



Corso Luigi Einaudi, 55 - Torino

Appunti universitari

Tesi di laurea

Cartoleria e cancelleria

Stampa file e fotocopie

Print on demand

Rilegature

NUMERO : 122

DATA : 01/07/2011

A P P U N T I

STUDENTE : Regis

MATERIA : Algoritmi e Programmazione Avanzata

Il presente lavoro nasce dall'impegno dell'autore ed è distribuito in accordo con il Centro Appunti.

Tutti i diritti sono riservati. È vietata qualsiasi riproduzione, copia totale o parziale, dei contenuti inseriti nel presente volume, ivi inclusa la memorizzazione, rielaborazione, diffusione o distribuzione dei contenuti stessi mediante qualunque supporto magnetico o cartaceo, piattaforma tecnologica o rete telematica, senza previa autorizzazione scritta dell'autore.

**ATTENZIONE: QUESTI APPUNTI SONO FATTI DA STUDENTIE NON SONO STATI VISIONATI DAL DOCENTE.
IL NOME DEL PROFESSORE, SERVE SOLO PER IDENTIFICARE IL CORSO.**

```
int f();    DICHIARAZIONE PROTOTIPO
int f() {   DEFINIZIONE FUNZIONE
  .....
}
```

[] parte intera approssimata per difetto } non influiscono sul comp. asintotico
[] parte intera approssimata per eccesso }

CLASSI di MEMORIZZAZIONE

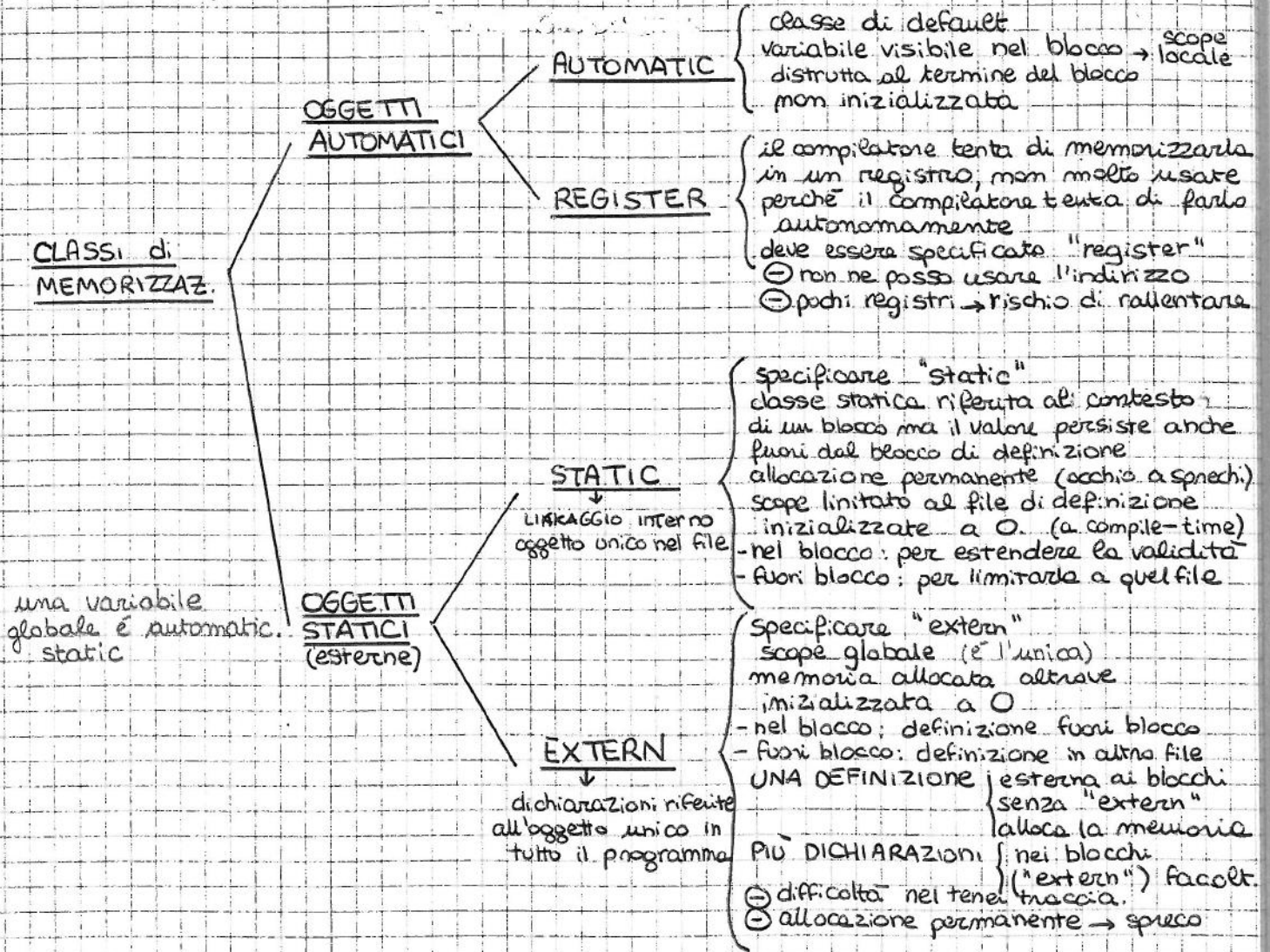
VARIABLE < classe di memorizzazione > < tipo > < nome >

variabile locale : definita in un blocco

variabile globale : definita fuori da un blocco (nella parte di codice successiva)

SCOPE (ambiente) : porzione di programma in cui la variabile è utilizzabile

locale o globale in base alla classe di memorizzazione



PREPROCESSORE : elabora il file sorgente prima di passarlo alla compilazione.

- definizione di MACRO (costanti o nomi simbolici) o MACRO CON ARGOMENTI, simili a chiamate a funzione ⊕ prestazione: espande il codice in linea così che non aumenta la dimensione del programma, ⊕ adattabili a qualunque tipo
 - ⊖ l'assenza di tipi può causare errori difficili da trovare
 - ⊖ problemi quando si passano espressioni: è necessario specificare parentesi.

• inclusione condizionale

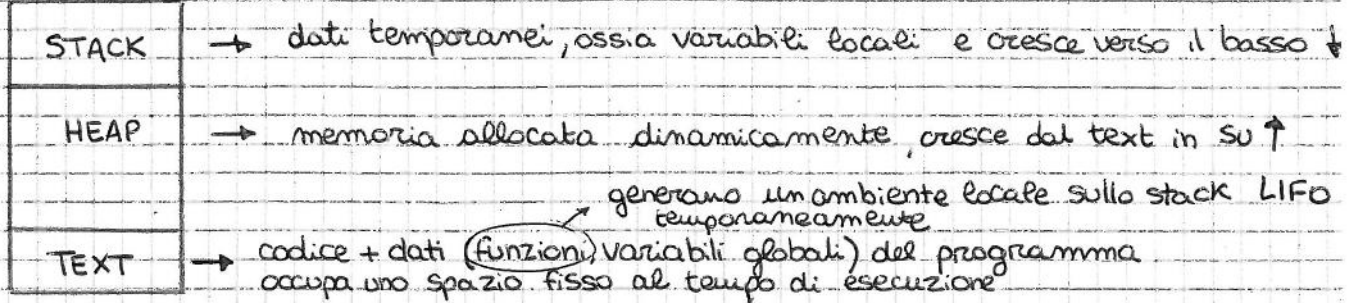
# if defined (<nome macro>)	eq.	# ifdef (nome)	# elif
# if <nome> == <valore>	eq.		# else
# if !defined (HEADER)	eq.	# ifndef (nome)	
# define HEADER		# endif	

MEMORIA DINAMICA

ALLOCAZIONE STATICA di MEMORIA: tramite dichiarazione di variabili

ALLOCAZIONE DINAMICA: su richiesta, per quantità definite al tempo di esecuzione (assegna un'area della memoria fisicamente separata da quella riservata alle variabili dichiarate staticamente).

SPAZIO di INDIRIZZAMENTO



void *malloc (size_t <size>) ^{al 1° byte} ritorna puntatore a area pari a "size" byte di contenuto indefinito con operatore cast. È possibile allocare qualunque tip

void free (void * <pointer>) libera memoria prima allocata con la malloc poi annulla il puntatore pointer = NULL

```

int *p;
p = malloc (sizeof(int));
free (p);
p = NULL;
X = *p;
    
```

OPERATORE di CAST
 p = (int *) malloc (sizeof(int));

NO!
 p ora punta a un'area di memoria che non è più nostra

area di memoria non valida! inoltre se non abbiamo annullato p c'è il rischio di accedere a un'area di memoria che può essere stata assegnata ad altri.

Fallimento della malloc: restituisce NULL se:

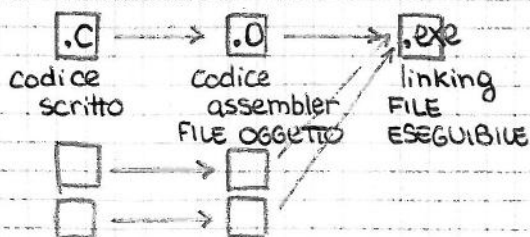
- non c'è più lo spazio richiesto disponibile in memoria;
- lo spazio ancora disponibile è frammentato (allora alloco multipli di un m piccolo o faccio dump su disco).

! È necessario un controllo sulla malloc

MODULARITÀ

il C consente la gestione di più moduli (file sorgenti) per maneggevolezza

→ MAIN + moduli che costituiscono la libreria.



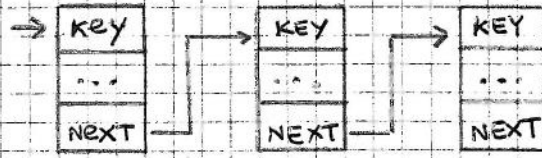
```

.H  dichiarazione prototipi # ifndef (.h)
    # define (.h)
    # endif
.C  contiene le definizioni (#include ".h"
MAIN #include ".h"
    
```

LISTE LINEARI:

- elementi sistemati secondo ordine lineare determinato da puntatori

```
struct item {
    int key;
    struct item *next;
};
```



le aree di memoria non sono contigue come nei vettori

- individuata dal puntatore alla testa (e opzionalmente una coda)
- uso di struct ricorsive che vengono create così:

```
{ struct item *h;
  h = (struct item *) malloc(sizeof(struct item));
  h -> key = 0;
  h -> next = NULL;
}
```

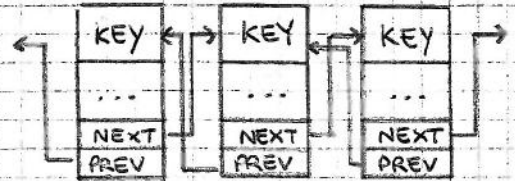
inizializzazione elemento generico e isolato.

LISTE LINKATE: lista di elementi dello stesso tipo connessi da puntatori;

LISTA SEMPLICE: quella rappresentata prima ⊕ semplice ⊖ solo passaggi avanti;

LISTA DOPPIO LINK: ogni elemento possiede puntatore a next e a previous

⊖ più complessa ⊕ consente passaggio avanti/ind

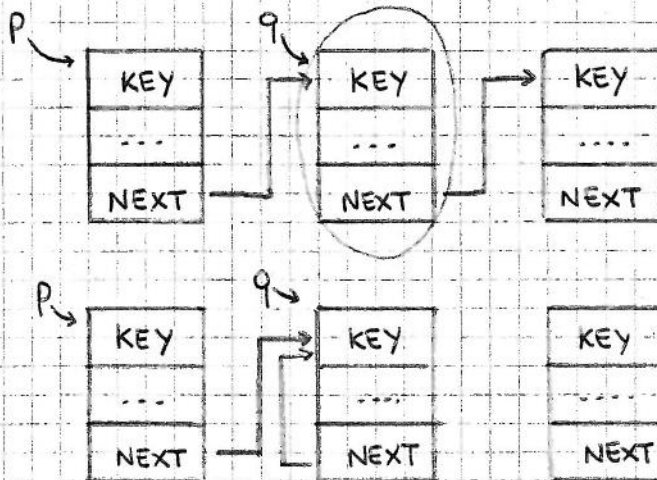


LISTA CON SENTINELLE: esiste un elemento fittizio corrispondente a lista vuota per gestire i casi speciali tipo lista vuota → evita il controllo head == NULL.

⊕ semplifica algoritmo e libera puntatore ⊖ occupa memoria (è necessario?)

LISTA CIRCOLARE: la coda si richiude sulla testa.

NB:



```
struct item *q;
```

```
q = (item *) malloc(sizeof(item));
```

per inserire un elem. devo invertire le operazioni!

```
p -> next = q;
q -> next = p -> next;
```

abbiamo creato una lista chiusa circolare ma attenzione! abbiamo perso un elemento! che ora non è puntato da alcun puntatore → perso aree di memoria e dati in essa contenuti per averla circolare

```
q -> next = p;
```

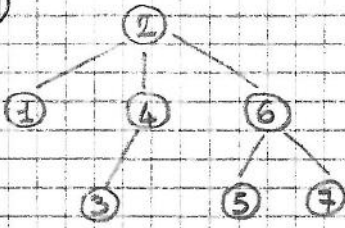
LISTA ORDINATA: elementi dati in ordine crescente/decrecente.

Primitive:

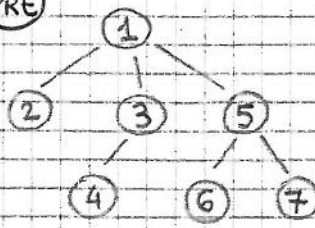
VISITA
 ↳ INORDER figlio sx, nodo, figlio dx
 ↳ PREORDER nodo, figlio sx, figlio dx
 ↳ POSTORDER figlio sx, figlio dx, nodo

SEARCH, INSERT, DELETE costruite adattando la visita all'operazione.

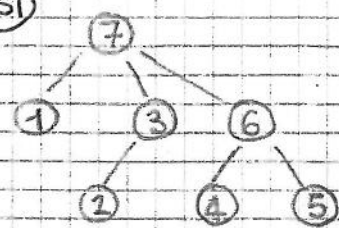
(N)



(PRE)



(POST)



```
void inorder (*head)
{ if (head == NULL) return;
  inorder(head->left);
  visit(head);
  inorder(head->right); }
```

```
void preorder (*head)
{ if (head == NULL) return;
  visit(head);
  preorder(head->left);
  preorder(head->right); }
```

```
void postorder (*head)
{ if (head == NULL) return;
  postorder(head->left);
  postorder(head->right);
  visit(head); }
```

HEAP

Albero binario in cui ogni figlio ha valore \leq al valore del padre

serve per implementare code prioritarie (con priorità associata al campo chiave).

È un albero quasi completo:
 ↳ sbilanciamento max = 1 livello
 ↳ completo su tutti i livelli tranne eventualmente sull'ultimo riempito da sinistra.

PROPRIETÀ:

- elemento maggiore nella radice; (max heap)
- ogni nodo interno ha valore \geq della discendenza;
- l'altezza di uno heap di n elementi è $\Theta(\log_2(n))$ SHAPE PROPERTY tutti i livelli completi eccetto ultimi

Rappresentabile come albero binario o come vettore (figli di i in $2i$ e $2i+1$)

length = dimensione max vettore
heapsize = n elementi dello heap

PRIMITIVE:

HEAPIFY

RICORSIVA

forza l'applicazione della proprietà dello heap; ricorsiva:
 dato l'elemento i
 ↳ i due sottoalberi di i sono heap;
 ↳ i può essere < dei figli (violazione proprietà!)
 ↳ fa scendere A[i] nella prima posizione utile
 tempo di esecuzione $O(h_i)$ con h_i altezza del nodo i (foglie: livello 0)
 ↳ complessità $O(\log_2 m)$

(prima confronta i due figli e sceglie il maggiore che poi viene confrontato col padre e scambiato se è maggiore del predecessore).

Si può usare la heapify bottom-up per convertire un vettore in uno heap.

ANALISI di COMPLESSITÀ di ALGORITMI

La complessità di un algoritmo deve essere valutata in termini di:
 tempo di calcolo $T(m)$ o di memoria occupata $S(m)$.

L'incidenza sul tempo e sulla memoria cresce al crescere della dimensione m dei dati in ingresso, quindi si conduce un'ANALISI ASINTOTICA, per $m \rightarrow \infty$.

L'analisi effettuata deve essere indipendente dal tipo di calcolatore e assumiamo che sia anche indipendente dai valori dei dati in ingresso.

Notazione asintotica (definisce una famiglia di funzioni)

Θ TETA GRANDE

$$\Theta(g(m)) = \{ f(m) : \exists c_1, c_2 > 0 \text{ e } m_0 \geq 0 \text{ t.c. } c_1 g(m) \leq f(m) \leq c_2 g(m) \forall m > m_0 \}$$

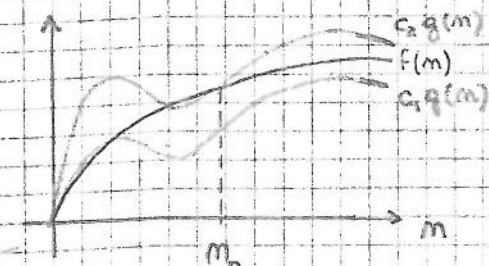
si dice $f(m) \in \Theta(g(m))$ o $f(m) = \Theta(g(m))$

anche se l'uguaglianza non è proprio stretta.

$$f(m) = am^2 + bm + c, \quad f(m) = \Theta(m^2)$$

dato un polinomio $p(m)$ di grado d $p(m) = \Theta(m^d)$.

$$\text{costanti} = \Theta(m^0) = \Theta(1)$$



sia O che $\Omega(g(m))$

esempio. $\frac{1}{2}m^2 - 3m = \Theta(m^2)$.

Devo trovare $c_1, c_2, m_0 \geq 0$ t.c. $c_1 m^2 \leq \frac{1}{2}m^2 - 3m \leq c_2 m^2 \forall m \geq m_0$

divido i due membri per m^2 : $c_1 \leq \frac{1}{2} - \frac{3}{m} \leq c_2$

$$\frac{1}{2} - \frac{3}{m} \leq c_2 \quad \forall m \geq 1 \text{ e } c_2 \geq \frac{1}{2}$$

$$\frac{1}{2} - \frac{3}{m} \geq c_1 \quad \forall m \geq 7, \text{ e } c_1 \leq \frac{1}{14} \quad \frac{7-6}{14} = \frac{1}{14} \leq \frac{1}{14}$$

$$\rightarrow c_1 = \frac{1}{14}, c_2 = \frac{1}{2}, m_0 = 7 \rightarrow \frac{1}{2}m^2 - 3m = \Theta(m^2)$$

Sicuramente esistono molte altre possibilità di scelta, ma ciò che conta è che almeno qualcuna esista.

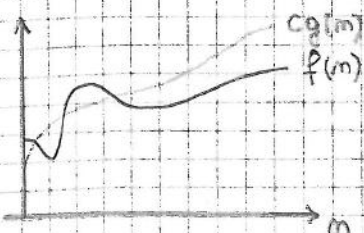
esempio. $6m^3 \neq \Theta(m^2)$

supponiamo che esistano c_2 e m_0 tali che $6m^3 \leq c_2 m^2, \forall m \geq m_0$

$\rightarrow m \leq \frac{c_2}{6}$ che è impossibile per un m arbitrariamente grande dato che c_2 è una costante arbitraria.

O GRANDE

$$O(g(m)) = \{ f(m) : \exists c > 0 \text{ e } \exists m_0 \geq 0 \text{ t.c. } \forall m > m_0 \quad f(m) \leq c g(m) \}$$



$$\Theta(g(m)) \subseteq O(g(m))$$

qualsunque funzione lineare $am+b = O(m^2)$

facilmente verificabile prendendo $c = a+|b|$ e $m_0 = 1$

Θ piccolo

limite superiore che non è asintoticamente stretto. $2m = \Theta(m^2)$

Esempi:

1. RICERCA LINEARE

```
for i=1 to m do
  if A[i] = Key return trovato;
return non trovato;
```

Caso migliore: key è il primo elemento $T(m) = O(1)$.

Caso peggiore: Key è l'ultimo elemento $T(m) = O(m)$ complessità lineare

Caso medio: assumiamo una distribuzione equiprobabile $P_i = \frac{1}{m}$
 $T(m) = \sum \frac{1}{m} i = \frac{1}{m} \sum i = \frac{1}{m} \frac{m(m+1)}{2} = \frac{m+1}{2} = O(m)$

2. INSERTION SORT ($t_j =$ volte che l'istruzione viene ripetuta)

```
for i=2 to m do
  key = A[j] : inserisci A[j] nella sequenza spostando
              a destra gli elementi maggiori di A[j]
  i = j-1
  while i > 0 and A[i] > key do
    A[i+1] = A[i]
    i = i-1
  A[i+1] = key
  j = j+1
```

COSTO	N. VOLTE
C_1	m
C_2	$m-1$
C_4	$m-1$
C_5	$\sum t_j (j=2..n)$
C_6	$\sum (t_j - 1)$ "
C_7	$\sum (t_j - 1)$ "
C_3	$m-1$
C_9	$m-1$

$T(m) = f + e \cdot m + g \sum_{j=2}^m t_j$

Caso migliore: vettore già ordinato: $t_j = 1 \forall j \rightarrow T(m) = f + e \cdot m + g(m-1) = \Theta(n)$

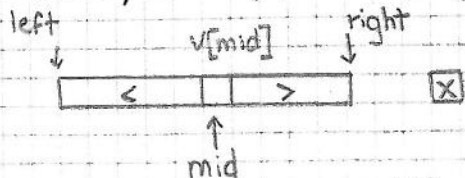
Caso peggiore: vettore in ordine inverso: $t_j = j \forall j \rightarrow T(m) = f + e \cdot m + g \left(\frac{m(m+1)}{2} - 1 \right)$

$T(m) = \Theta(m^2)$

3. RICERCA BINARIA ($v, left, right, x$)

```
if (left > right) return FALSE
mid = (left + right) / 2
if x = v[mid] return TRUE
if (x < v[mid])
  binarysearch(v, left, mid-1, x)
if (x > v[mid])
  binarysearch(v, mid+1, right, x)
```

COSTO	n. VOLTE
C_1	1
C_2	1
C_3	1
Problema con $dim = \frac{n}{2}$?
Problema con $dim = \frac{n}{2}$?



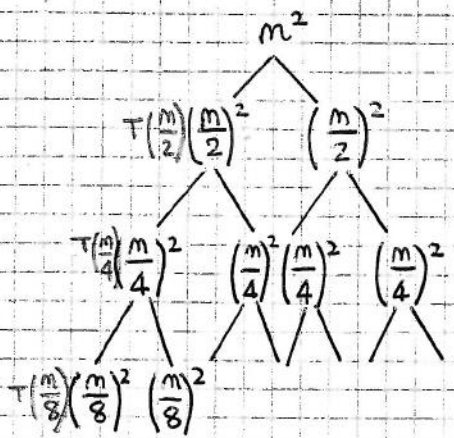
$T(m) = C_1 + C_2 + C_3 + T\left(\frac{m}{2}\right) = a + T\left(\frac{m}{2}\right)$

$T(m) = \Theta(\log_2 m)$

ALBERO della RICORSIONE utile per visualizzare il metodo iterativo

1. si espande livello per livello
 ↳ nodo = costo suddivisione
 ↳ figli = termine ricorsivo.
2. si sommano i valori su ogni livello (valori in totale pari all'altezza dell'albero)
 → "costo complessivo"

esempio. $T(m) = 2T\left(\frac{m}{2}\right) + m^2$



COSTO m^2

$$\left(\frac{m}{2}\right)^2 + \left(\frac{m}{2}\right)^2 = 2 \frac{m^2}{4} = \frac{m^2}{2}$$

$$\left(\frac{m}{4}\right)^2 \cdot 4 = \frac{m^2}{4}$$

$$\left(\frac{m}{8}\right)^2 \cdot 8 = \frac{m^2}{8}$$

...

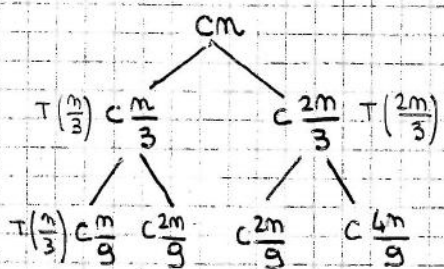
TOTALE = $\theta(m^2)$

m. livelli = $\log_2 m$ ricavato dal caso base come prima.

$$T(m) = \sum_{i=0}^{\log_2 m} \frac{m^2}{2^i} = \sum \frac{m^2}{2^i} = m^2 \sum \frac{1}{2^i} = \theta(m^2)$$

↳ converge a 1

esempio. $T(m) = T\left(\frac{m}{3}\right) + T\left(\frac{2m}{3}\right) + cm$



COSTO cm

cm

cm

TOTALE = $cm \cdot h$

h altezza dell'albero la ricavo dal caso base del ramo più lungo dell'albero dato che questa volta sarà sbilanciato.

$$\left(\frac{2}{3}\right)^i m = 1 \quad \left(\frac{3}{2}\right)^i = m \quad i = \log_{\frac{3}{2}} m$$

$$T(m) = \sum_{i=0}^{\log_{\frac{3}{2}} m} cm = cm \sum_{i=0}^{\log_{\frac{3}{2}} m} 1 = cm \log_{\frac{3}{2}} m = \theta(m \log m)$$

esempio: $T(m) = 2T\left(\frac{m}{2}\right) + m$

$a=2 \quad b=2 \quad f=m$

$m \log_2 b = m \rightarrow$ caso 2: $T(m) = \Theta(m \log m)$

$T(m) = 3T\left(\frac{m}{4}\right) + m^2$

$a=3 \quad b=4 \quad f(m)=m^2$

$m \log_4 3 \rightarrow$ caso 3: $T(m) = \Theta(m^2)$

Ricorrenze lineari

$T(m) = b_1 T(m-1) + b_2 T(m-2) + \dots + b_k T(m-k)$

con k condizioni iniziali per $T(0), \dots, T(k-1)$

Metodo:

1. Si costruisce l'equazione caratteristica $x^k - b_1 x^{k-1} - b_2 x^{k-2} - \dots - b_k = 0$
2. se ne calcolano le soluzioni r_1, \dots, r_k .
3. La soluzione è $T(m) = C_1 r_1^m + C_2 r_2^m + \dots + C_k r_k^m$.
con C_1, \dots, C_k ricavati dalle condizioni iniziali.

esempio: $T(m) = 2T(m-1)$

$T(0) = 1 \quad k=1$

equazione caratteristica $x - 2 = 0 \quad x=2$

$T(m) = C_1 2^m \rightarrow T(0) = 1, \quad 1 = C_1 \rightarrow T_m = 2^m$

$T(m) = T(m-1) + T(m-2)$

$T(0) = 1$

$T(1) = 1$

$k=2 \rightarrow$ eq. caratteristica $x^2 - x - 1 = 0 \rightarrow x = \frac{1 \pm \sqrt{5}}{2}$

$T(m) = C_1 \left(\frac{1+\sqrt{5}}{2}\right)^m + C_2 \left(\frac{1-\sqrt{5}}{2}\right)^m$

$$\begin{cases} C_1 + C_2 = 1 \\ C_1 \left(\frac{1+\sqrt{5}}{2}\right) + C_2 \left(\frac{1-\sqrt{5}}{2}\right) = 1 \end{cases} \quad \begin{matrix} C_1 = 0,723 \\ C_2 = 0,276 \end{matrix}$$

$T(m) \approx \left(\frac{1+\sqrt{5}}{2}\right)^m$

IMPLEMENTAZIONE

```
void selsort (int v[], int m) {
    int i, j, min;
    for (i=0, i<m-1; i++) {
        min=i;
        for (j=i+1; j<m; j++) {
            if (v[j]<v[min])
                min=j;
        }
        swap (v[min], v[i]);
    }
}
```

esempio

7	11	3	4	9	8	2	1
1	11	3	4	9	8	2	7
1	2	3	4	9	8	11	7
1	2	3	4	7	8	11	9
1	2	3	4	7	8	9	11

HEAPSORT (ALGORITMO LOGARITMICO $O(m \log m)$)

Opera come selection sort ma è basato su una struttura dati più efficiente.

1. Buildheap a partire da un vettore;
2. Sconto della radice e modifica dello heap per forzare un ordine (iteraz. heapify

PSEUDOCODICE

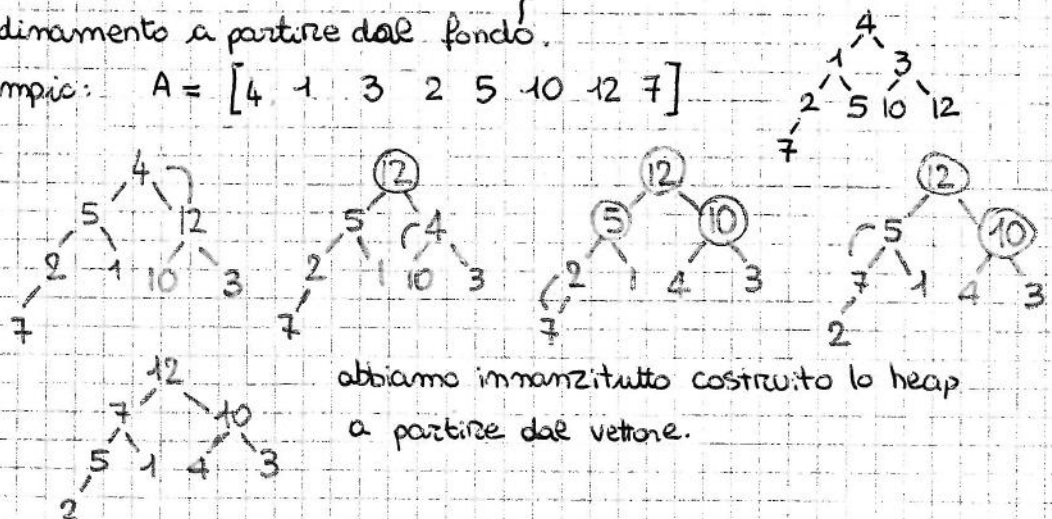
```
Heapsort (A) {
    Buildheap (A, m);
    for i=m to 2 {
        scambia (A[1], A[i]);
        m=m-1;
        Heapify (A, 1);
    }
}
```

IMPLEMENTAZIONE

```
void Heapsort (int v[], int dim) {
    int i;
    heap_size = dim;
    buildheap (v, heap_size);
    for (i=dim; i>=2; i--) {
        swap (v, 1, i);
        heapify (v, 1, --heap_size);
    }
}
```

Si costruisce l'ordinamento a partire dal fondo.

esempio: $A = [4, 1, 3, 2, 5, 10, 12, 7]$



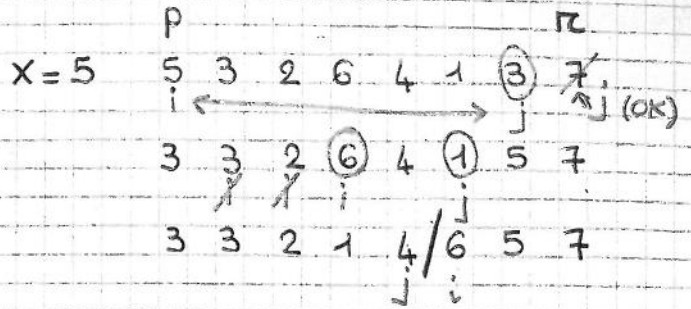
abbiamo innanzitutto costruito lo heap a partire dal vettore.

L'elemento "spartiacque" del vettore viene scelto con algoritmo di partizionamento.
 q sarà l'indice della posizione in cui si troverà alla fine l'elemento minore o uguale al pivot di valore più grande.

IMPLEMENTAZIONE

```

Partition (A, p, r) {
    x = A[p];
    i = p-1;
    j = r+1;
    while {
    
```



```

    l'elemento da dx < x → do j = j-1;
                        while (A[j] >= x);
    1° elemento da sx > x → do i = i+1;
                        while (A[i] <= x);
    se l'array non è stato visto completamente scambiato → if (i < j)
                                                                scambia (A[i], A[j]);
                                                                else return j;
    }
    
```

La scelta avrebbe sicuramente effetti positivi sull'algoritmo, ma costa troppo, così si prende il primo elemento del vettore.

COMPLESSITÀ (PARTITION) = $O(m)$.

Analisi del quicksort

ipotesi bilanciata

CASO MIGLIORE: si producono due partizioni bilanciate.

Diagram showing a binary tree of recursive calls for a balanced partitioning of size n. The tree has levels n-1, n/2-1, n/4-1, ..., 2, 1. The cost of each level is shown as a bar chart.

$T(m) = 2T(\frac{m}{2}) + O(m)$ partition
 costo = $m + \frac{m}{2} \cdot 2 + \frac{m}{4} \cdot 4 + \dots + \frac{m}{m} \cdot m = \sum_{i=0}^{\log_2 m} 2^i \cdot \frac{m}{2^i} = m(\log_2 m + 1)$
 $m^{\log_2 2} = m$; $f(m) = m \rightarrow T(m) = O(m \log_2 m)$

CASO PEGGIORE: si produce una partizione in cui una parte ha un solo elemento mentre l'altra ha i restanti m-1 elementi.

Diagram showing a skewed binary tree of recursive calls for a worst-case partitioning of size m. The tree has levels m-1, m-2, m-3, ..., 2, 1. The cost of each level is shown as a bar chart.

$T(m) = T(m-1) + O(m)$
 costo = $(m-1) + (m-2) + \dots + 2 + 1 = O(m^2)$.
 si ha solo nel caso di sbilanciamento a valore costante $T(n) = T(n-1) + O(n)$
 si ha quando il vettore è già ordinato. In questo caso è meglio l'insertion!

CASO MEDIO: coincide con il caso migliore (fortunatamente)

anche nel caso di sbilanciamento a proporzionalità costante complessità $O(m \log_2 n)$ (per qualunque sbilanciamento tra 10% e 90%).

COUNTING SORT (ALGORITMO LINEARE $O(m)$)

Si assume che si conosca l'intervallo in cui sono compresi i valori $[1, \dots, k]$

COMPLESSITÀ = $O(k+m) = O(m)$ se $k=O(m)$

non usa i confronti per l'ordinamento

Compromesso velocità/memoria: usa 2 vettori ausiliari $\left\{ \begin{array}{l} \text{output ordinato } B \\ \text{vettore conteggi } C \end{array} \right.$
 se $k \gg n$ memoria richiesta può essere eccessiva

1. Determina per ogni elemento il n° di elementi minori di x
2. Usa questa informazione per porre l'elemento al posto giusto

IMPLEMENTAZIONE

Countingsort (A, m) {

for $i=1$ to k | inizializzo il vettore C
 $C[i]=0$

for $i=1$ to m] $C[i]$ contiene tutte le occorrenze
 $C[A[i]] = C[A[i]] + 1;$ | (n° elementi uguali ad $A[i]$)
 e segna n.

for $i=2$ to k]
 $C[i] = C[i] + C[i-1];$ | n° elementi totali che ci sono prima di quello

for $i=m$ to 1

$B[C[A[i]]] = A[i];$ → inserisce nella posizione corretta

$C[A[i]] = C[A[i]] - 1$ } → toglie l'occorrenza dell'elemento inserito

esempio. A 1 2 3 4 5 6 7 8
 3 6 4 1 3 4 1 4

le range in cui sono compresi i valori è $[1, 6]$

C 1 2 3 4 5 6
 0 0 0 0 0 0 (inizializzato)
 +1 +1 +1 +1
 +1 +1 +1
 +1

$A[m] = 4$

$C[A[m]] = C[4] = 7$

C 2 0 2 3 0 1 $C[A[i]] = C[A[i]] + 1$

C 2 2 4 7 7 8 $C[i] = C[i] + C[i-1]$
 ~~2~~ ~~2~~ ~~4~~ ~~7~~ ~~7~~ ~~8~~

a questo punto inizio ad ordinare nel vettore

B 1 2 3 4 5 6 7 8
 1 1 3 3 4 4 4 6

↑
 1° elemento inserito (si mette dal fondo)

(NB) ORDINAMENTO STABILE: preserva l'ordine relativo dei dati con chiavi uguali all'interno della sequenza da ordinare!

Heapsellect applicata alla ricerca del mediano ricade nella complessità dell'ordinamento ($k = \frac{m}{2} = O(m)$) $\rightarrow T(m) = O(m \log m)$

Selezione randomizzata

approccio divide-et-impera simile al quicksort, ma non serve cercare in entrambe le partizioni: basta in una di esse.

Random select. (A, p, r, k) cerca k -esimo elemento più piccolo in A

if ($p=r$) return $A[p]$

$q = \text{randompartition}(A, p, r)$

$s = q - p + 1$ (indice di q in $[p \dots r]$)

if ($k \leq s$)

return Randomselect (A, p, q, k)

else return Randomselect ($A, q+1, r, k-s$)



se $k < s$ è a sinistra,
se no è a destra

COMPLESSITÀ

CASO MIGLIORE $T(m) = T(\frac{m}{2}) + \theta(m) = \theta(m)$ qualunque rapporto x partizione

CASO PEGGIORE $T(m) = T(m-1) + \theta(m) = \theta(m^2)$

CASO MEDIO $T(m) \leq \frac{2}{m} \sum T(k) + \theta(m) = O(m)$

Randompartition = versione partition in cui la scelta pivot è random.

NB. Quando si ricorre sulla partizione destra si cerca l' i -esimo.

MASSIMO e MINIMO: non richiedono confronti per come è stato costruito!

```
Tree_min (T) {
    while (left[T] ≠ NULL)
        T = left[T];
    return T; }
```

```
Tree_max(T) {
    while (right[T] ≠ NULL)
        T = right[T];
    return T; }
```

SUCCESSORE E PREDECESSORE *successore = chiave min sottoalbero dx*

```
Tree_successor(x) {
    if (right[x] ≠ NULL)
        return Tree_min(right[x]);
    y = p[x];
    while (y ≠ NULL and x = right[y])
        x = y;
    y = p[y];
    return y; }
```

se x ha figlio dx, successore è minimo del sottoalbero dx

risale l'albero fino a trovare un nodo che è figlio sx del padre di x

CANCELLAZIONE

- 3 casi:
1. nodo senza figli → elimina
 2. nodo ha un figlio → collega padre del nodo al figlio del nodo poi elimina
 3. nodo ha 2 figli → cerca il successore (che ha solo figlio dx) elimina successore (come caso 2) copia campi valore del successore nel nodo da eliminare.

```
Tree_Delete (T, z) {
    if (left[z] = NULL or right[z] = NULL)
        y = z;
    else y = Successor(z);
    if (left[y] ≠ NULL)
        x = left[y];
    else x = right[y];
    if (x ≠ NULL)
        p[x] = p[y];
    if (p[y] = NULL) root[T] = x;
    else if (y = left[p[y]]) left[p[y]] = x;
        else right[p[y]] = x;
    if (y ≠ z) chiave[z] = chiave[y];
    return y; }
```


TABELLE HASH

Sono tabelle ad accesso diretto, in cui la ricerca non è più basata sul confronto. ~~Costo~~ L'efficienza non è più legata all'universo di dati ($|U|$) quanto piuttosto all'insieme di chiavi effettivamente usate (dimensione = m).

Se fattore di carico $\alpha = \frac{m}{|U|}$ è basso \rightarrow tabelle hash $[|U| = m]$

Si tratta di una generalizzazione di vettori: vettore di puntatori a struct:

- a ogni chiave corrisponde una posizione nella tabella (un indice)
- la tabella è di dimensione $m \ll |U|$ (dim. pari all'universo delle chiavi)
- funzione hash $h: U \rightarrow \{0, \dots, m-1\}$ calcola indice a partire da chiave
indice = $h(k)$.
- elemento chiave k viene memorizzato nella posizione $h(k)$.
- \rightarrow il range degli indici si riduce da $|U|=m$ a m .
- \rightarrow tempi di esecuzione $O(1)$.

Funzione hash non è iniettiva \rightarrow COLLISIONI (chiavi distinte - stesso valore hash)

per ridurre le collisioni:

- scelta della funzione hash;
- tecniche di gestione delle collisioni $\left\{ \begin{array}{l} \text{concatenazione (liste di collisione)} \\ \text{indirizzamento aperto.} \end{array} \right.$

Funzione hash

caratteristiche:

CRITERIO di UNIFORMITÀ SEMPLICE: il valore hash di una chiave k è uno dei valori

$0, \dots, m-1$ in modo equiprobabile $\sum_{k: h(k)=j} P(k) = \frac{1}{m}$
scegliendo una chiave a caso questa finisce nella j -esima cella.

In genere la distribuzione di probabilità delle chiavi P non è nota; nella pratica si usano tecniche euristiche che permettono di creare funzioni hash che si comportano bene con buona probabilità.

Una buona funzione hash deve usare TUTTE le cifre della chiave.

METODO della DIVISIONE

$$h(k) = k \bmod m$$

$m \neq 2^p$, di solito m primo.

parte frazionaria di kA

METODO della MOLTIPLICAZIONE/DIVISIONE

$$h(k) = \left\lfloor m \left(\frac{kA}{m} \bmod 1 \right) \right\rfloor \quad k \in [0, 1]$$

di solito $m = 2^p$ per semplicità

Se si hanno chiavi non numeriche:

caratteri: riconducibili a dati numerici (codice ASCII)

stringhe: rappresentazione come dato numerico in base

Gestione delle collisioni - indirizzamento aperto. ($m \leq m, \alpha \leq 1$)

Tutti gli elementi vengono memorizzati in tabella. La collisione è risolta cercando un'altra cella libera.

→ La funzione hash deve essere anche funzione delle posizioni esaminate $h(k, i)$, prende una chiave e un indice e genera una nuova posizione.

$$h: U \times \{0, 1, \dots, m-1\} \rightarrow \{0, 1, \dots, m-1\}$$

data una chiave k : la prima volta $h(k, 0)$, poi $h(k, 1)$, la generica $h(k, i)$.

Hash-Insert (T, k) {

```

i=0;
repeat j=h(k,i)
  if (T[j]=NULL){
    T[j]=k;
    return j;
  }
  else i=i+1;
until i=m;
error "overflow";}
    
```

Hash-Search (T, k) {

```

i=0;
repeat j=h(k,i)
  if (T[j]=k){
    return j;
    i=i+1;}
until i=m or T[j]=NULL;
return NULL;}
    
```

è hash primaria

1) SCANSIONE LINEARE

$$h(k, i) = (h'(k) + i) \% m$$

problema di concatenazione primaria: le collisioni vengono distribuite ma si concentrano attorno a una posizione di hash.

Esempio.

0	22
1	88
2	
3	
4	4
5	15
6	17
7	59
8	
9	
10	31
	10

$m=11; \{22, 4, 31, 10, 15, 17, 88, 59\}$

$h(k, i) = (k \bmod 11 + i) \% 11$

$22 \bmod 11 = 0 \quad i=0$ libera ✓

$4 \bmod 11 = 4 \quad i=4$ libera ✓

$31 \bmod 11 = 9 \quad i=9$ libera ✓

$10 \bmod 11 = 10 \quad i=10$ libera ✓

$15 \bmod 11 = 4 \quad i=4$ occupata X $i=5$ ✓

$17 \bmod 11 = 6 \quad i=6$ libera ✓

$88 \bmod 11 = 0 \quad i=0$ occupata X $i=0+1=1$ ✓

$59 \bmod 11 = 4 \quad i=4$ occupata X $i=4+1=5$ X

$i=4+2=6$ X $i=4+3=7$ ✓

PROGRAMMAZIONE DINAMICA

- Divide in sottoproblemi (non necessariamente indipendenti)
- Li risolve una sola volta e memorizza la soluzione in una tabella (così si evita di risolvere più volte lo stesso problema, da esponenziale a polinomiale)
- Soluzione TIPICA: esplorazione delle soluzioni in modo bottom-up
- Applicazione ai casi: se c'è più di una soluzione, a cui è associabile una preferenza e si vuole determinare quella con indice ottimo.

1. Caratterizzazione della struttura di una soluzione ottima, in termini di sottoproblemi: SOTTOSTRUTTURA OTTIMA, costituita da sottosoluzioni ottime.

(Non tutti i problemi sono caratterizzabili in questo modo!)

2. Definizione ricorsiva del valore di una soluzione ottima (ricorrenza; la soluzione ottima contiene le sottosoluzioni ottime ai sottoproblemi).

3. Calcolo del valore di una soluzione ottima "bottom-up" (prima i + semplici) si usa una tabella per memorizzare le soluzioni ai sottoproblemi, evitando di ripetere quelli già fatti.

4. Costruzione della soluzione ottima (unica).

esempio

Parentesizzazione ottima nella moltiplicazione tra matrici (vogliamo minimizzare il numero di moltiplicazioni) sfruttando la proprietà associativa.

C_0 righe A_1

C_1 colonne A_1 $P(m) = m!$ modi di calcolare il prodotto di m matrici $A_1 \times A_2 \times \dots \times A_m$.

C_{i-1} righe A_i • Combinando la parentesizzazione delle prime k matrici e poi

C_i colonne A_i delle rimanenti $m-k$ $P(m) = \sum_{k=1}^{m-1} P(k) P(m-k) = \Omega\left(\frac{4^m}{m^{3/2}}\right)$

= righe A_{i+1} (esponenziale!)

① Caratterizzazione soluzione ottima: divido il problema in due sottoproblemi

→ prodotto prime k matrici $A_1 \dots A_k$

→ prodotto delle altre $m-k$ matrici $A_{k+1} \dots A_m$ } devono godere anche di parentesizz. ottimali

la soluzione $A_{1\dots m}$ è il prodotto di $A_{1\dots k}$ e $A_{k+1\dots m}$

il costo del prodotto $A_{1\dots m}$ è la somma del costo dei prodotti

$A_{1\dots k}$ e $A_{k+1\dots m}$ più il costo del prodotto finale $= C_0 E_k C_m$.

PROP. Ogni soluzione ottima al problema contiene le soluzioni ottime dei due sottoproblemi (dim.)

REQUISITO della SOTTOSTRUTTURA OTTIMA

e

della SOVRAPPOSIZIONE di SOTTOPROBLEMI

alcuni sottoproblemi si devono ripetere

④ $S[i, j]$ contiene il valore di k su cui spezzare il prodotto $A_{i \dots j}$

$\rightarrow A_{i \dots j} = A_{i \dots k} \times A_{k+1 \dots j}$

ALGORITMO RICORSIVO

MCM(A, S[], i, j)

if (j > i)

then $k = S[i, j]$

$X = \text{MCM}(A, S, i, k)$

$Y = \text{MCM}(A, S, k+1, j)$

return Prod. Matrici (X, Y)

else return $A[i, j]$

$A[]$ vettore di lunghezza n
contiene le matrici

$[A_1 \dots A_n]$

$S[]$ matrice $n \times n$
contiene k ottimo per
ogni coppia di indici

VARIANTE: RICORSIONE CON MEMORIZZAZIONE

↑ tipico del divide & conquer ↑ tipico della programmazione dinamica

Quando risolvo un sottoproblema per la prima volta lo memorizzo in tabella
poi per ogni sottoproblema controllo se è già stato risolto

SI uso il risultato
NO calcolo e memorizzo

la complessità non può superare il numero di casi distinti.

Mem MCO (c[])

$m = \text{lung}(c) - 1$

for $i = 1$ to m

do for $j = i$ to m

do $m[i, j] = \infty$

return cercaMC(c, 1, m)

CERCA MC (c[], i, j)

if $m[i, j] < \infty$

return $m[i, j]$

if (i = j)

$m[i, j] = 0$

else for $k = i$ to $j - 1$

do costo = $\text{cercaMC}(c, i, k) + \text{cercaMC}(c, k+1, j) + c[i-1]c[k]c[j]$

if costo < $m[i, j]$

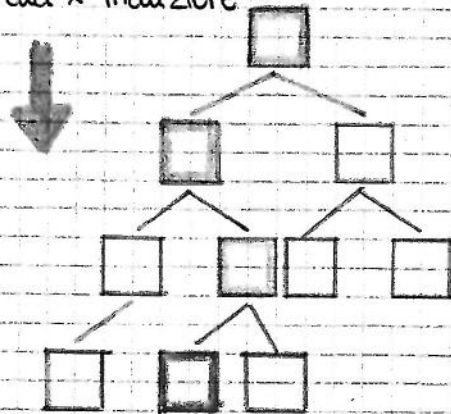
$m[i, j] = \text{costo}$

return $m[i, j]$;

La variante presenta dei vantaggi quando non debbano esser risolti tutti i sottoproblemi (tratta solo quelli necessari).

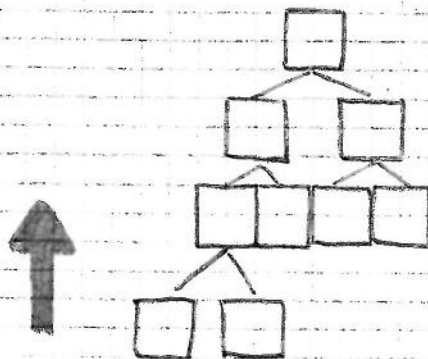
G
R
E
E
D
Y

- la scelta fatta da un algoritmo greedy può dipendere dalle scelte fatte fino a quel momento;
- non dipende dalle scelte future ma tantomeno di tutte le soluzioni del sottoproblema.
- prende la decisione e cambia il percorso verso la soluzione dopo la decisione **NON RICONSIDERA MAI LE DECISIONI GIÀ PRESE**
- procede in modo iterativo top-down, viene generata una e una sola sequenza di soluzioni.
- difficoltà: dimostrare che greedy permette di trovare un ottimo globale va dimostrata x induzione



P
R
O
D
I
N
A
M
I
C
A

- ha in comune la sottostruttura ottima
- procede in bottom-up
- esaustiva e garantisce di trovare una soluzione
- prende le decisioni basandosi su tutte le decisioni prese allo stadio precedente.
- le decisioni dipendono dalla soluzione ai sottoproblemi (tabella)



tutti i modi hanno un valore.

else if ($w = p_i$)

$$N(i, w) = \max \{ N(i-1, w), c_i \}$$

$$\text{else } N(i, w) = N(i-1, w) \}$$

esempio $N(i, w) = \max \{ N(i-1, w), N(i-1, w - p_i) + c_i \}$

1 c_i 24 p_i 6

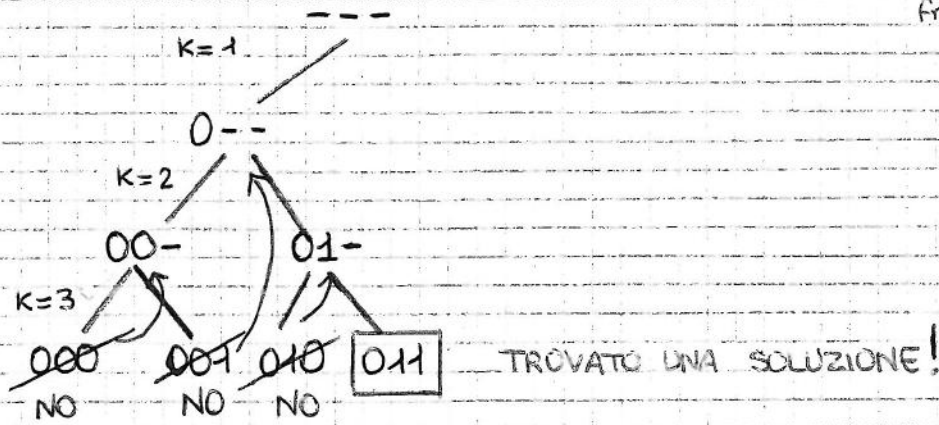
2 12 2

3 25 5

11

$i \backslash w$	0	1	2	3	4	5	6	7	8	9	10	11
0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	24	24	24	24	24	24
2	0	0	12	12	12	12	24	24	36	36	36	36
3	0	0	12	12	12	25	25	37	37	37	37	49

esempio. $m=3$



TUTTE LE SOLUZIONI

Backtrack (K)

for ($i = 1 \dots m$)

choose K-th candidate

if (acceptable)

store it

if ($K < m$)

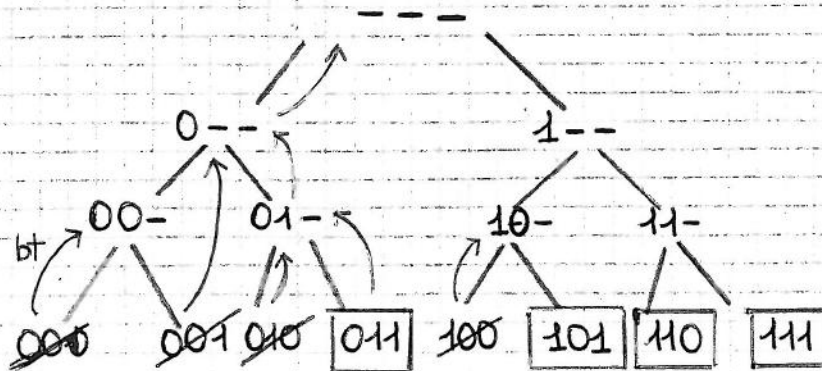
backtrack ($K+1$)

else

print solution

Remove from solution

esempio



GRAFICI

Sono rilevanti l'insieme di oggetti e le connessioni.

$$G = (V, E)$$

V = vertici, E = archi (definito come la coppia di vertici che unisce)

un grafo è orientato se E è una relazione binaria tra i vertici (archi orientati)

un grafo è non orientato se gli archi non sono orientati

arco $e = (w, v) \in E$

INCIDENZA

e incidente su v

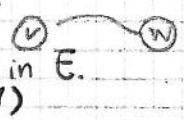
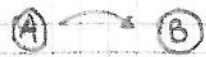
e incidente da w

v è incidente su e

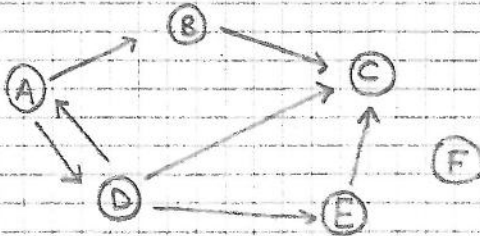
(A, B) è incidente da A su B

(grafi non orientati) e incide su v e w

a un altro v se l'arco che li collega è in E .



Un vertice w si dice adiacente



- B adiacente ad A
- C adiacente a B e D
- A adiacente a D e viceversa
- B non è adiacente a D né a C
- F non è adiacente ad alcun vertice.

PROP. In un grafo non orientato l'adiacenza è simmetrica.

Grado di un vertice = n. di archi che da esso dipartono;

in un grafo orientato si distinguono il grado entrante e il grado uscente

Grafi pesati: ogni arco ha un peso associato definito da una funzione di costo (quando tra due vertici non esiste arco \rightarrow costo infinito).

Sottografo di G $H = (V^*, E^*) / V^* \subseteq V, E^* \subseteq E$ e $E^* \subseteq V^* \times V^*$

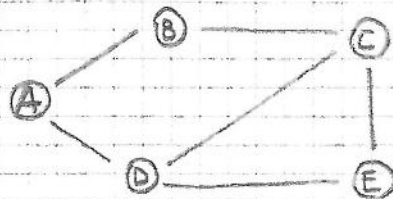
sottografo indotto da V^* è $H = (V^*, E^*)$ tale che $E^* = \{(v, w) \in E / v, w \in V^*\}$.

Percorso sequenza di vertici $\langle w_1, w_2, \dots, w_m \rangle$ tale che $(w_i, w_{i+1}) \in E \quad 1 \leq i \leq m-1$
la lunghezza sarà $m-1$.

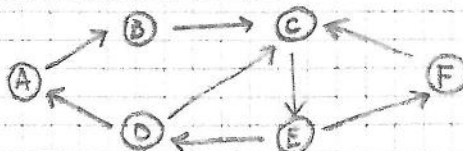
Percorso semplice se tutti i suoi vertici sono distinti

Nodo w raggiungibile da v tramite p se esiste un percorso tra v e w $V \xrightarrow{p} W$

Un grafo non orientato è connesso se esiste un percorso da ogni vertice a ogni altro vertice



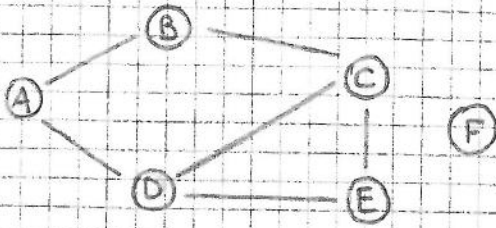
Un grafo orientato è fortemente connesso se esiste un percorso da ogni vertice a ogni altro vertice



RAPPRESENTAZIONE di GRAFI

CON MATRICE di ADIACENZA $|V|^2$

$$M(v,w) = \begin{cases} 1 & \text{se } (v,w) \in E \\ 0 & \text{altrimenti} \end{cases}$$

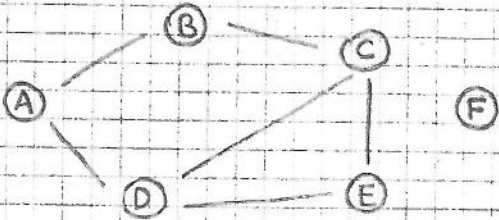


	A	B	C	D	E	F
A	0	1	0	1	0	0
B	1	0	1	0	0	0
C	0	1	0	1	1	0
D	1	0	1	0	1	0
E	0	0	1	1	0	0
F	0	0	0	0	0	0

matrice simmetrica se grafo non orient

CON LISTE di ADIACENZA $a|V| + 2b|E|$

$L(v) = \text{lista di } w \text{ tale che } (v,w) \in E \text{ per } v \in V$



A	→ B	→ D			
B	→ A	→ C			
C	→ B	→ D	→ E		
D	→ A	→ C	→ E		
E	→ C	→ D			
F					

se grafo orientato spazio = $a|V| + b|E|$

CON MATRICE di INCIDENZA $\Theta(|V| \times |E|)$

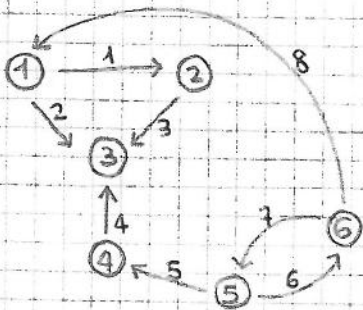
avendo numerato i vertici $1, 2, \dots, |V|$

$$D = (d_{ij}) \begin{cases} +1 & \text{se } e_j \text{ entra in } v_i \\ -1 & \text{se } e_j \text{ esce da } v_i \\ 0 & \text{altrimenti} \end{cases}$$

dim = $|V| \times |E|$



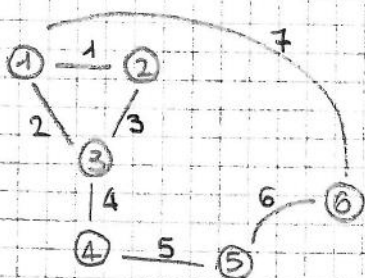
se il grafo non è orientato non si distingue il verso → matrice simmetrica



vertici

	1	2	3	4	5	6	7	8
1		-1	-1	0	0	0	0	+1
2	+1		0	-1	0	0	0	0
3	0	+1		+1	0	0	0	0
4	0	0	0		-1	+1	0	0
5	0	0	0	0		-1	-1	+1
6	0	0	0	0	0		+1	-1

← archi



	1	2	3	4	5	6	7
1		1	1	0	0	0	1
2	1		0	1	0	0	0
3	0	1		1	1	0	0
4	0	0	0		1	1	0
5	0	0	0	0		1	1
6	0	0	0	0	0		1

VISITE di GRAFI = attraversamento di tutti i vertici a partire dalla sorgente

BIANCO = vertice inesplorato;

GRIGIO = vertice esplorato ma ancora utile per visitare le adiacenze;

NERO = vertice esplorato con tutte le sue adiacenze → non è più utile.

$O(N)$ Initialize (G)
 for ogni $u \in V$ do
 color[u] = white

 $O(N+E)$ Visit (G, S) S = sorgente
 color[S] = gray
 while ci sono vertici grigi do
 u = vertice grigio
 if $\exists v$ bianco adiacente a u
 then color[v] = gray ($P[v] = u$)
 else color[u] = black

OPZIONALE:
 permette di ricordare l'arco percorso per visitare v, ossia il suo predecessore
 al termine tutti i vertici neri hanno un predecessore

sottografo dei predecessori $G_P = (V_P, E_P)$ è un albero o una foresta.

$$V_P = \{ n \in V : P[n] \neq \text{NULL} \} \cup \{s\}$$

$$E_P = \{ (P[n], n) \in E : n \in V_P - \{s\} \}$$

al termine di visit (G, S) V_P è l'insieme di tutti i vertici neri (quelli raggiungibili da S)

Per gestire l'insieme di vertici grigi si può usare una struttura dati D con le sue primitive.

Visit (G, S)

D = create_empty

color[S] = gray

add(D, S)

while not-empty(D) do

 u = first(D)

 if $\exists v$ bianco adiacente a u

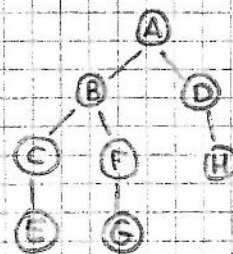
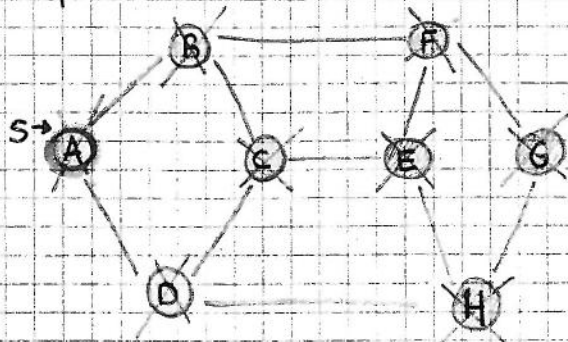
 then color[v] = gray

 add(D, v)

 else color[u] = black

 remove_first(D)

esempio.



Q

A	B	D	C	F	H	E	G
0	1	1	2	2	2	3	3

DFS - DEPTH-FIRST-SEARCH

VISITA IN PROFONDITA'

si scoprono prima i vertici adiacenti all'ultimo adiacente scoperto; generalizzazione della visita in preordine per alberi binari.



- Viene processato il vertice s ;
- Ricorsivamente s : processiamo tutti i vertici adiacenti a s ; necessar evitare di riprocessare vertici già visitati (attenzione ai cicli)
- il sottografo dei predecessori può formare una foresta di alberi perché la visita può esser ripetuta da più sorgenti.

$$G_p = (V, E_p)$$

$$E_p = \{ (p[v], v) \in E : v \in V \text{ e } p[v] \neq \text{NULL} \}$$

- Usa uno stack per memorizzare i nodi grigi.

DFS (G, S)

S = make-empty-stack

color[s] = gray

push(S, s)

while not-empty(s) do

 while c'è un v adiacente a top(s) non visto do

 if color[v] = white

 color[v] = gray

 p[v] = top(s)

 push(S, v)

 color[top(s)] = black

 pop(s)

DFS con GRAFI ORIENTATI

CLASSIFICAZIONE di ARCHI



dell'albero (tree, T)

all'indietro (backward, B)

collega a un antenato

in avanti (forward, F)

collega a un discendente

di attraversamento (cross, C)

se collega vertici che non sono antenato/discendente

in base al colore del vertice raggiunto

1. BIANCO $\rightarrow (u, v)$ è T dell'albero

2. GRIGIO $\rightarrow (u, v)$ è B all'indietro

3. NERO \rightarrow (visita di v è terminata) $\rightarrow (u, v)$ è F se $d[u] < d[v]$
 $\rightarrow (u, v)$ è C se $d[v] < d[u]$

Applicazioni:

- ordinamento topologico di un DAG
- calcolo delle componenti fortemente connesse.

ORDINAMENTO TOPOLOGICO di un DAG

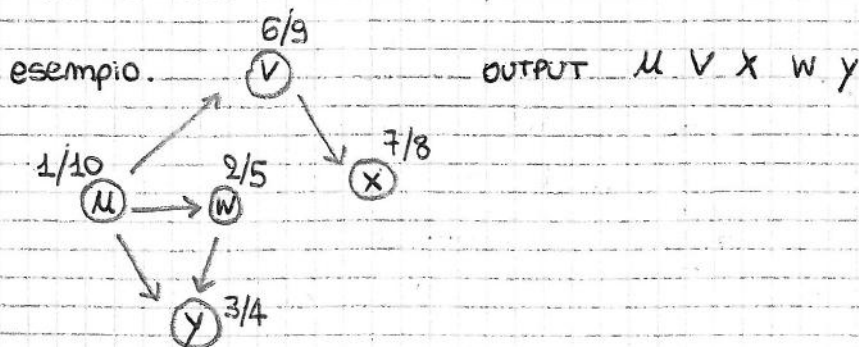
ordinamento lineare di tutti i vertici tale che se esiste l'arco (u, v) in E , allora u precede v . (Non è unico!)

TOPSORT (G)

Chiama DFS(G) per calcolare i tempi $f[v]$

Ogni volta che un vertice è "terminato" aggiungilo in testa a una lista

Ritorna la lista di vertici



CALCOLO delle COMPONENTI CONNESSE

componente connessa = sottografo massimale connesso di un grafo non orientato (massimale = non è possibile aggiungere vertici o archi lasciandolo connesso)

CC COMPONENTI CONNESSE (di grafi non orientati) = classi di equivalenza dei vertici rispetto alla relazione di connettività. ("raggiungibile da")

SCC COMPONENTI FORTEMENTE CONNESSE (di grafi orientati) = classi di equivalenza rispetto alla relazione di connettività forte ("esiste un cammino tra")

ALBERI di COPERTURA MINIMI MST

Dato un grafo non orientato, connesso, con archi pesati

ALBERO di COPERTURA MINIMO = sottografo aciclico $T \subseteq E$ che connette tutti i vertici minimizzando il peso totale. $W(T) = \sum W(u,v)$

T è un albero perché è aciclico e connesso (tocca tutti i vertici)

Non è unico!

$A \subseteq T$

Idea Accrescere un sottoinsieme A di archi di un albero di copertura aggiungendo ogni volta un ARCO SICURO (mantiene la proprietà per A di essere un sottoinsieme di archi di un albero di copertura)

A inizialmente è vuoto

Generic MST (G, w)

$A = \emptyset$

while A non forma un albero di copertura

do trova un arco sicuro (u, v)

$A = A \cup \{(u, v)\}$

approccio greedy

return A

Taglio = partizione dei vertici V

Arco attraversa il taglio = gli estremi dell'arco sono in partizioni diverse

Taglio rispetta un insieme di archi = nessun arco attraversa il taglio

Arco leggero = arco con peso minimo tra quelli che attraversano un dato taglio

Teorema

$G(V, E)$ grafo non orientato, connesso, con una fz di peso w

$A \subseteq E$ contenuto in qualche albero di copertura minimo per G .

$(S, V-S)$ è un taglio che rispetta A

(u, v) arco leggero che attraversa il taglio

⇒ (u, v) è sicuro per A .

Oss. A è sempre aciclico altrimenti un MST T conterrebbe un ciclo

$G_A = (V, A)$ è una foresta, e ogni sua cc è un albero. gli archi sicuri connettono cc

Corollario

$G(V, E)$ grafo non orientato, connesso, con una fz di peso w

$A \subseteq E$ contenuto in un albero di copertura minimo per G

C componente connessa (albero) nella foresta $G_A = (V, A)$

(u, v) arco leggero che connette C a qualche componente in G_A

⇒ (u, v) è sicuro per A .

Algoritmo di Prim

- mantiene A come singolo albero (una singola componente connessa)
- albero parte da un vertice arbitrario (radice) e cresce fino a coprire tutto.
- ogni volta si aggiunge un arco leggero che collega un vertice in A con uno in $V-A$ (sicuro per il corollario, dato che A è una CC)
- greedy perché aggiunge sempre l'arco di peso minimo.

MST-PRIM (G, w)

$A = \{r\}$ albero formato da un solo nodo

ordina gli archi di E per peso w non decrescente

while (A ha meno di $|V|$ nodi) do

 trova l'arco (u, v) di costo minimo incidente su A

$A = A \cup \{(u, v)\}$

return A .

EFFICIENZA dipende da scelta del nuovo arco. Soluzione:

1. $Q = V - A$ (coda con priorità con tutti i vertici che non sono nell'albero in costruzione)
2. la priorità è basata su $key[n^*] = \text{minimo tra i pesi degli archi che collegano } n^* \text{ a un qualunque vertice dell'albero in costruzione } A$.
3. per ogni nodo si introduce parent $p[x]$ che serve per ricostruire l'albero. [visita con l'aggiunta dell'ordine di priorità!].

$A = \{(n^*, p[n^*]) : n^* \in V - \{r\} - Q\}$

alle fine $Q = \emptyset$, $A = \{(n^*, p[n^*]) : n^* \in V - \{r\}\}$.

MST-PRIM (G, w, r) con coda di priorità

$Q = V[G]$ tutti i vertici

for $\forall u$ in Q

 do $key[u] = \infty$

$key[r] = 0$

$p[r] = \text{NULL}$

while $Q \neq \emptyset$

 do $u = \text{EXTRACT-MIN}(Q)$

 for ogni n^* in $\text{Adj}[u]$

 if $n^* \in Q$ e $w(u, n^*) < key[n^*]$

 then $p[n^*] = u$

$key[n^*] = w(u, n^*)$.

CAMMINI MINIMI

costo del cammino $w(p) = \sum_{i=1}^k w(v_{i-1}, v_i)$ $p = (v_0, \dots, v_k)$

peso di un cammino minimo $\delta(u, v) = \begin{cases} \min (w(p) \cdot u \rightarrow v) & \text{se } \exists \text{ cammino} \\ \infty & \text{altrimenti} \end{cases}$

cammino p da u a v minimo se $w(p) = \delta(u, v)$
 può non essere definito in presenza di archi o cicli negativi.

Peso unitario \rightarrow BFS

ALBERO dei CAMMINI MINIMI: sottografo orientato $G'(V', E')$ di G

- G' albero con radice in S
- V' insieme vertici raggiungibili da S vertici distinti
- $\forall v \in V'$ l'unico percorso semplice da S a v è un percorso minimo.

Proprietà algoritmi \rightarrow sottostruttura ottima cammini minimi
 tecnica rilassamento

Lemma 1 - SOTTOSTRUTTURA OTTIMA dei CAMMINI MINIMI

$G = (V, E)$ grafo ~~non~~ orientato con funzione di peso w

$p = \langle v_1, \dots, v_k \rangle$ percorso minimo tra v_1 e v_k

$p_{ij} = \langle v_i, \dots, v_j \rangle$ sottopercorso di p tra v_i e v_j $\forall i, j$

$\Rightarrow p_{ij}$ è un percorso minimo tra v_i e v_j

Lemma 1

$G = (V, E)$ grafo pesato orientato con funzione di peso w.

percorso minimo p decomponibile in $S \xrightarrow{p'} u \rightarrow v$ per qualche u

\Rightarrow peso del percorso minimo tra S e v è $\delta(S, v) = \delta(S, u) + w(u, v)$

Lemma 2

$G = (V, E)$ grafo orientato con funzione di peso w; s è sorgente

$\Rightarrow \delta(s, v) \leq \delta(s, u) + w(u, v)$

Rilassamento verifica se è possibile ottenere un percorso migliore tra s e v passando per u.

Relax (u, v, w)

if $d[v] > d[u] + w(u, v)$

then $d[v] = d[u] + w(u, v)$

$p[v] = u$

\nearrow dist. da sorgente
 $d[v]$ estremo superiore lunghezza del percorso minimo tra s e v
 inizializzato a ∞

$p[v]$ predecessore di v

$w(u, v)$ peso arco (u, v)

cerca di dim. costo cammino minimo - se riesce \rightarrow NUOVO PREDECESSORE

Algoritmo Dijkstra

PESI NON-NEGATIVI

- Mantiene un insieme S di vertici i cui pesi del percorso minimo sono già stati determinati $d[u] = \delta(s, u)$
- seleziona a turno il vertice u in $V-S$ col minimo valore $d[u]$
- inserisce u in S
- rilassa tutti gli archi uscenti da u .
- usa una coda a priorità Q che contiene i vertici usando i valori $d[]$ come chiave

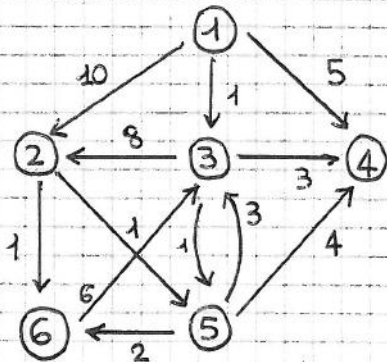
Dijkstra (G, s)

- $O(V)$ Inizializza (G, s) ($d = \infty$ e $p = \text{NULL}$, $d[s] = 0$)
- $O(V)$ $S = \emptyset$
- $Q = V(G)$
- $O(V)$ while ($Q \neq \emptyset$)
- $O(\log V)$ $u = \text{Delete_min}(Q)$
- $S = S \cup \{u\}$
- $O(E)$ for each vertex v adiacente a u
- $O(\log V)$ Relax (u, v, w) cambia valore di $d[] \rightarrow$ corrisponde a Decrease-key ($d[v], d[u] + w(u, v)$) sulla coda a priorità

TOT = $O((E+V) \log V)$

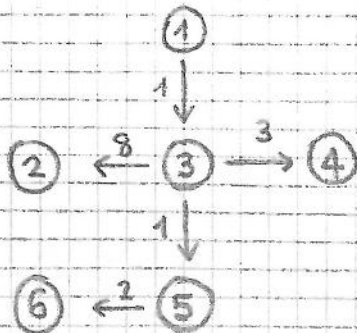
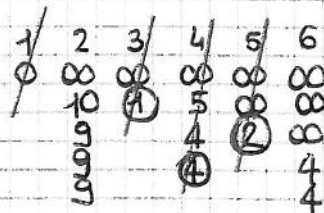
COMPLESSITA' $O(E \log V)$ la struttura dati è efficiente x mantenere priorità

esempio



S =

1	3	5	4	6	2
0	1	2	4	4	9



Algoritmo di Floyd-Warshall

ALGO di PR. DINAMICA
CAMMINI MINIMI TRA TUTE LE COPPIE

Archi anche negativi (ma cicli no)

$G = (V, E, w)$ grafo orientato pesato con $|V|=n$ rappre. con matrice di adiac.

$$W[i, j] = \begin{cases} 0 & \text{se } i=j \\ w(i, j) & \text{se } i \neq j \text{ e } (i, j) \in E \\ \infty & \text{se } (i, j) \notin E \end{cases} \quad \text{MATRICE di ADIACENZA}$$

$$P_{ij} = \begin{cases} \text{NULL} & \text{se } i=j \text{ o se } i \neq j \text{ e } j \text{ non è raggiungibile da } i \\ k & \text{se } i \neq j \text{ e } j \text{ raggiungibile da } i \text{ e } k \text{ predecess. di } j \text{ lungo il percorso minimo.} \end{cases}$$

↑ ogni riga P_i induce un sottografo dei predecessori

1. PROP. SOTTOSTRUTTURA OTTIMA dato un cammino ottimo p ogni suo sottocamm. è ottimo.

Data $W = w_{ij}$ considero il percorso minimo p tra i e j (supponiamo $p = m$ archi)

se $i \neq j$ $p = i \xrightarrow{P'} k \rightarrow j$ dove P' ha al più $m-1$ archi ed è il percorso minimo tra i e k $\delta(i, j) = \delta(i, k) + w_{kj}$

2. DEFINIZIONE RICORSIVA → devo trovare $d_{ij}^{(m-1)} \forall i, j$.

$d_{ij}^{(m)}$ = peso percorso min tra V_i e V_j usando al più m archi.

caso base: $m=0$ $d_{ij}^{(0)} = \begin{cases} 0 & i=j \\ \infty & i \neq j \end{cases}$

caso generico: $m \neq 0$ $d_{ij}^{(m-1)}$ matrice dei percorsi minimi con al più $m-1$ archi
→ percorso minimo tra V_i e V_j :

- contiene $m-1$ archi → costo $d_{ij}^{(m-1)}$
- contiene m archi → costo $d_{ij}^{(m)} = d_{ik} + w_{kj}$

ossia

$$d_{ij}^{(m)} = \min \left(d_{ij}^{(m-1)}, \min_{1 \leq k \leq m} \{ d_{ik}^{(m-1)} + w_{kj} \} \right) = \min_{1 \leq k \leq m} \{ d_{ik}^{(m-1)} + w_{kj} \}$$

Se il grafo non contiene cicli di peso negativo, ogni percorso minimo è un percorso semplice con al più $|V|-1$ archi

3. CALCOLO del VALORE della SOLUZIONE OTTIMA

algoritmo riceve $W = w_{ij}$ matrice di adiacenza del grafo pesato.

Vengono calcolate $D^{(0)} = D^{(1)} = D^{(2)} = \dots = D^{(|V|-1)}$

$D^{(|V|-1)}$ contiene i pesi dei percorsi minimi

$D^{(1)}$ coincide con W

$$D^{(k)} = (d_{ij}^{(k)})$$

INTRACTABLE PROBLEMS & NP COMPLETENESS

PROBLEMI TRATTABILI = problemi risolvibili in tempo polinomiale da algoritmi
(esponenti piccoli di solito)

PROBLEMI INTRATABILI = risolvibili solo con algoritmi con tempo di esecuzione $> m^k$

non tutti i problemi sono risolvibili con una soluzione polinomiale, la conoscenza della NP COMPLETEZZA è importante per definire un algoritmo, di solito si ricorre ad algoritmi approssimati (euristici): lavorano bene nella maggior parte dei casi. La teoria della NP-completezza è limitata ad un tipo di problema:

PROBLEMI DI DECISIONE: problemi con risposta binaria $\{0,1\}$.

I problemi di ottimizzazione possono essere facilmente ricondotti a problemi di decisione (con artifici atti a restringere il campo delle possibilità).

Finora abbiamo considerato solamente algoritmi deterministici, per i quali cioè il risultato di ciascuna operazione è unicamente definita (l'hardware è determ).

Per la teoria della NP COMPLETEZZA rimuoviamo questa restrizione ed assumiamo che il risultato delle operazioni non sia univocamente determinato.

→ ALGORITMI NON DETERMINISTICI

non esistono macchine non deterministiche **OP** ma danno un aiuto intuitivo

per capire quali problemi ^(non) siamo \varnothing risolvibili con algoritmi determu.

- choice** (x_1, \dots, x_n) ritorna un elemento arbitrario (non random)
sceglie il valore giusto
- failure** () segnala una terminazione senza successo
- success** () segnala una terminazione con successo

Nota bene che non ritornano un valore, failure e success.

ND search (A, n, k) {

$j = \text{choice}(1, m)$

if ($A[j] == k$)

print j;
success;

else
print '0';
failure; }

cerca k nell'array A

cerca j t.c. $A[j] = k$ oppure $j = 0$
 $k \notin A$

sostituisce il ciclo for deterministico.

$O(1)$

(corrisp deterministico $O(m)$)

Concettualmente la choice corrisponde a fare tante copie dell'algoritmo quante sono le possibilità di scelta; la prima copia che raggiunge il valore atteso termina anche tutte le altre (spiegazione deterministica non funziona così!).

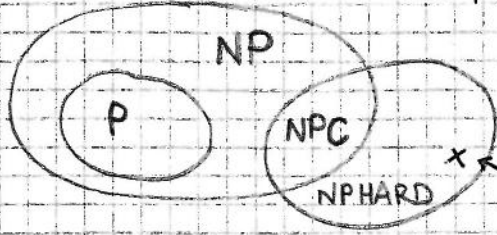
NP-HARD per ogni altro problema π in NP, $\pi \leq \pi$ (almeno difficile quanto un NP)

NP-COMPLETE { NP (i più difficili problemi in NP)
 per ogni altro problema A' in NP $A' \leq A$

devo verificare che è NP e che NPC altro più di a questo problema.

$NP \cap P = \emptyset$

$NP-C = NP \cap NP-HARD$



non verificabile in tempo polinomiale (problemi che non sono di decisione).

SAT BOOLEAN SATISFIABILITY $O(m)$ (D: complex esponenziale)

- problema NPC (il primo)
- tutti i problemi NP \leq SAT.
- trovare i valori (vero/falso) per a_i così che soddisfi l'equazione booleana. (and, or, not)

```

ND sat (E, m) {
  for (i=1 to m) do
     $x_i = \text{choice}(\text{true}, \text{false})$ 
  if  $\{E(x_1, \dots, x_n)\}$  is true
    success
  else failure
}
    
```

CDP CLIQUE DECISION PROBLEM

- Dato un grafo non orientato, dire se contiene sottografi connessi di dim = k.
- problema NPC (è NP: può essere verificato in tempo polinomiale).

3CNF SAT

- 3-SAT = { F / F è conjunctive normal form : ogni clausola ha 3 lettere e F è soddisfabile }
- 3-SAT \leq CDP
 trasformiamo le clause 3CNF in un grafo; ogni k-clique corrisponde ad un'assegnazione accettabile.

HAMILTONIAN CYCLE

percorre n vertici di G visitando tutti i vertici una volta e tornando a quello di partenza.

SAT \leq DIRECTED HAMILTONIAN CYCLE

TSP TRAVELLING SALESPERSON PROBLEM

Deve visitare n vertici visitando ciascuno una volta sola e ritornando quello di partenza, ce ne è uno lungo al massimo d?

