Divisão e Conquista

SCC 218 - Alg. Avançados e Aplicações

Esta aula não é sobre:

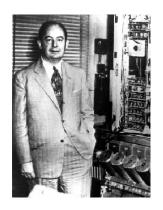
- Quicksort
- Mergesort
- Busca binária (não o modelo tradicional de busca em um vetor ordenado!)

- Aqui veremos estratégias interessantes para resolver problemas baseado no paradigma de divisão e conquista:
 - Sim... vamos repartir o problema em "pedaços" menores de forma a reduzir a complexidade e, no final, obter o resultado desejado
 - \circ Nossa esperança: o que era O(n^2), agora passa a ser O($n \log n$) !!!

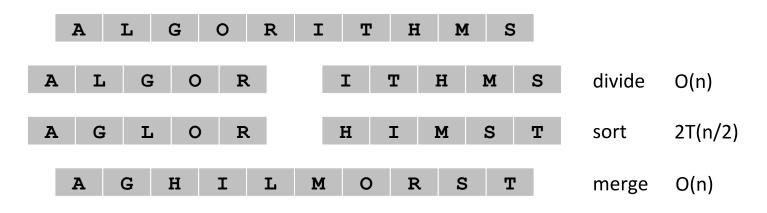
Mas caso vc se esqueceu...!

• Mergesort:

- divida o vetor em duas metades
- o recursivamente ordene cada metade
- o mescle ambas metades de forma a obter o resultado!



Jon von Neumann (1945)

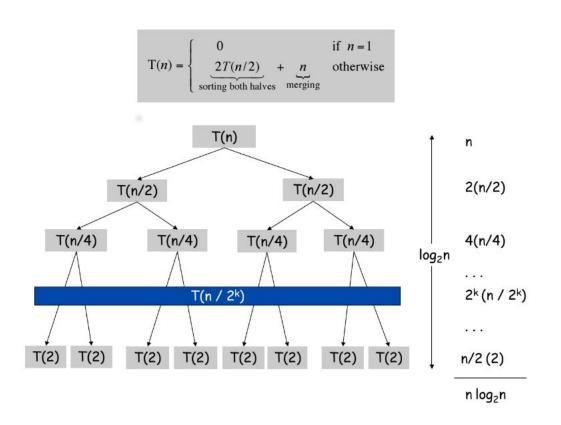


Complexidade - Divisão e Conquista

- Seja T(n) o tempo de execução do pior caso para uma entrada n.
- Se n é par, a estratégia é de ordem O(n) para separar os valores em 2 metades
- Leva T(n/2) para resolver cada metade (pior caso, claro...)
- Finalmente, O(n) para combinar as soluções provenientes das 2 **chamadas recursivas**.
- O tempo de execução T(n) satisfaz a seguinte relação de recorrência:

>>>>
$$T(n) \le 2T(n/2) + cn$$
. (para n > 2)
>>>> $T(2) \le c$ informalmente, fazemos: $T(n) \le 2T(n/2) + O(n)$

Prova por Árvore de Recursão



Complexidade: Divisão e Conquista

- nível 0: problema de tamanho n: c(n) + recursões...
- nível 1: 2 problemas tam = cn/2: 2 cn/2 = cn + recursões subsequentes
- nível 2: 4 problemas tam = n/4: 4 cn/4 = cn + recursões....
- qual o padrão?
 - No nível j da recursão, o nro de subproblemas dobra j vezes >>> 2^j
 - \circ O tamanho de cada subproblema encolhe por um fator de 2^{j} . Tam subprob = n / 2^{j}
 - O nível j contribui com um tempo de $2^{j}(cn/2^{j}) = cn$
- Somando todos os níveis da recursão
 - vimos que cn se repete em todos os níveis.
 - O nro de vezes que a entrada deve dividida ao meio de n até 2 é log₂n
- Somando-se cn para todos os log n níveis de recursão dá: O(n log n)

- Análise de ranking é algo muito usado hoje em dia (web). Vários sites usam o que se chama de <u>filtragem colaborativa</u> para tentar casar (dar matching) suas preferências.
 - o livros, filmes, músicas, etc..
 - feito isso, vc recebe uma série de dicas sobre o que consumir.

- Basicamente, o que procuramos aqui é comparar 2 rankings.
 - Eu elenco/"ranqueio" n filmes de que gosto rotulando-os entre 1, ..., n
 - 1 é o de que mais gosto e **n** o de que menos gosto.
 - Agora ordene estes rótulos de acordo com a preferência do outro
 - Basta contar agora quantos pares estão "fora de ordem".

A métrica de similaridade: o número de inversões entre 2 rankings.

- Meu rank: 1, 2, ..., n.
- Outro rank: $a_1, a_2, ..., a_n$.
- Filmes i e j invertidos se i < j, mas a_i > a_i.

Filmes

	Α	В	С	D	E	
Eu	1	2	3	4	5	
Outro	1	3	4	2	5	

<u>Inversões</u> 3-2, 4-2

- Um algoritmo força bruta para isso seria como
- E qual a complexidade?

- Um algoritmo força bruta para isso teria que complexidade?
- Força Bruta: verifique todos $\Theta(n^2)$ pares i, j.

Suponha o vetor:

1 5 4 8 10 2 6 9 12 11 3 7

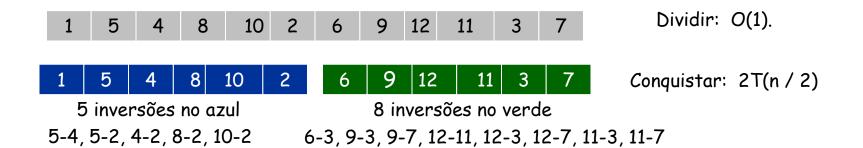
Suponha o vetor:

• Dividir: separar a lista em duas metades.



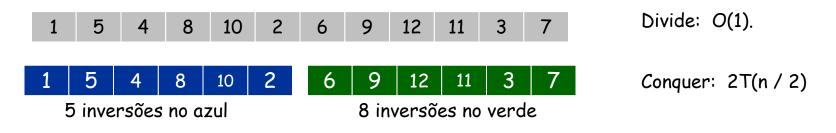
Suponha o vetor:

- Dividir: separar a lista em duas metades.
- Conquistar: recursivamente, contar as inversões em cada metade.



Suponha o vetor:

- Dividir: separar a lista em duas metades.
- Conquistar: recursivamente, contar as inversões em cada metade.
- Combinar: contar inversões para a_i and b_i em metades distintas, e retorne a soma das 3 quantidades.



Combinar: ???

9 inversões azul-verde

5-3, 4-3, 8-6, 8-3, 8-7, 10-6, 10-9, 10-3, 10-7

Total = 5 + 8 + 9 = 22.

Contando inversões: Combinar

Combinar: contar inversões em ambas metades

- Assuma que cada metade está ordenada.
- a_i e b_i estão em metades distintas.
- Combine as duas metades ordenadas em um único vetor ordenado.

isso preserva invariância quanto à ordenação!





13 inversões azul-verde: 6 + 3 + 2 + 2 + 0 + 0

$$T(n) \leq T\Big(\left\lfloor n/2\right\rfloor\Big) + T\Big(\left\lceil n/2\right\rceil\Big) + O(n) \implies \mathsf{T}(n) = O(n\log n)$$

Contando Inversões: Implementação

Pré-condição. [Merge-and-Count] A e B ordenados.

Pós-condição. [Sort-and-Count] Lordenado.

```
Sort-and-Count(L) {
   if list L has one element
      return 0 and the list L
   Divide the list into two halves A and B
   (r_{\lambda}, A) \leftarrow Sort-and-Count(A)
   (r_B, B) \leftarrow Sort-and-Count(B)
   (r , L) ← Merge-and-Count(A, B)
   return r = r_A + r_B + r and the sorted list L
```

• O Problema. Dados n pontos em um plano, encontre o par com a menor distância Euclidiana entre eles, ou seja, encontre o par mais próximo!

- O Problema. Dados n pontos em um plano, encontre o par com a menor distância Euclidiana entre eles, ou seja, encontre o par mais próximo!
- Aplicações em Geometria Computacional
 - CG
 - Visão Computacional
 - GIS (Geographic Information Systems)
 - Modelagem Molecular

- Qual o algoritmo força bruta e sua complexidade ?
- E qual seria o algoritmo e complexidade se o problema fosse 1D, isto é, todos os pontos estivessem em uma única linha?

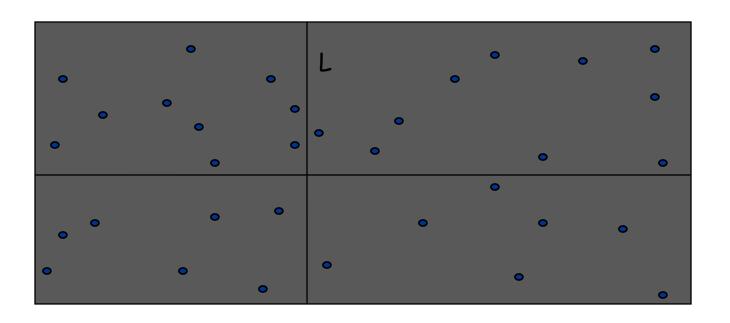
- O Problema. Dados n pontos em um plano, encontre o par com a menor distância
 Euclidiana entre eles, ou seja, encontre o par mais próximo!
- Aplicações em Geometria Computacional
 - CG
 - Visão Computacional
 - GIS (Geographic Information Systems)
 - Modelagem Molecular

- Qual o algoritmo força bruta e sua complexidade?
 - \circ Verifique todos os pares de pontos p e q em $\Theta(n^2)$ comparações.
- E qual seria o algoritmo e complexidade de o problema fosse 1D, isto é, todos os pontos estivessem em uma única linha?
 - Ordene o pontos O(n log n). A seguir, percorra a lista calculando a distância de um ponto com o próximo. Devolva o valor mínimo.

- Notação
 - P é um conjunto de pontos $\{p_1, ..., p_n\}$, onde p_i tem coordenada (x_i, y_i)
 - Sejam p_i e p_i pertencentes a P.
 - \bullet d(p_i,p_i) é a distância Euclidiana entre eles
 - Assuma que não há 2 pontos com as mesmas coordenadas y ou coordenadas x;
 - Esta restrição pode ser eliminada se aplicarmos uma rotação aos pontos de forma a tornar esta afirmação verdadeira e usar o algoritmo apresentado!

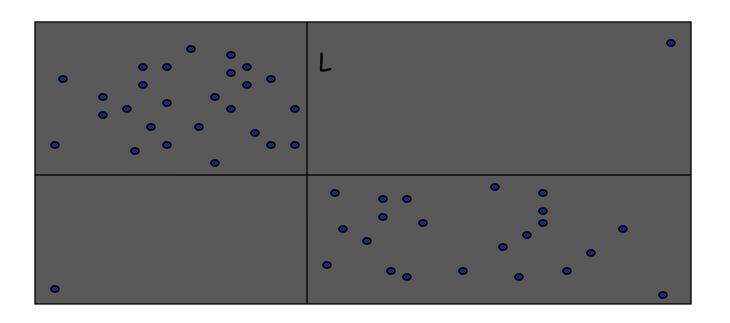
Pares de Pontos mais Próximos: Primeira Tentativa

Divisão. Sub-divida a região em 4 quadrantes



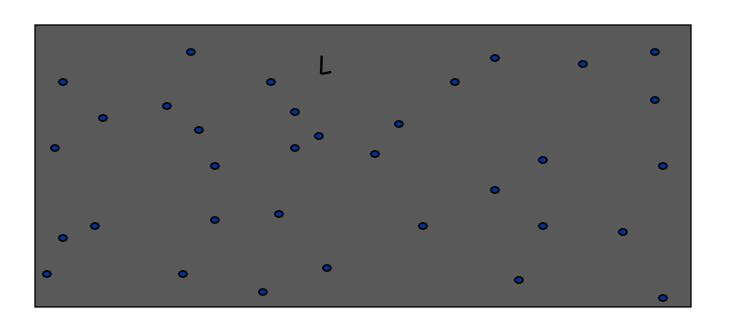
Pares de Pontos mais Próximos: Primeira Tentativa

Divisão. Sub-divida a região em 4 quadrantes Obstáculo. Impossível garantir n/4 pontos em cada parte.



Algoritmo.

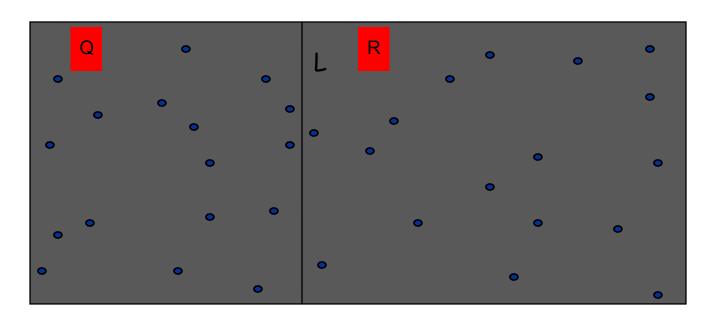
- . Antes da recursão: Ordenar os pontos em função do eixo \mathbf{x} . (Lista P_{x}). Veremos que será preciso uma lista P_{y} , ordenada pelo eixo \mathbf{y} .
 - As listas deverão guardar as posições de cada ponto....



Algoritmo.

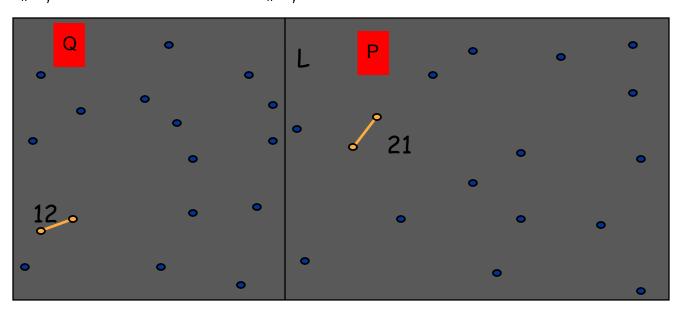
. Divisão: desenhe uma linha vertical L de forma que aproximadamente metade dos pontos estejam em cada lado. Utilize $P_{_{\rm x}}$ para isso.

Crie as listas
$$Q_x$$
, Q_y , R_x e R_z >> $O(n)$



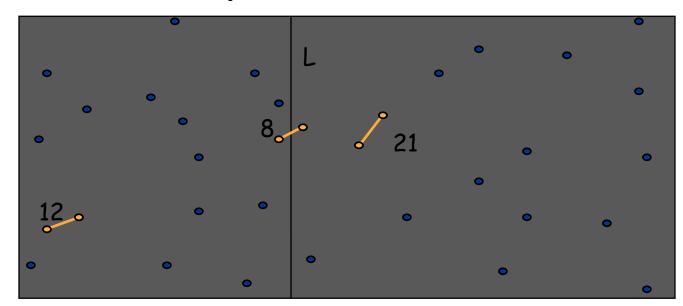
Algoritmo.

- Divisão: desenhe uma linha vertical L de forma que aproximadamente metade dos pontos estejam em cada lado.
- Conquista: Recursivamente, encontre o par mais próximo para cada lado.
 - \circ (q_x, q_v) par mais próximo em Q. (p_x, p_v) par mais próximo em P.

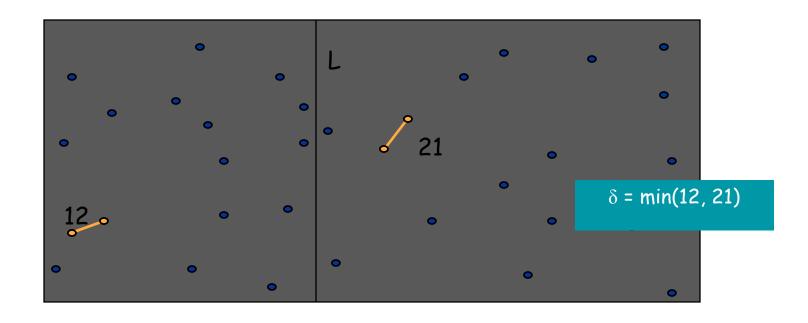


Algoritmo.

- Divisão: desenhe uma linha vertical L de forma que aproximadamente metade dos pontos estejam em cada lado.
- Conquista: encontre o par mais próximo para cada lado recursivamente.
- União: encontre o par mais próximo com um ponto em cada lado \leftarrow parece $\Theta(n^2)$
- Retorne a melhor das 3 soluções.

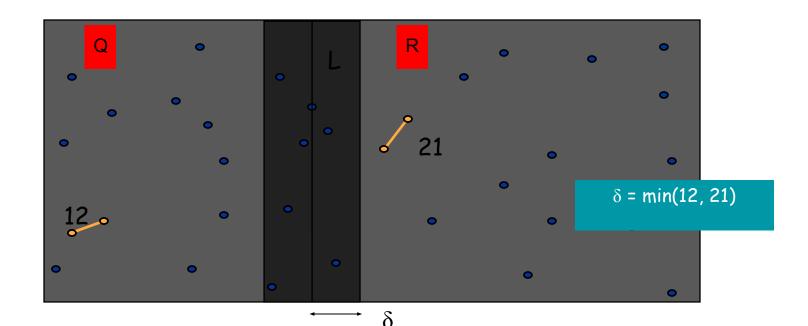


Seja
$$\delta$$
 = min (d((q_x, q_y)), d((p_x, p_y)))



Seja
$$\delta$$
 = min (d((q_x, q_y)), d((p_x, p_y)))

• Observação: somente é preciso considerar os pontos dentro de δ da linha L. Obs: L é a reta posicionada em x*, ponto mais à direita de Q.



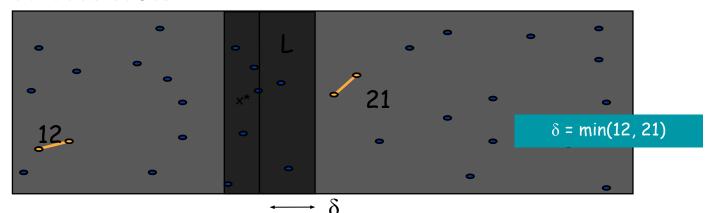
- Se existe $q \in Q$ e $r \in R$ tal que $d(r,q) < \delta$, então cada q e r está a uma distância não superior a δ de L
- Prova:

$$\circ$$
 q = (q_x, q_y) e r = (r_x, r_y). Por definição, sabemos que q_x \le x* \le r_x. Então temos:

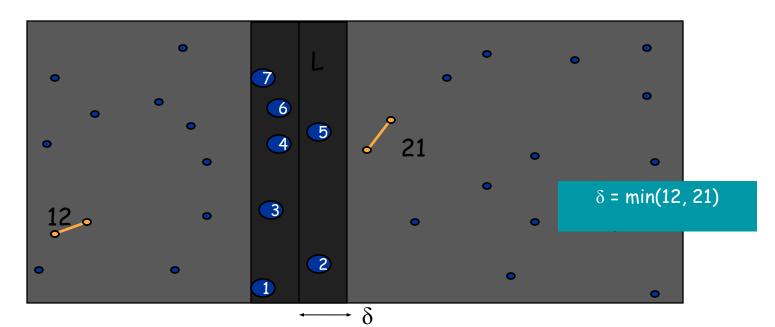
$$x^* - q_x \le r_x - q_x \le d(r,q) < \delta$$
 e também

$$r_x - x^* \le r_x - q_x \le d(r,q) < \delta$$

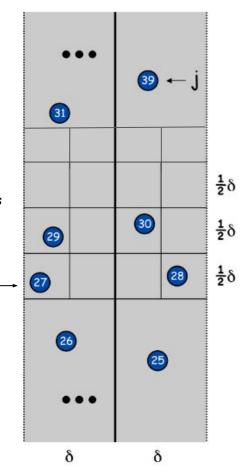
Portanto, cada q e r tem uma coordenada x dentro do limite δ em x* que é o mesmo que dizer que está a uma distância δ de L.



- Portanto, podemos restringir a busca de r e q entre os pontos que estão em uma **porção da faixa** P, que é a faixa escura na figura abaixo... (pontos 1, 2, 3,...,7).
- Quantos pontos pode haver nesta faixa?
- Mas por quanto pontos precisamos procurar???
 - Seja S \subseteq P, esta região em que os pontos estão ordenados crescentemente pela coord. y. Usando a lista P_{v} , isso pode ser feito em O(n)...
 - Se existe um par de pontos (q,r) cuja distância $< \delta$, somente se existe s, s' \in S



- Seja s_i um ponto na faixa 2δ , sendo i^a a menor coordenada y. (ponto 27)
- Se s_i , $s \in S$ e $d(s_i$, $s) < \square$, então s_i , s estão a no máximo 15 posições um do outro, na lista S_v .
- Observe a figura ao lado...
- Sobre a linha L, desenhamos quadrados de lado = δ/2. Uma linha consiste de 4 caixas cujos lados horizontais têm a mesma coordenada ^{2 linhas} y.
- Não há dois pontos numa mesma caixa, pois isso contradiz nossa definição de ☐ mínimo em Q ou em R.
- Suponha que s_i e s_j tem $d(s_i$ e s_j) < \square e que estejam a 16 posições de distância em S_{ij} .
- Lembre que s_i tem a menor coord. y. Como há apenas um ponto por caixa, deve haver ao menos 3 faixas de caixas entre s_i e s_i.
- Mas qq par de pontos separados por 3 linhas de caixas deveriam estar a uma distância de no mínimo 3 □/2 - uma contradição



Algoritmo

```
Closest-Pair(p<sub>1</sub>, ..., p<sub>n</sub>) {
   Compute separation line L such that half the points
   are on one side and half on the other side.
                                                                                O(n \log n)
   \delta_1 = Closest-Pair(left half)
   \delta_2 = Closest-Pair(right half)
                                                                                2T(n / 2)
   \delta = \min(\delta_1, \delta_2)
   Delete all points further than \delta from separation line L
                                                                                O(n)
   Sort remaining points by y-coordinate.
                                                                                O(n \log n)
   Scan points in y-order and compare distance between
                                                                                 O(n)
   each point and next 11 neighbors. If any of these
   distances is less than \delta, update \delta.
   return \delta.
```

```
Closest-Pair(P)
  Construct P_x and P_v (O(n \log n) time)
  (p_0^*, p_1^*) = \text{Closest-Pair-Rec}(P_x, P_y)
Closest-Pair-Rec (P_x, P_y)
  If |P| \leq 3 then
    find closest pair by measuring all pairwise distances
  Endif
  Construct Q_x, Q_y, R_x, R_y (O(n) time)
  (q_0^*, q_1^*) = \text{Closest-Pair-Rec}(Q_x, Q_y)
  (r_0^*, r_1^*) = \text{Closest-Pair-Rec}(R_x, R_y)
  \delta = \min(d(q_0^*, q_1^*), d(r_0^*, r_1^*))
  x^* = maximum x-coordinate of a point in set Q
  L = \{(x,y) : x = x^*\}
  S = points in P within distance \delta of L.
  Construct S_v (O(n) time)
  For each point s \in S_{\nu}, compute distance from s
     to each of next 15 points in Sy
      Let s, s' be pair achieving minimum of these distances
      (O(n) \text{ time})
  If d(s,s') < \delta then
      Return (s,s')
  Else if d(q_0^*, q_1^*) < d(r_0^*, r_1^*) then
      Return (q_0^*, q_1^*)
         Else
            Return (r_0^*, r_1^*)
         Endif
```

Pares de Pontos mais Próximos: Análise

Tempo de execução.

$$T(n) \le 2T(n/2) + O(n \log n) \Rightarrow T(n) = O(n \log^2 n)$$

Q. Podemos calcular em O(n log n)?

Resp. Sim. Não ordenar pontos na faixa a partir do zero a cada vez.

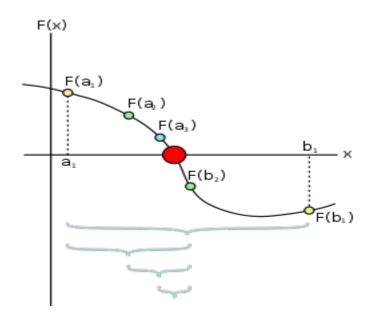
- Cada recursão retorna duas listas: todos os pontos ordenados pela coordenada y, e todos os pontos ordenados pela coordenada x.
- · Ordene por merging as duas listas pré-ordenadas.

$$T(n) \le 2T(n/2) + O(n) \implies T(n) = O(n \log n)$$

Problemas resolvidos com conceitos de busca binária

- Suponha que você precisa pagar um empréstimo com juros. Por exemplo, um empréstimo de \$1000, em dois meses, com juros de 10%.
- Qual valor deve ser pago para que, ao final de dois meses, o saldo seja zero?
 - \$1000 x 1.1 \$576.19 = \$523.81
 - \circ \$523.81 x 1.1 \$576.19 = \$0
- Como obter o valor de \$576.19?

- É preciso estimar um intervalo [a, b] onde a raiz se encontra
- f(a) e f(b) devem ter sinais opostos.
- f deve ser contínua.
- No problema citado, qual é f??
- •f(1000, 2, 10%, d) = ??



а	b	d = (a+b) / 2	f(1000,2,10%,d)	ação
0.01 550.005	1100.00 1100.00	550.005 825.0025	54.9895 -522.50525	aumentar diminuir
550.005	825.0025	687.50375	-233.757875	diminuir
		 576.190476	 error < ε	 parar

$$f(1000,2,10\%,d) = 1000 \times 1.1 - d + (1000 \times 1.1 - d) \times 1.1 - d$$

Complexidade: $O(\log_2(b-a)/\epsilon)$

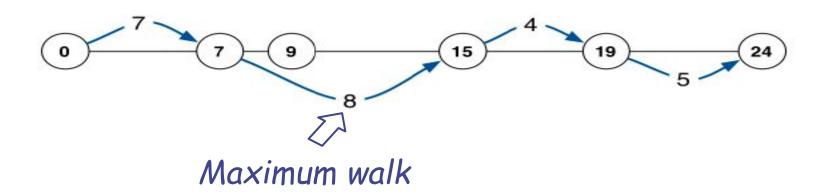
```
double bisection(double lo, double hi){
  while (hi - lo > eps){
    double center = (lo + hi) / 2;
    if (f(lo) * f(center) <= 0){
       hi = center;
    } else {
       lo = center;
    }
}
return (lo+hi)/2;
}</pre>
```

This Winter we are going on a trip along the Appalachian Trail. The trail is a continuous marked footpath that goes from Katahdin in Maine to Springer Mountain in Georgia, a distance of about 2160 miles. Even though our trip will only consider some part of the trail, it will be our first real backpacking experience and an excellent opportunity to acquire winter camping skills.

Part of the experience is also the route planning of the trip. We have a list of all possible campsites ($N \le 600$) that we can use along the way and we want to do this trip so that we only stop K nights to camp ($K \le 300$). We also know in advance the distance between consecutive campsites and we are only allowed to camp at a campsite.

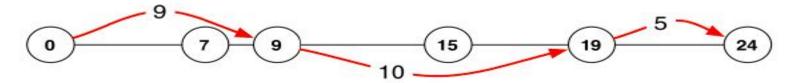
Our goal is to plan the trip so that we <u>minimise the maximum amount of walking done in a single day</u>. In other words, if our trip involves 2 nights (3 days of walking), and we walk 9, 10, 5 miles on each day respectively, the cost (maximum amount of walking done in one day) is 10. Another schedule that involves walking 9, 6, 9 miles on each day has cost 9.

• Por exemplo, para K = 3 e os acampamentos a seguir, a resposta é 8.



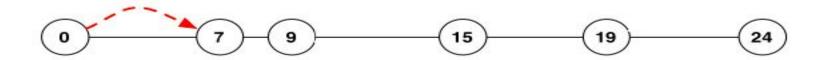
• Iremos fazer uma busca binária no tamanho do passo *P. P* pode assumir valores entre 0 e a maior distância, que neste caso é 24.

Sendo assim, a primeira inspeção é para P = 12.

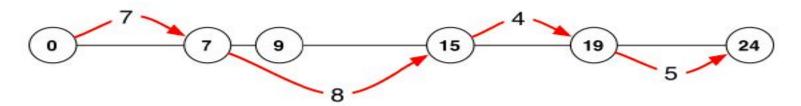


• Para P = 12 é possível chegar ao destino com 2 paradas, logo, o próximo intervalo passa a ser [0, 11].

• Para o intervalo de [0, 11], o elemento central é 5. Como com 5 não e possível chegar ao fim, o intervalo é atualizado para [6, 11].



• Para o intervalo [6, 11], o elemento central é 8. Com 8 e possível chegar ao fim em 3 paradas, logo, o intervalo é atualizado para [6, 7].



• Nas iterações seguintes todos os elementos inspecionados falham. Sendo assim, a nossa resposta é 8 (o último elemento que passou pela verificação).