código cuida imediatamente do subvetor A[l..j-1] e trata A[j+1..r] somente depois que A[l..j-1] já está ordenado.
- Dependendo do valor de j nas sucessivas invocações da função, a pilha de execução pode crescer muito. Atingindo altura igual ao número de elementos do vetor.
 - Uso de memória adicional no pior caso: O(n)
- Esse fenômeno não altera o tempo de execução do algoritmo, mas pode esgotar o espaço de memória (estouro de pilha).



- Para controlar o crescimento da pilha de execução é preciso tomar duas providências:
- (1) Cuidar primeiro do menor dos subvetores A[l..j-1] e A[j+1..r] e
- (2) Eliminar a segunda invocação recursiva da função quicksort



- Para controlar o crescimento da pilha de execução é preciso tomar duas providências:
 - (1) Cuidar primeiro do menor dos subvetores A[l..j-1] e A[j+1..r] e
 - (2) Eliminar a segunda invocação recursiva da função quicksort
- Se tomarmos estas providências, o tamanho do subvetor que está no topo da pilha de execução será sempre menor que a metade do tamanho do subvetor que está logo abaixo na pilha.
- Assim, se a função for aplicada a um vetor com n elementos, a altura da pilha não passará de $\log_2 n$

Passo 2:



Eliminar a segunda invocação recursiva da função quicksort

Velho:

```
1 void quicksort (int A[], int 1, int r) {
2    if (1 < r) {
3        int j = partition(A, 1, r);
4        quicksort(A, 1, j-1);
5        quicksort(A, j+1, r);
6    }
7 }</pre>
```

Novo:

```
1 void quicksort_2 (int A[], int 1, int r) {
2    while(1 < r) {
3         int j = partition(A, 1, r);
4         quicksort_2 (A, 1, j-1);
5         1 = j + 1;
6    }
7 }</pre>
```

Passo 1:



Cuidar primeiro do menor subvetor

```
1 void quicksort_3 (int A[], int 1, int r) {
2    while(l < r) {
3         int j = partition(A, 1, r);
4         if(j-l < r-j) {
5             quicksort_2(A, 1, j-1);
6             l = j + 1;
7         }
8         else {
9             quicksort_2(A, j+1, r);
10             r = j - 1;
11         }
12     }
13 }</pre>
```



O QuickSort é um algoritmo de ordenação $O(n^2)$



O QuickSort é um algoritmo de ordenação $O(n^2)$

• Mas ele pode ser rápido na prática



- O QuickSort é um algoritmo de ordenação $O(n^2)$
- Mas ele pode ser rápido na prática
- Leva tempo $O(n \lg n)$ (em média) para ordenar uma permutação aleatória



O QuickSort é um algoritmo de ordenação $O(n^2)$

- Mas ele pode ser rápido na prática
- Leva tempo $O(n \lg n)$ (em média) para ordenar uma permutação aleatória
- Precisa de espaço adicional O(n) para a pilha de recursão

Comparação Assintótica



Algoritmo	Melhor Caso	Caso Médio	Pior Caso	Memória
BubbleSort	O(n)	$O(n^2)$	$O(n^2)$	O(1)
InsertionSort	O(n)	$O(n^2)$	$O(n^2)$	O(1)
SelectionSort	$O(n^2)$	$O(n^2)$	$O(n^2)$	O(1)

Comparação Assintótica



Algoritmo	Melhor Caso	Caso Médio	Pior Caso	Memória
BubbleSort	O(n)	$O(n^2)$	$O(n^2)$	O(1)
InsertionSort	O(n)	$O(n^2)$	$O(n^2)$	O(1)
SelectionSort	$O(n^2)$	$O(n^2)$	$O(n^2)$	O(1)
QuickSort	$O(n \lg n)$	$O(n \lg n)$	$O(n^2)$	O(n)





- (1) Escreva uma função que rearranje um vetor V[p..r] de números inteiros de modo que os elementos negativos e nulos fiquem à esquerda e os positivos fiquem à direita. Em outras palavras, rearranje o vetor de modo que tenhamos $V[p..j-1] \leq 0$ e V[j..r] > 0 para algum j em $\{p,\ldots,r+1\}$.
 - o Procure escrever uma função eficiente que não use vetor auxiliar.
- (2) Digamos que um vetor V[p..r] está arrumado se existe j em p..r que satisfaz:

$$V[p..j-1] \le V[j] < V[j+1..r]$$

Escreva um algoritmo que decida se V[p..r] está arrumado. Em caso afirmativo, o seu algoritmo deve devolver o valor de j.



- (3) Escreva uma implementação do algoritmo Quicksort que evite aplicar a função a vetores com menos do que dois elementos.
- (4) A função separa produz um rearranjo estável do vetor?
- (5) A função quicksort produz uma ordenação estável?
- (6) (recursão em cauda) Mostre que a segunda invocação da função quicksort pode ser eliminada se trocarmos o if por um while apropriado.
- (7) Escreva uma versão do algoritmo quicksort que rearranje uma lista duplamente encadeada de modo que ela fique em ordem crescente. Sua função não deve alocar novas células na memória.



Faça uma versão do QuickSort que seja boa para quando há muitos elementos repetidos no vetor.

• A ideia é particionar o vetor em três partes: menores, iguais e maiores que o pivô





- Provar: No início de cada iteração do laço for valem os seguintes invariantes:
 - (a) A[p..r] é uma permutação do vetor original,
 - (b) $A[p..j-1] \le c < A[j..k-1]$,
 - (c) A[r] = c
- (d) $p \leq j \leq k$.



FIM