

# Indice

<b>Tecnica di programmazione dinamica . . . . .</b>	<b>3</b>
Longest Common Subsequence . . . . .	3
Weighted Interval Scheduling . . . . .	5
Hateville . . . . .	8
Longest Increasing Subsequence . . . . .	10
Longest Increasing Common Subsequence . . . . .	12
Sottovettore di peso massimo . . . . .	14
Knapsack 0/1 . . . . .	16
Chiusura palindroma . . . . .	19
Distanza di modifica . . . . .	20
Problema Interleaving . . . . .	23
Cammino su un grafo . . . . .	25
 <b>Tecnica greedy . . . . .</b>	 <b>31</b>
Introduzione alla tecnica greedy . . . . .	31
Interval scheduling . . . . .	31
Zaino frazionario . . . . .	32
Matroidi e sistemi di indipendenza . . . . .	34
Algoritmo Greedy Standard . . . . .	36
Strutture dati per insiemi disgiunti . . . . .	38
Minimum Spanning Tree . . . . .	39
Cammini minimi da sorgente unica . . . . .	44
 <b>Visite di grafi . . . . .</b>	 <b>47</b>
Visita in ampiezza . . . . .	47
Visita in profondità . . . . .	48
Ordinamento topologico . . . . .	50



# Capitolo 1

## Tecnica di programmazione dinamica

### 1.1 Longest Common Subsequence

#### 1.1.1 Definizione del problema

Data una sequenza  $S$ , si dice **sottosequenza** di  $S$  un qualsiasi sottoinsieme degli elementi di  $S$ , disposti secondo l'ordine in cui appaiono. Più formalmente, una sottosequenza di lunghezza  $n$  di una sequenza  $S$  è definita da un insieme di indici  $i_1 < i_2 < \dots < i_k$ , tale per cui  $k \leq n$ ; la sottosequenza che si ricava da questi indici è  $S[i_1] \mid S[i_2] \mid \dots \mid S[i_k]$ .

Date due sequenze  $X$  e  $Y$ , una sottosequenza si dice comune ad entrambe le sequenze se appare sia in  $X$  che in  $Y$ . Il problema della sottosequenza comune più lunga consiste nel trovare, in due stringhe  $X$  ed  $Y$ , qual'è (se esiste) la sottosequenza comune ad entrambe che sia di lunghezza maggiore, abbreviata con *LCS* (*longest common subsequence*).

#### 1.1.2 Programmazione dinamica: sottostruttura ottima

**Proprietà della sottostruttura ottima per il problema Longest Common Subsequence.** Siano  $X = \langle x_1, x_2, \dots, x_m \rangle$  e  $Y = \langle y_1, y_2, \dots, y_n \rangle$  due sequenze. Sia poi  $Z = \langle z_1, z_2, \dots, z_k \rangle$  una LCS di  $X$  e di  $Y$ . Possono presentarsi tre situazioni:

1. Se  $x_m = y_n$ , allora  $z_k = x_m = y_n$  e  $Z_{k-1}$  è una LCS di  $X_{m-1}$  e  $Y_{n-1}$ ;
2. Se  $x_m \neq y_n$ , allora  $z_k \neq x_m$  implica che  $Z$  è una LCS di  $X_{m-1}$  e  $Y$ ;
3. Se  $x_m \neq y_n$ , allora  $z_k \neq y_n$  implica che  $Z$  è una LCS di  $X$  e  $Y_{n-1}$ ;

**Dimostrazione.** La dimostrazione procede caso per caso:

1. Si assuma per assurdo che  $z_k \neq x_m$ . Si accodi allora  $x_m = y_n$  a  $Z$ , ottenendo la stringa  $Z' = \langle z_1, z_2, \dots, z_k, x_m \rangle$ . Tale stringa ha lunghezza  $k + 1$  e, essendo  $x_m = y_n$ , è una LCS per  $X$  e  $Y$ . Ma allora  $Z$  non può essere una LCS per  $X$  e  $Y$ , perché  $Z$  ha lunghezza  $k$  mentre  $Z'$  ha lunghezza  $k + 1$ . Questo è in contraddizione con l'ipotesi di partenza secondo cui  $Z$  è effettivamente una LCS, pertanto deve aversi  $x_m = y_n = z_k$ .  
Dato che per ipotesi  $x_m = y_n$ , il prefisso  $Z_{k-1}$  è una sottosequenza comune di  $X_{m-1}$  e  $Y_{n-1}$  avente lunghezza  $k-1$ . Si vuole dimostrare che  $Z_{k-1}$  è una LCS di  $X_{m-1}$  e  $Y_{n-1}$ . Si supponga per assurdo che questo non sia vero, e che esista quindi una sequenza  $W$  che è LCS di  $X_{m-1}$  e  $Y_{n-1}$  avente lunghezza maggiore di  $k-1$ . Allora, accodando  $x_m = y_n$  a  $W$  si ottiene la sequenza  $W'$ , che è sottosequenza comune di  $X$  e di  $Y$  avente lunghezza maggiore di  $k$ . Questo è però in contraddizione con il fatto che  $Z$  sia una LCS per  $X$  e  $Y$ , dato che questa è lunga  $k$ . Occorre allora assumere che  $Z_{k-1}$  è una LCS di  $X_{m-1}$  e  $Y_{n-1}$ .
2. Si supponga per assurdo che, se  $z_k \neq x_m$ , allora  $Z$  non sia una LCS di  $X_{m-1}$  e  $Y$ . Deve allora esistere una sequenza  $W$  di lunghezza maggiore di  $k$  che è LCS di  $X_{m-1}$  e  $Y$ . Ma allora  $W$  è anche LCS per  $X$  e  $Y$ , contraddicendo l'ipotesi che lo sia  $Z$ . Occorre allora assumere che  $Z$  sia una LCS di  $X_{m-1}$  e  $Y$ .
3. La dimostrazione è simmetrica a quella del punto precedente.

#### 1.1.3 Programmazione dinamica: equazione di ricorrenza

Date due sequenze  $X$  e  $Y$  rispettivamente di lunghezza  $m$  e  $n$ , siano  $X_i = X[1 : i]$  e  $Y_j = Y[1 : j]$  i prefissi di lunghezza  $i$  e  $j$  delle rispettive sequenze. La soluzione  $S_{i,j}$  per l' $i, j$ -esima istanza del problema corrisponde a trovare la soluzione ottimale per il problema LCS rispetto alle sottosequenze  $X_i$  e  $Y_j$ . La soluzione per le intere sequenze  $X$  e  $Y$  è la soluzione per l'istanza  $X_m, Y_n$ . La lunghezza della sottosequenza comune più lunga soluzione per la  $i, j$ -esima istanza del problema viene indicata con  $\text{Opt}(i, j)$ .

Il caso base è semplice da determinare; è evidente come due sequenze vuote non abbiano alcun elemento in comune. Inoltre, una qualsiasi sequenza non vuota non ha alcun carattere in comune con la sequenza vuota.

$$S_{i,j} = \langle \rangle \text{ se } i = 0 \vee j = 0$$

$$\text{Opt}(i, j) = 0 \text{ se } i = 0 \vee j = 0$$

Per quanto riguarda il passo ricorsivo, si assuma di avere a disposizione tutte le soluzioni nella forma  $S_{a,b}$ , con  $1 \leq a < i$  e  $1 \leq b < j$ , e di voler calcolare  $S_{i,j}$ . Possono verificarsi solamente due situazioni, mutualmente esclusive: l' $i$ -esimo carattere di  $X_i$  è uguale al  $j$ -esimo carattere di  $Y_j$  oppure l' $i$ -esimo carattere di  $X_i$  è diverso dal  $j$ -esimo carattere di  $Y_j$ .

Nel primo caso, per definizione, tale carattere deve essere parte della sottosequenza comune più lunga. Pertanto, in questo caso, la soluzione ottimale per la  $i, j$ -esima istanza si ottiene semplicemente accodando l'ultimo carattere alla soluzione ottimale dell'istanza precedente, ovvero  $S_{i-1,j-1}$ .

Nel secondo caso, la soluzione ottimale per la  $i, j$ -esima istanza deve necessariamente essere di lunghezza maggiore o uguale (ma non minore) alla lunghezza di tutte le soluzioni per istanze precedenti, e non può, per definizione, contenere l' $i$ -esimo elemento di  $X$  (o il  $j$ -esimo elemento di  $Y$ ). In particolare, le uniche soluzioni candidate ad essere le soluzioni per la  $i, j$ -esima istanza sono la  $i-1, j$ -esima e la  $i, j-1$ -esima, ovvero  $S_{i-1,j}$  e  $S_{i,j-1}$ , perché a loro volta comprendono tutte le istanze a loro precedenti. Dato che sono entrambe valide, il criterio di scelta consiste nello scegliere fra le due la soluzione con lunghezza maggiore.

$$S_{i,j} = \begin{cases} S_{i-1,j-1} \cup \{x_i\} & \text{se } X[i] = Y[j] \\ S_{i-1,j} & \text{se } X[i] \neq Y[j] \wedge \text{Opt}(i-1, j) > \text{Opt}(i, j-1) \\ S_{i,j-1} & \text{se } X[i] \neq Y[j] \wedge \text{Opt}(i, j-1) > \text{Opt}(i-1, j) \end{cases}$$

$$\text{Opt}(i, j) = \begin{cases} \text{Opt}(i-1, j-1) + 1 & \text{se } X[i] = Y[j] \\ \max\{\text{Opt}(i-1, j), \text{Opt}(i, j-1)\} & \text{se } X[i] \neq Y[j] \end{cases}$$

#### 1.1.4 Programmazione dinamica: implementazione bottom-up

L'algoritmo bottom-up viene costruito a partire dall'equazione di ricorrenza sfruttando una tabella  $c$ . In ciascuna cella  $c[i, j]$  viene riportato il valore della soluzione ottimale per la  $i, j$ -esima istanza del problema, che viene utilizzata per calcolare i valori ottimali per le istanze successive. L'algoritmo riceve in input le due sequenze  $X$  e  $Y$  e restituisce in output la tabella  $c$ .

Si noti come le celle che hanno 0 come uno dei due indici possano venire riempite immediatamente con 0, come da caso base dell'equazione di ricorrenza.

```

procedure LCS(X, Y)
  for i = 0 to |X| do
    c[i, 0] = 0
  for j = 0 to |Y| do
    c[0, j] = 0

  for i = 1 to |X| do
    for j = 1 to |Y| do
      if X[i] == Y[j] then
        c[i, j] = 1 + c[i - 1, j - 1]
      else
        c[i, j] = max(c[i, j - 1], c[i - 1, j])

  return c

```

È facile notare come il tempo di esecuzione dell'algoritmo sia  $O(nm)$ , dove  $m$  e  $n$  sono le lunghezze di rispettivamente la stringa  $X$  e la stringa  $Y$ . I primi due cicli eseguono una operazione immediata rispettivamente per  $m$  e per  $n$  volte, dopodiché si presenta un doppio ciclo innestato: il ciclo esterno esegue il ciclo interno  $m$  volte, mentre il ciclo interno esegue  $n$  volte un blocco di decisione il cui tempo di esecuzione può considerarsi immediato. Si ha allora che, asintoticamente,  $O(m) + O(n) + O(mn) = O(mn)$ .

	ε	S	A	T	U	R	D	A	Y
ε	0	0	0	0	0	0	0	0	0
S	0	1	1	1	1	1	1	1	1
U	0	1	1	1	2	2	2	2	2
N	0	1	1	1	2	2	2	2	2
D	0	1	1	1	2	2	3	3	3
A	0	1	2	2	2	2	3	4	4
Y	0	1	2	2	2	2	3	4	5

### 1.1.5 Programmazione dinamica: ricostruzione di una soluzione

Una volta calcolata la lunghezza della LCS, é possibile individuarne una ripercorrendo la tabella a ritroso. In particolare, questo viene fatto estendendo l'algoritmo per il calcolo della lunghezza introducendo una nuova tabella  $b$  di puntatori. In particolare, nella cella  $(i, j)$  della tabella  $b$  viene inserito:

- Un puntatore  $\nwarrow$  quando l' $i$ -esimo elemento della prima sequenza é uguale al  $j$ -esimo elemento della seconda;
- un puntatore  $\uparrow$  quando l' $i$ -esimo elemento della prima sequenza é distinto dal  $j$ -esimo elemento della seconda ed il valore nella cella  $[i-1, j]$  della tabella principale é maggiore di quello nella cella  $[i, j-1]$ ;
- Un puntatore  $\leftarrow$  altrimenti.

Alla vecchia procedura opportunamente modificata ne viene aggiunta un'altra, PRINT-LCS. Questa ha in input la tabella  $b$ , la stringa  $X$  (o la stringa  $Y$ , é indifferente) e due indici di posizione  $i$  e  $j$ , mentre in output ha una delle possibili LCS per la stringa in input. La prima chiamata alla procedura ha  $|X|$  come valore per  $i$  e  $|Y|$  come valore per  $j$ , in modo da ottenere la soluzione per l'istanza  $|X|, |Y|$ . La procedura ripercorre la tabella  $b$  dalla cella  $(i, j)$  verso la cella  $(0, 0)$ ; ogni volta che si incontra il puntatore  $\nwarrow$ , il valore delle due sequenze in corrispondenza della cella di partenza é un elemento della soluzione.

Si noti come possano esistere piú percorsi all'interno della tabella. Il numero di percorsi dipende dal numero  $k$  di biforcazioni. Ogni biforcazione indica una possibile soluzione al problema LCS, eventualmente ridonanti. Il numero di percorsi non può essere superiore a  $2^k$ , ed il numero di possibili soluzioni distinte é non superiore al numero di possibili percorsi.

```

procedure LCS(X, Y)
  for i = 0 to |X| do
    c[i, 0] = 0
  for j = 0 to |Y| do
    c[0, j] = 0

  for i = 1 to |X| do
    for j = 1 to |Y| do
      if X[i] == Y[j] then
        c[i, j] = 1 + c[i - 1, j - 1]
        b[i, j] = "\nwarrow"
      else if c[i - 1, j] > c[i, j - 1]
        c[i, j] = c[i - 1, j]
        b[i, j] = "\u2191"
      else
        c[i, j] = c[i, j - 1]
        b[i, j] = "\u2190"

  return c, b

procedure PRINT-LCS(b, X, i, j)
  if i == 0 or j == 0
    return

  if b[i, j] == "\nwarrow"
    PRINT-LCS(b, X, i - 1, j - 1)
    print X[i]
  else if b[i, j] == "\u2191"
    PRINT-LCS(b, X, i - 1, j)
  else
    PRINT-LCS(b, X, i, j - 1)

```

Sebbene la procedura sia ricorsiva, il suo tempo di esecuzione é comunque proporzionale alle dimensioni della tabella  $b$ , perché la ricorsione é una tail-recursion.

## 1.2 Weighted Interval Scheduling

### 1.2.1 Definizione del problema

Sia dato un insieme  $X$  costituito da "attività". Ciascuna attività  $i \in X$  é definita a partire da una tripla  $(s_i, f_i, v_i)$ , dove i tre valori indicano rispettivamente il tempo di inizio dell'attività, il tempo di fine ed il suo peso. Due attività  $i$  e  $j$  distinte si dicono *compatibili* se  $i$  termina prima che  $j$  inizi, ovvero se  $f_i \leq s_j$ .

$$i \in X = (s_i, f_i, v_i)$$

$$\text{Comp}(i, j) = \begin{cases} T & \text{se } f_i \leq s_j \\ F & \text{se } f_i > s_j \end{cases}$$

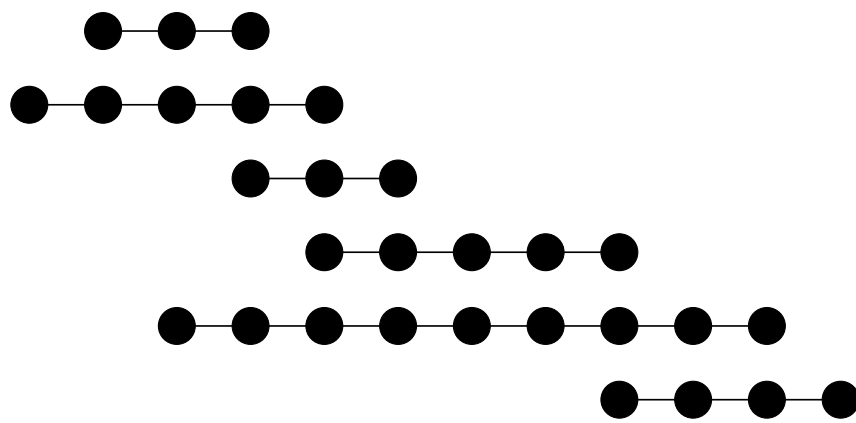
La nozioni di peso e di compatibilità vengono poi estese dai singoli elementi di  $X$  a sottoinsiemi  $A$  di  $X$ . Il peso di un insieme  $A \subseteq X$  é dato dalla somma dei pesi di tutte le attività di cui é costituito (per convenzione, se  $A$  é l'insieme vuoto, il suo peso é 0). Similmente, un insieme  $A \subseteq X$  si dice compatibile se contiene solo ed esclusivamente elementi di  $X$  tutti compatibili fra di loro o, equivalentemente, non contiene nemmeno una coppia di elementi di  $X$  che non sono fra loro compatibili.

$$V(A) = \begin{cases} \sum_{i=1}^{|A|} v_{a_i} \in A & \text{se } A \neq \emptyset \\ 0 & \text{se } A = \emptyset \end{cases}$$

$$\text{Comp}(A) = \begin{cases} T & \text{se } \nexists i, j \in A \text{ t.c. } f_i > s_j \\ F & \text{se } \exists i, j \in A \text{ t.c. } f_i > s_j \end{cases}$$

Il problema **Weighted Interval Scheduling** richiede di individuare qual'è, dato un insieme di attività, il sottoinsieme di attività fra loro compatibili che ha peso maggiore. Il problema può essere risolto in maniera banale osservando l'insieme potenza dell'insieme di attività (che ne contiene tutti i possibili sottoinsiemi), scartando tutti gli insiemi di cui è costituito che non sono compatibili, calcolando il peso di tutti gli insiemi compatibili e cercando il peso massimo.

Attività	Tempo iniziale	Tempo finale	Peso
$a_1$	1	3	10
$a_2$	2	0	4
$a_3$	8	3	5
$a_4$	1	4	8
$a_5$	1	2	10
$a_6$	3	8	11



Il sottoinsieme  $A = \{a_2, a_4, a_6\}$  è un insieme compatibile, ma non è il sottoinsieme con peso maggiore. Si osserva infatti che il sottoinsieme con queste caratteristiche è  $B = \{a_1, a_3, a_6\}$ , che ha peso pari a 21.

### 1.2.2 Programmazione dinamica: sottostruttura ottima

**Proprietà della sottostruttura ottima per il problema Weighted Interval Scheduling.** Sia data una istanza  $X_i = \{x_1, x_2, \dots, x_i\}$  del problema Weighted Interval Scheduling, con  $i \geq 1$ . Assunto di avere a disposizione tutte le soluzioni per le istanze precedenti ad  $X_i$ , ovvero  $S_1, S_2, \dots, S_{i-1}$ , si ha:

$$S_i = \begin{cases} S_{p(i)} \cup \{i\} & \text{se } i \in S_i \\ S_{i-1} & \text{se } i \notin S_i \end{cases}$$

**Dimostrazione.** L' $i$ -esimo elemento dell'istanza  $X_i$  può oppure non può fare parte della  $i$ -esima soluzione. Si distinguono allora due casi:

- $i \notin S_i$ . Si supponga per assurdo che  $S_{i-1}$  non sia la soluzione ottimale del problema per l'istanza  $X_i$ . Se così è, allora la soluzione ottimale per tale istanza (quale che sia) deve necessariamente avere un valore maggiore di quello della soluzione ottimale  $S_{i-1}$ , ovvero deve valere  $\text{Opt}(i) > \text{Opt}(i-1)$ .  
Se, come da ipotesi,  $i$  non appartiene ad  $S_i$ , deve aversi  $S_i \subseteq X_{i-1}$ . Ma allora  $S_i$  è una potenziale soluzione per la  $i-1$ -esima istanza del problema. Ricordando però che  $\text{Opt}(i) > \text{Opt}(i-1)$ , questo equivale a dire che  $S_i$  è una soluzione migliore per la  $i-1$ -esima istanza del problema di quanto lo sia  $S_{i-1}$ , e questo è in contraddizione con l'ipotesi che la soluzione ottimale per l'istanza  $X_{i-1}$  sia  $S_{i-1}$ . Occorre allora assumere che  $S_{i-1}$  sia l'effettiva soluzione ottimale per la  $i$ -esima istanza del problema quando  $i \notin S_i$ .
- $i \in S_i$ . Si supponga per assurdo che  $S_{p(i)} \cup \{i\}$  non sia la soluzione ottimale al problema per l'istanza  $X_i$ . Se così è, allora la soluzione ottimale per tale istanza (quale che sia) deve avere un valore maggiore di quello della soluzione ottimale  $S_{p(i)} \cup \{i\}$ , ovvero deve valere  $\text{Opt}(i) > \text{Opt}(p(i)) + v_i$ .  
Dato che, per ipotesi,  $i$  appartiene ad  $S_i$ , è possibile scomporre tale soluzione in  $S_h \cup \{i\}$ , con  $S_h \neq S_{p(i)}$ , così come è possibile scomporre  $\text{Opt}(i)$  in  $\text{Opt}(h) + v_i$ . Essendo però  $\text{Opt}(i) > \text{Opt}(p(i)) + v_i$ , deve valere  $\text{Opt}(h) + v_i > \text{Opt}(p(i)) + v_i$ , cioè  $\text{Opt}(h) > \text{Opt}(p(i))$ .  
Dovendo essere  $S_h \cup \{x_i\}$  un insieme costituito da elementi mutualmente compatibili, deve necessariamente valere  $S_h \subseteq X_{p(i)}$ . Ma allora  $S_h$  è una potenziale soluzione per la  $p(i)$ -esima istanza del problema. Ricordando però che  $\text{Opt}(h) > \text{Opt}(p(i))$ , questo equivale a dire che  $S_h$  è una soluzione migliore per la  $p(i)$ -esima istanza del problema di quanto lo sia  $S_{p(i)}$ , e questo è in contraddizione con l'ipotesi che la soluzione ottimale per l'istanza  $X_{p(i)}$  sia  $S_{p(i)}$ . Occorre allora assumere che  $S_{p(i)} \cup \{i\}$  sia l'effettiva soluzione ottimale per la  $i$ -esima istanza del problema quando  $i \in S_i$ .

### 1.2.3 Programmazione dinamica: equazione di ricorrenza

Dato un insieme di attività  $X$  avente cardinalità  $n$ , ordinate per tempo di fine, sia  $X_i$  l'insieme costituito dalle prime  $i$  attività di  $X$ . La soluzione  $S_i$  per la  $i$ -esima istanza del problema corrisponde a trovare la soluzione ottimale per il problema WIS rispetto al sottoinsieme  $X_i$ . La soluzione per l'intero insieme  $X$  è la soluzione per l'istanza  $X_n$ . Il valore totale della soluzione ottimale per la  $i$ -esima istanza del problema viene indicata con  $\text{Opt}(i)$ .

Il caso base è semplice da determinare; l'istanza  $X_0$ , che corrisponde all'insieme vuoto, ha per soluzione l'insieme vuoto stesso, che per definizione ha associato il valore 0.

$$S_0 = \emptyset$$

$$\text{Opt}(0) = 0$$

Per quanto riguarda il passo ricorsivo, si assuma di avere a disposizione tutte le soluzioni  $S_1, S_2, \dots, S_{i-1}$  e di voler calcolare  $S_i$ . Possono verificarsi solamente due situazioni, mutualmente esclusive: l' $i$ -esima attività fa parte della soluzione ottimale  $S_i$  oppure non ne fa parte.

Nel primo caso, è evidente come  $S_i$  e  $S_{i-1}$  debbano essere lo stesso insieme. Questo perché, essendo sia  $S_i$  sia  $S_{i-1}$  soluzioni ottimali, se  $i$  non è uno dei componenti della soluzione ottimale dell'istanza  $X_i$ , allora la soluzione ottimale di tale istanza dovrà essere la stessa dell'insieme  $X_i$  a cui viene tolto l'elemento  $i$ , ovvero  $X_{i-1}$ , la cui soluzione è proprio  $S_{i-1}$ .

Nel secondo caso, la soluzione ottimale dell'istanza  $X_i$  deve essere un insieme che contiene  $i$  e tutte le attività precedenti ad  $i$  che sono con questa compatibili. Tuttavia, dato che l'insieme  $S_i$  è soprainsieme di tutti gli insiemi  $S_j$  con  $j < i$ , per individuare il sottoinsieme di  $S_i$  che contiene tutte le attività compatibili con  $i$  è sufficiente cercare quello di indice maggiore, perché per definizione conterrà anche tutti i sottoinsiemi di elementi compatibili con  $i$  a loro volta più piccoli. L'indice con queste caratteristiche viene indicato con  $p(i)$ .

$$p(i) = \max\{j \mid j < i \wedge \text{Comp}(i, j) = T\}$$

Le equazioni di ricorrenza complete per gli insiemi  $S_i$  e per i valori  $\text{Opt}(i)$  possono allora essere scritte in questa forma:

$$S_i = \begin{cases} S_{p(i)} \cup \{i\} & \text{se } i \in S_i \\ S_{i-1} & \text{se } i \notin S_i \end{cases} \quad V(S_i) = \text{Opt}(i) = \begin{cases} \text{Opt}(p(i)) + v_i & \text{se } i \in S_i \\ \text{Opt}(i-1) & \text{se } i \notin S_i \end{cases}$$

Naturalmente non è possibile sapere a priori se l' $i$ -esimo elemento appartiene oppure non appartiene ad  $S_i$ ; l'unica informazione nota è il peso complessivo degli insiemi  $S_j$  con  $j < i$ . Dovendo però ricavare l'insieme avente valore ottimale, è sufficiente scegliere di volta in volta l'insieme  $S$  che permette di avere il valore massimo. Ovvero, se il valore dell'insieme  $S_{p(i)}$  a cui viene unito  $\{i\}$  è maggiore del valore dell'insieme  $S_{i-1}$ , allora  $S_{p(i)} \cup \{i\}$  è la scelta migliore, altrimenti è l'insieme  $S_{i-1}$  ad esserlo.

$$S_i = \begin{cases} S_{p(i)} \cup \{i\} & \text{se } \text{Opt}(p(i)) + v_i \geq \text{Opt}(i-1) \\ S_{i-1} & \text{se } \text{Opt}(p(i)) + v_i < \text{Opt}(i-1) \end{cases} \quad \text{Opt}(i) = \max\{\text{Opt}(p(i)) + v_i; \text{Opt}(i-1)\}$$

### 1.2.4 Programmazione dinamica: implementazione bottom-up

L'algoritmo bottom-up viene costruito a partire dall'equazione di ricorrenza sfruttando un vettore monodimensionale  $c$ . In ciascuna cella  $c[i]$  viene riportato il valore della soluzione ottimale per la  $i$ -esima istanza del problema, che viene utilizzata per calcolare i valori ottimali per le istanze successive. Tale procedura ha in input i valori associati alle attività (ordinate per tempo di fine) e restituisce in output il vettore  $c$ .

Si noti come la prima cella del vettore  $c$  possa venire riempito immediatamente con 0, come da caso base dell'equazione di ricorrenza.

```

procedure WIS(X)
  c[0] = 0

  for i = 1 to |X| do
    c[i] = max(c[i - 1], c[p[i]] + X[i])

  return c

```

È facile notare come il tempo di esecuzione dell'algoritmo sia  $O(n)$ , dove  $n$  è la lunghezza del vettore  $X$ . Questo perché è presente un ciclo che esegue una istruzione in tempo immediato esattamente  $|X|$  volte. Naturalmente, questo è possibile solamente se il vettore  $p$  è noto, altrimenti il tempo di esecuzione salirebbe a  $O(n^2)$ .

### 1.2.5 Programmazione dinamica: ricostruzione di una soluzione

Una volta calcolato il valore dell'insieme ottimale di intervalli, é possibile individuarne uno ripercorrendo la tabella a ritroso. In particolare, questo viene fatto introducendo una nuova procedura, PRINT-WIS, che ha in input la tabella  $c$ , il vettore  $X$  ed un indice  $i$ , mentre in output ha uno dei possibili insiemi di attività mutualmente compatibili aventi massimo valore totale. La prima chiamata alla procedura ha  $|X|$  come valore per  $i$ , in modo da ottenere la soluzione per l'istanza  $|X|$ .

La procedura ripercorre il vettore  $c$  dalla cella  $i$  verso la cella 0; ogni volta che si ha  $c[p[i]] + X[i] \geq c[i-1]$ , si ha che l' $i$ -esima attività é un elemento della soluzione.

```

procedure WIS(X)
  c[0] = 0

  for i = 1 to |X| do
    c[i] = max(c[i - 1], c[p[i]] + X[i])

  return c

procedure PRINT-WIS(c, X, i)
  if i ≠ 0
    if c[p[i]] + X[i] ≥ c[i - 1]
      PRINT-WIS(c, X, p[i])
      print i
    else
      PRINT-WIS(c, X, i - 1)
  else
    return

```

Sebbene la procedura sia ricorsiva, il suo tempo di esecuzione é comunque proporzionale alla lunghezza del vettore  $c$ , perché la ricorsione é una tail-recursion.

## 1.3 Hateville

### 1.3.1 Definizione del problema

Si consideri una via di una città, composta da  $n$  case. Per la costruzione di un ospedale si richiede agli  $n$  residenti una donazione. Ciascun  $i$ -esimo residente é disposto a donare una certa somma di denaro  $d_i$ , ma a condizione che entrambi i suoi vicini, l' $i-1$ -esimo e il  $i+1$ -esimo residente, si rifiutino di partecipare. Il problema **Hateville** richiede di trovare l'insieme di residenti tale per cui nessuno di questi é vicino di un altro ed al contempo permette di ricavare la massima quantità di denaro dalle donazioni.

Sia  $X = \{1, 2, \dots, n\}$  un insieme di  $n$  entità: a ciascuno é associato un valore  $d_i$ . Per qualsiasi insieme  $A \subseteq X$  é possibile definire un valore  $D(A)$  come la somma dei valori di tutte le entità costituiscono  $A$ . Due entità  $x_i$  e  $x_j$ , con  $i > j$ , si dicono *compatibili* se  $x_j \neq x_{i-1}$  e  $x_j \neq x_{i+1}$ . La nozione di compatibilità viene estesa ad un qualsiasi insieme  $A \subseteq X$  assegnandovi un valore booleano  $\text{Comp}(A)$ , che ha valore di verità  $T$  se tutte le entità che compongono  $A$  sono fra loro compatibili, e  $F$  altrimenti.

$$D(A) = \sum_{x_i \in A} d_i \qquad \text{Comp}(A) = \begin{cases} T & \text{se } x_{i-1} \notin A \wedge x_{i+1} \notin A \forall x_i \in A \\ F & \text{altrimenti} \end{cases}$$

La risoluzione del problema é data dall'insieme  $A^* \subset X$  formato esclusivamente da elementi fra loro compatibili che ha il massimo valore fra tutti gli insiemi con questa caratteristica:

$$\text{Comp}(A^*) = T \wedge D(A^*) = \max_{A \subset X \text{ t.c. } \text{Comp}(A)=T} \{D(A)\}$$

### 1.3.2 Programmazione dinamica: sottostruttura ottima



**Proprietá della sottostruttura ottima per il problema Hateville.** Sia data una istanza  $X_i = \{x_1, x_2, \dots, x_i\}$  del problema Hateville, con  $i \geq 2$ . Assunto di avere a disposizione tutte le soluzioni per le istanze precedenti ad  $X_i$ , ovvero  $S_1, S_2, \dots, S_{i-1}$ , si ha:

$$S_i = \begin{cases} S_{i-2} \cup \{i\} & \text{se } x_i \in S_i \\ S_{i-1} & \text{se } x_i \notin S_i \end{cases}$$

**Dimostrazione.** L' $i$ -esimo elemento dell'istanza  $X_i$  può oppure non può fare parte della  $i$ -esima soluzione. Si distinguono allora due casi:

- $x_i \notin S_i$ . Si supponga per assurdo che  $S_{i-1}$  non sia la soluzione ottimale del problema per l'istanza  $X_i$ . Se così é, allora la soluzione ottimale per tale istanza (quale che sia) deve necessariamente avere un valore maggiore di quello della soluzione ottimale  $S_{i-1}$ , ovvero deve valere  $\text{Opt}(i) > \text{Opt}(i-1)$ .  
Se, come da ipotesi,  $x_i$  non appartiene ad  $S_i$ , deve aversi  $S_i \subseteq X_{i-1}$ . Ma allora  $S_i$  é una potenziale soluzione per la  $i-1$ -esima istanza del problema. Ricordando però che  $\text{Opt}(i) > \text{Opt}(i-1)$ , questo equivale a dire che  $S_i$  é una soluzione migliore per la  $i-1$ -esima istanza del problema di quanto lo sia  $S_{i-1}$ , e questo é in contraddizione con l'ipotesi che la soluzione ottimale per l'istanza  $X_{i-1}$  sia  $S_{i-1}$ . Occorre allora assumere che  $S_{i-1}$  sia l'effettiva soluzione ottimale per la  $i$ -esima istanza del problema quando  $x_i \notin S_i$ .
- $x_i \in S_i$ . Si supponga per assurdo che  $S_{i-2} \cup \{x_i\}$  non sia la soluzione ottimale al problema per l'istanza  $X_i$ . Se così é, allora la soluzione ottimale per tale istanza (quale che sia) deve avere un valore maggiore di quello della soluzione ottimale  $S_{i-2} \cup x_i$ , ovvero deve valere  $\text{Opt}(i) > \text{Opt}(i-2) + d_i$ .  
Dato che, per ipotesi,  $x_i$  appartiene ad  $S_i$ , é possibile scomporre tale soluzione in  $S_h \cup \{x_i\}$ , con  $S_h \neq S_{i-2}$ , così come é possibile scomporre  $\text{Opt}(i)$  in  $\text{Opt}(h) + d_i$ . Essendo però  $\text{Opt}(i) > \text{Opt}(i-2) + d_i$ , deve valere  $\text{Opt}(h) + d_i > \text{Opt}(i-2) + d_i$ , cioè  $\text{Opt}(h) > \text{Opt}(i-2)$ .  
Dovendo essere  $S_h \cup \{x_i\}$  un insieme costituito da elementi mutualmente compatibili, la soluzione  $S_h$  non può contenere  $x_{i-1}$  (così come non può contenere  $x_i$ ), pertanto si ha  $S_h \subseteq X_{i-2}$ . Ma allora  $S_h$  é una potenziale soluzione per la  $i-2$ -esima istanza del problema. Ricordando però che  $\text{Opt}(h) > \text{Opt}(i-2)$ , questo equivale a dire che  $S_h$  é una soluzione migliore per la  $i-2$ -esima istanza del problema di quanto lo sia  $S_{i-2}$ , e questo é in contraddizione con l'ipotesi che la soluzione ottimale per l'istanza  $X_{i-2}$  sia  $S_{i-2}$ . Occorre allora assumere che  $S_{i-2} \cup \{x_i\}$  sia l'effettiva soluzione ottimale per la  $i$ -esima istanza del problema quando  $x_i \in S_i$ .

### 1.3.3 Programmazione dinamica: equazione di ricorrenza

Il problema può essere risolto mediante programmazione dinamica. Si consideri  $X_i = \{x_1, x_2, \dots, x_i\} \subseteq X$ , l'insieme costituito dai primi  $i$  elementi di  $X$ , che corrisponde alla  $i$ -esima istanza del problema. A tale insieme é associato l'insieme  $S_i = \{x_1, x_2, \dots, x_i\} \subseteq X_i$ , la soluzione ottimale per la  $i$ -esima istanza del problema. Il valore della soluzione  $S_i$  viene indicato con  $\text{Opt}(i)$ . La soluzione finale del problema é  $S_n$ , in quanto associata all'istanza  $X_n$ , che coincide con l'intero insieme  $X$ .

L'insieme  $S_i$  ed il valore  $\text{Opt}(i)$  vengono calcolati a partire da una equazione di ricorrenza. Il caso base di tale equazione é immediato: si consideri l'istanza  $X_0$  del problema, che non contiene alcuna entità. Tale insieme é certamente composto esclusivamente da entità fra loro compatibili, non contenendone alcuno. Inoltre, per lo stesso motivo, il valore di tale insieme é nullo.

$$S_0 = X_0 = \emptyset$$

$$D(S_0) = \text{Opt}(i) = 0$$

Per quanto riguarda la relazione di ricorrenza, si consideri una generica istanza  $X_i = \{x_1, x_2, \dots, x_i\}$  del problema, alla quale viene associata la soluzione ottimale  $S_i \subseteq \{x_1, x_2, \dots, x_i\}$  assumendo di avere a disposizione tutte le soluzioni ottimali  $S_j \subseteq \{x_1, x_2, \dots, x_j\}$  tali per cui  $j < i$ . Possono presentarsi due casistiche:  $x_i$  può oppure non può fare parte di  $S_i$ . Se  $x_i \notin S_i$ , allora é possibile assumere che la soluzione ottimale per la  $i$ -esima istanza sia la medesima della  $i-1$ -istanza.

Se invece  $x_i \in S_i$ , allora la soluzione ottimale  $S_{i-1} \subseteq \{x_1, x_2, \dots, x_{i-1}\}$  potrebbe dover venire esclusa, perché  $x_i$  e  $x_{i-1}$  sono elementi certamente incompatibili e  $S_{i-1}$  potrebbe contenere  $x_{i-1}$ . D'altro canto,  $x_i$  e  $x_{i-2}$  sono certamente compatibili, pertanto  $S_{i-2}$  é una possibile soluzione per la  $i$ -esima istanza, dato che non può contenere  $x_{i-1}$ . Tuttavia, proprio per il fatto che  $x_i$  e  $x_{i-2}$  sono compatibili, anche  $S_{i-2} \cup \{x_i\}$  é una soluzione, ed é certamente una soluzione migliore di  $S_{i-2}$ .

$$S_i = \begin{cases} S_{i-2} \cup \{x_i\} & \text{se } x_i \in S_i \\ S_{i-1} & \text{se } x_i \notin S_i \end{cases}$$

$$\text{Opt}(i) = \begin{cases} \text{Opt}(i-2) + d_i & \text{se } x_i \in S_i \\ \text{Opt}(i-1) & \text{se } x_i \notin S_i \end{cases}$$

Naturalmente non é possibile sapere a priori se  $x_i$  fa oppure non fa parte di  $S_i$ , tuttavia é possibile scegliere se includere oppure non includere  $x_i$  nella  $i$ -esima soluzione ottimale in base a quale delle due scelte rende maggiore il valore di  $S_i$ :

$$S_i = \begin{cases} S_{i-1} & \text{se } \text{Opt}(i-1) > \text{Opt}(i-2) + d_i \\ S_{i-2} \cup \{x_i\} & \text{altrimenti} \end{cases} \quad \text{Opt}(i) = \max\{\text{Opt}(i-1), \text{Opt}(i-2) + d_i\}$$

### 1.3.4 Programmazione dinamica: implementazione bottom-up

L'algoritmo bottom-up viene costruito a partire dall'equazione di ricorrenza sfruttando un vettore  $c$ . In ciascuna cella  $c[i]$  viene riportato il valore della soluzione ottimale per la  $i$ -esima istanza del problema, che viene utilizzata per calcolare i valori ottimali per le istanze successive. L'algoritmo riceve in input il vettore  $A$  e la funzione peso  $W$  e restituisce in output il vettore  $c$ .

Si noti come la prima cella del vettore  $c$  possa venire riempito immediatamente con 0, mentre la seconda con il valore associato al primo elemento di  $A$ , come da caso base dell'equazione di ricorrenza.

```
procedure HATEVILLE(A, W)
  c[0] = 0
  c[1] = W(A[1])

  for i = 2 to |A|
    c[i] = max{c[i - 1], c[i - 2] + W(A[i - 1])}

  return c
```

É facile notare come il tempo di esecuzione dell'algoritmo sia  $O(n)$ , dove  $n$  é la lunghezza del vettore  $A$ . Questo perché é presente un ciclo che esegue una istruzione in tempo immediato esattamente  $|A|$  volte.

### 1.3.5 Programmazione dinamica: ricostruzione di una soluzione

## 1.4 Longest Increasing Subsequence

### 1.4.1 Definizione del problema

Il problema della **Longest Increasing Subsequence (LIS)** prevede di trovare la piú lunga sottosequenza di una sequenza i cui elementi (nell'ordine in cui si trovano) hanno valore strettamente crescente.

Sia data la sequenza  $X = \langle 14, 2, 4, 2, 7, 0, 13, 21, 11 \rangle$ . Una sottosequenza strettamente crescente di  $X$  é  $\langle 2, 7, 11 \rangle$ , ma non é la piú lunga; tale sottosequenza é infatti  $\langle 2, 4, 7, 13, 21 \rangle$ .

Attaccare il problema LIS direttamente tramite programmazione dinamica é complesso. Un miglior approccio consiste nel restringere i requisiti del problema per considerarne una versione semplificata.

Una versione piú semplice del problema LIS, chiamato **LIS vincolato**, prevede di trovare la piú lunga sottosequenza di una sequenza i cui elementi (nell'ordine in cui si trovano) hanno valore strettamente crescente e dove l'ultimo elemento di tale sottosequenza coincide con l'ultimo elemento della sequenza originaria.

Sia data la sequenza  $X = \langle 14, 1, 4, 6, 13, 15, 0, 13, 29, 8 \rangle$ . La piú lunga sottosequenza strettamente crescente di  $X$  si rivela essere  $\langle 1, 4, 6, 13, 15, 29 \rangle$ . D'altro canto, la piú lunga sottosequenza strettamente crescente di  $X$  vincolata si rivela invece essere  $\langle 1, 4, 6, 8 \rangle$ .

### 1.4.2 Programmazione dinamica: sottostruttura ottima

**Proprietá della sottostruttura ottima per il problema Longest Increasing Subsequence.** Sia  $X$  una sequenza di  $m$  numeri interi e sia  $X_i$  un suo prefisso di lunghezza  $i$ , con  $1 \leq i \leq m$ . Sia  $Z^i$  una LIS di  $X_i$ , che termina con  $x_i$ . Sia infine  $W_i$  l'insieme di tutte le sottosequenze crescenti (non necessariamente le piú lunghe) di  $X_j$  che finiscono con  $x_j$  e a cui é possibile concatenare  $x_i$  ottenendo ancora una sequenza crescente, ovvero:

$$W_i = \bigcup_{1 \leq j < i; x_j < x_i} \{W \text{ sottosequenza crescente di } X_j \text{ che termina con } x_j\}$$

Allora vale che  $Z^i = Z^* \mid x_i$ , con  $Z^* \in W_i$  e  $|Z^*| = \max_{W \in W_i} |W|$ .

**Dimostrazione.** Si supponga per assurdo che  $Z^i = Z^* \mid x_i$  non sia la soluzione per l' $i$ -esimo problema. Questa deve allora essere  $Z^i = Z' \mid x_i$ , dove  $Z'$  é una sottosequenza crescente distinta da  $Z^*$ . Naturalmente deve aversi  $|Z| > |Z^*|$ , perché se  $Z^i$  si ottiene a partire da  $Z'$  e non da  $Z^*$  e  $Z^i$  é la sottosequenza crescente piú lunga, allora  $Z'$  deve essere piú lunga di  $Z^*$ . Sia  $z'$  l'elemento in coda a  $Z'$ . É chiaro che  $z' < x_i$ , perché se é stato possibile accodare  $x_i$  a  $Z'$  ottenendo ancora una sequenza crescente allora  $z'$  deve venire prima di  $x_i$ . Sia poi  $h < i$  il piú grande indice tale che  $x_h = z'$ . Di conseguenza, per come é stato definito  $W_i$ , si ottiene che  $Z' \in W_i$ . Infatti,  $Z'$  é una sottosequenza crescente di  $X_h$ , la quale termina con  $x_h < x_i$ . Ciò porta però ad una contraddizione: infatti si ha  $|Z'| > |Z^*|$  ma al contempo  $|Z^*| = \max_{W \in W_i} |W|$ .

### 1.4.3 Programmazione dinamica: equazione di ricorrenza

Data una sequenza  $X$  di lunghezza  $n$ , sia  $X_i = X[1 : i]$  il prefisso di lunghezza  $i$  di tale sequenza. La soluzione  $S_i$  per l' $i$ -esima istanza del problema corrisponde a trovare la soluzione ottimale per il problema LIS vincolato rispetto alla sottosequenza  $X_i$ . La soluzione per l'intera sequenza  $X$  é la soluzione per l'istanza  $X_n$ . La lunghezza della sottosequenza strettamente crescente piú lunga soluzione per la  $i$ -esima istanza del problema viene indicata con  $\text{Opt}(i)$ .

Il caso base dell'equazione di ricorrenza é immediato: se la sequenza é la sequenza vuota, allora la sottosequenza comune piú lunga vincolata é essa stessa la sequenza vuota.

$$S_0 = \varepsilon$$

$$\text{Opt}(0) = |S_0| = 0$$

Per quanto riguarda il passo ricorsivo, si assuma di avere a disposizione tutte le soluzioni  $S_{i-1}, S_{i-2}, \dots$  e di voler calcolare  $S_i$ . Dato che il problema LIS vincolato richiede di trovare la sottosequenza strettamente crescente piú lunga che termini con l'ultimo termine della sequenza, deve per forza aversi  $x_i \in S_i$ .

Una generica soluzione per la  $i$ -esima istanza del problema deve essere costruita accodando  $x_i$  ad sottosequenza strettamente crescente tale che tutti gli elementi di tale sottosequenza vengono prima, in ordine lessicografico, di  $x_i$ . Questo perché l'aggiunta di  $x_i$  in coda ad una tale sottosequenza restituisce ancora una sottosequenza strettamente crescente. La soluzione ottimale  $S_i$  é data dalla sottosequenza con queste caratteristiche piú lunga possibile a cui viene aggiunto  $x_i$  in coda.

La sottosequenza strettamente crescente piú lunga possibile con tutti gli elementi che vengono prima, in ordine lessicografico, di  $x_i$  sará a sua volta soluzione di una certa istanza  $h$  del problema, con  $h < i$ . La soluzione  $S_i$  é allora il risultato dell'accodare l'elemento  $x_i$  alla soluzione  $S_h$ . Se  $x_i$  viene dopo tutti gli altri elementi di  $X_i$  nell'ordine lessicografico, allora l'elemento  $x_h$  non esiste. In questo caso, la soluzione per la  $i$ -esima istanza del problema é formata dal solo elemento  $x_i$ .

$$S_i = \begin{cases} \max\{S_h \text{ t.c. } 1 \leq h < i \wedge x_h < x_i\} \cup \{x_i\} & \text{se } \exists x_h \\ \langle x_i \rangle & \text{se } \nexists x_h \end{cases} \quad \text{Opt}(i) = \begin{cases} \max\{\text{Opt}(h) \text{ t.c. } 1 \leq h < i \wedge x_h < x_i\} + 1 & \text{se } \exists x_{p(i)} \\ 1 & \text{se } \nexists x_{p(i)} \end{cases}$$

### 1.4.4 Programmazione dinamica: implementazione bottom-up

L'algoritmo bottom-up viene costruito a partire dall'equazione di ricorrenza sfruttando un vettore  $c$ . In ciascuna cella  $c[i]$  viene riportato il valore della soluzione ottimale per la  $i$ -esima istanza del problema, che viene utilizzata per calcolare i valori ottimali per le istanze successive. L'algoritmo riceve in input la sequenza  $X$  e restituisce in output il vettore  $c$ .

Si noti come tutte le celle del vettore  $c$  possono essere inizializzate subito ad 1, perché questo é il minimo valore che in una cella può essere presente.

```
procedure LIS(X)
  for i = 0 to |X| do
```

```

    c[i] = 1

    for i = 0 to |X| do
        p = 0
        for j = 0 to i do
            if (X[j] < X[i]) and (c[j] >= c[p])
                p = j
        if (p ≠ 0)
            c[i] = c[p] + 1

    return c

```

É facile notare come il tempo di esecuzione dell'algorithmo sia  $O(n^2)$ , dove  $n$  é la lunghezza della sequenza  $X$ . Il primo ciclo esegue una operazione immediata per  $n$  volte, dopodiché si presenta un doppio ciclo innestato: il ciclo esterno esegue il ciclo interno  $n$  volte, mentre il ciclo interno esegue al più  $n$  volte (dovendo potenzialmente ripercorrere l'intera sequenza a ritroso) un blocco di decisione il cui tempo di esecuzione può considerarsi immediato. Si ha allora che, asintoticamente,  $O(n) + O(n^2) = O(n^2)$ .

1.4.5 Programmazione dinamica: ricostruzione della soluzione

Una volta calcolata la lunghezza della LIS, é possibile individuarne una ripercorrendo il vettore a ritroso. In particolare, questo viene fatto estendendo l'algorithmo per il calcolo della lunghezza introducendo un vettore  $h$  Nella  $i$ -esima cella di  $h$  viene riportato l'indice della cella di  $c$  in cui si trova l'elemento predecessore dell'elemento nella  $i$ -esima cella di  $c$ . Se l'elemento predecessore non esiste, nella  $i$ -esima cella di  $h$  viene riportato 0.

Alla vecchia procedura opportunamente modificata ne viene aggiunta un'altra, PRINT-LIS. Questa ha in input il vettore  $h$ , la sequenza  $X$  ed un indice di posizione  $i$ , mentre in output ha una delle possibili LIS per la stringa in input. La prima chiamata alla procedura ha  $|X|$  come valore per  $i$ , in modo da ottenere la soluzione per l'istanza  $|X|$ .

La procedura ripercorre il vettore  $h$  dalla cella  $i$  verso la cella 0; ogni volta che si incontra un elemento non nullo di  $h$ , l'elemento della sequenza in tale posizione é un elemento della soluzione.

```

procedure LIS(X)
    for i = 0 to |X| do
        c[i] = 1
        h[i] = 0

    for i = 0 to |X| do
        p = 0
        for j = 0 to i do
            if (X[j] < X[i]) and (c[j] >= c[p])
                p = j
        if (p ≠ 0)
            c[i] = c[p] + 1
            h[i] = p

    return c, h

```

```

procedure PRINT-LIS(h, X, i)
    if h[i] ≠ 0 then
        PRINT-LIS(h, X, h[i])
    print X[i]

```

Sebbene la procedura sia ricorsiva, il suo tempo di esecuzione é comunque proporzionale alle dimensioni del vettore  $h$ , perché la ricorsione é una tail-recursion.

<14, 2, 4, 2, 7, 0, 13, 21, 20>

$X_0$	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$
0	1	1	2	1	3	1	4	5	5
0	0	0	2	0	3	0	5	7	7

1.4.6 Osservazioni

Il problema LIS non vincolato può essere ridotto al problema LIS vincolato. Infatti, la soluzione del problema LIS non vincolato non é altro che la maggior soluzione parziale del problema LIS vincolato.

1.5 Longest Increasing Common Subsequence

### 1.5.1 Definizione del problema

Il problema della **Longest Increasing Common Subsequence (LICS)** prevede di trovare la più lunga sottosequenza comune a due sequenze i cui elementi (nell'ordine in cui si trovano) hanno valore strettamente crescente.

Siano date le sequenze  $X = \langle 14, 2, 4, 2, 7, 0, 13, 21, 11 \rangle$  e  $Y = \langle 13, 2, 6, 5, 4, 11, 0 \rangle$ . Una sottosequenza strettamente crescente comune ad  $X$  e a  $Y$  di lunghezza massima è  $\langle 2, 4, 11 \rangle$ .

Attaccare il problema LICS direttamente tramite programmazione dinamica è complesso. Un miglior approccio consiste nel restringere i requisiti del problema per considerarne una versione semplificata.

Una versione più semplice del problema LICS, chiamato **LICS vincolato**, prevede di trovare la più lunga sottosequenza comune a due sequenze i cui elementi (nell'ordine in cui si trovano) hanno valore strettamente crescente e dove l'ultimo elemento di tale sottosequenza coincide con l'ultimo elemento di entrambe le sequenze originarie.

Siano date le sequenze  $X = \langle 14, 2, 4, 6, 13, 15, 0 \rangle$  e  $Y = \langle 13, 2, 6, 5, 4, 11, 0 \rangle$ . Una sottosequenza strettamente crescente di massima lunghezza comune ad  $X$  e ad  $Y$  è  $\langle 2, 4 \rangle$ . D'altro canto, la sottosequenza strettamente crescente di massima lunghezza comune ad  $X$  e ad  $Y$  vincolata è  $\langle 0 \rangle$ .

### 1.5.2 Programmazione dinamica: sottostruttura ottima

**Proprietà della sottostruttura ottima per il problema Longest Increasing Common Subsequence.** Sia  $X$  una sequenza di  $m$  numeri interi e sia  $X_i$  un suo prefisso di lunghezza  $i$ , con  $1 \leq i \leq m$ . Sia  $Y$  una sequenza di  $n$  numeri interi e sia  $Y_j$  un suo prefisso di lunghezza  $j$ , con  $1 \leq j \leq n$ .

Sia  $Z^{i,j}$  una LICS di  $X_i$  e di  $Y_j$  tale che termini con  $x_i = y_j$ . Sia infine  $W_{i,j}$  l'insieme di tutte le sottosequenze crescenti comuni (non necessariamente le più lunghe) di  $X_h$  e di  $Y_k$  che finiscono con  $x_h = y_k$  e a cui è possibile concatenare  $x_i$  (o  $y_j$ ) ottenendo ancora una sequenza crescente comune, ovvero:

$$W_{i,j} = \bigcup_{1 \leq h < i; 1 \leq k < j; x_h = y_k < x_i = y_j} \{W \text{ sottosequenza comune crescente di } X_h \text{ e di } Y_k \text{ che termina con } x_h = y_k\}$$

Allora vale che  $Z^{i,j} = Z^* \mid x_i = Z^* \mid y_j$ , con  $Z^* \in W_{i,j}$  e  $|Z^*| = \max_{W \in W_{i,j}} |W|$ .

**Dimostrazione.** Si supponga per assurdo che  $Z^i = Z^* \mid x_i = Z^* \mid y_j$  non sia la soluzione per l' $i, j$ -esimo problema. Questa deve allora essere  $Z^i = Z' \mid x_i = Z' \mid y_j$ , dove  $Z'$  è una sottosequenza crescente comune distinta da  $Z^*$ . Naturalmente deve aversi  $|Z'| > |Z^*|$ , perché se  $Z^i$  si ottiene a partire da  $Z'$  e non da  $Z^*$  e  $Z^i$  è la sottosequenza crescente comune più lunga, allora  $Z'$  deve essere più lunga di  $Z^*$ .

Sia  $z'$  l'elemento in coda a  $Z'$ . È chiaro che  $z' < x_i = y_j$ , perché se è stato possibile accodare  $x_i = y_j$  a  $Z'$  ottenendo ancora una sequenza crescente allora  $z'$  deve venire prima di  $x_i = y_j$ .

Siano poi  $r < i$  e  $s < j$  la più grande coppia di indici tali per cui  $x_r = y_s = z'$ . Di conseguenza, per come è stato definito  $W_{i,j}$ , si ottiene che  $Z' \in W_{i,j}$ . Infatti,  $Z'$  è una sottosequenza crescente comune di  $X_r$  e di  $Y_s$ , la quale termina con  $x_r < x_i = y_j$ . Ciò porta però ad una contraddizione: infatti si ha  $|Z'| > |Z^*|$  ma al contempo  $|Z^*| = \max_{W \in W_{i,j}} |W|$ .

### 1.5.3 Programmazione dinamica: equazione di ricorrenza

Date due sequenze  $X$  e  $Y$  di lunghezza rispettivamente  $n$  e  $m$ , siano  $X_i = X[1 : i]$  e  $Y_j = Y[1 : j]$  i prefissi di lunghezza  $i$  e  $j$  rispettivamente della sequenza  $X$  e della sequenza  $Y$ . La soluzione  $S_{i,j}$  per l' $i, j$ -esima istanza del problema corrisponde a trovare la soluzione ottimale per il problema LICS vincolato rispetto alle sottosequenze  $X_i$  e  $Y_j$ . La soluzione per le intere sequenze  $X$  e  $Y$  è la soluzione per le istanze  $X_n$  e  $Y_m$ . La lunghezza della sottosequenza strettamente crescente più lunga comune soluzione per la  $i, j$ -esima istanza del problema viene indicata con  $\text{Opt}(i, j)$ .

Certamente, la soluzione per tutte le coppie di indici  $i, j$  con  $i = 0$  oppure  $j = 0$  si rivela immediatamente essere la stringa vuota, perché questa è l'unica sottosequenza che una sequenza generica può avere in comune con la stringa vuota stessa.

Tuttavia, anche nel caso in cui  $x_i \neq y_j$ , ovvero quando le due sottosequenze  $X_i$  e  $Y_j$  non terminano con lo stesso carattere, la sottosequenza crescente comune più lunga è la stringa vuota. Questo perché il problema chiede di trovare una sottosequenza che termini

con il carattere (uguale) delle due stringhe, e se tale carattere é distinto l'unica soluzione accettabile non può che essere la stringa vuota.

Dato che, per definizione, l'ultimo elemento della stringa vuota (non esistendo) non può coincidere con l'ultimo carattere dell'altra sequenza, il caso  $i = 0 \vee j = 0$  rientra di fatto nel caso  $x_i \neq y_j$ , pertanto é sufficiente considerare solamente quest'ultimo come caso base.

$$S_{i,j} = \varepsilon \text{ se } x_i \neq y_j$$

$$\text{Opt}(i, j) = 0 \text{ se } x_i \neq y_j$$

Per quanto riguarda il passo ricorsivo, si assuma di avere a disposizione tutte le soluzioni nella forma  $S_{a,b}$ , con  $1 \leq a < i$  e  $1 \leq b < j$ , e di voler calcolare  $S_{i,j}$ . Dato che il problema LICS vincolato richiede di trovare la sottosequenza strettamente crescente comune piú lunga che termini con l'ultimo termine di entrambe le sequenze, deve per forza aversi  $x_i = y_j \in S_{i,j}$ .

Una generica soluzione per la  $i, j$ -esima istanza del problema deve essere costruita accodando  $x_i = y_j$  ad sottosequenza strettamente crescente comune alle due stringhe tale che tutti gli elementi di tale sottosequenza vengono prima, in ordine lessicografico, di  $x_i = y_j$ . Questo perché l'aggiunta di  $x_i = y_j$  in coda ad una tale sottosequenza restituisce ancora una sottosequenza strettamente crescente comune alle due. La soluzione ottimale  $S_{i,j}$  é data dalla sottosequenza con queste caratteristiche piú lunga possibile a cui viene aggiunto  $x_i = y_j$  in coda.

La sottosequenza strettamente crescente comune piú lunga possibile con tutti gli elementi che vengono prima, in ordine lessicografico, di  $x_i = y_j$  sarà a sua volta soluzione di una certa istanza  $h, k$  del problema, con  $h < i$  e  $k < j$ . La soluzione  $S_{i,j}$  é allora il risultato dell'accodare l'elemento  $x_i = y_j$  alla soluzione  $S_{h,k}$ . Se  $x_i = y_j$  viene dopo tutti gli altri elementi di  $X_i$  nell'ordine lessicografico, allora l'elemento  $x_h$  non esiste. In questo caso, la soluzione per la  $i, j$ -esima istanza del problema é formata dal solo elemento  $x_i = y_j$ .

$$S_{i,j} = \begin{cases} \max\{S_{h,k} \mid 1 \leq h < i, 1 \leq k < j, x_h < x_i\} \mid x_i & \text{se } \exists h, k \\ x_i & \text{se } \nexists h, k \end{cases} \quad \text{Opt}(i, j) = \begin{cases} \max\{\text{Opt}(h, k) \mid 1 \leq h < i, 1 \leq k < j, x_h < x_i\} + 1 & \text{se } \exists h, k \\ 1 & \text{se } \nexists h, k \end{cases}$$

#### 1.5.4 Programmazione dinamica: implementazione bottom-up

L'algoritmo bottom-up viene costruito a partire dall'equazione di ricorrenza sfruttando una tabella  $c$ . In ciascuna cella  $c[i, j]$  viene riportato il valore della soluzione ottimale per la  $i, j$ -esima istanza del problema, che viene utilizzata per calcolare i valori ottimali per le istanze successive. L'algoritmo riceve in input le due sequenze  $X$  e  $Y$  e restituisce in output la tabella  $c$  ed il valore ottimale.

```

procedure LICS(X, Y)
  max = 0
  for i = 1 to m do
    for j = 1 to n do
      if X[i] ≠ Y[j] then
        c[i, j] = 0
      else
        t = 0
        for h = 1 to i - 1 do
          for k = 1 to j - 1 do
            if (X[h] < X[i]) ∧ (c[h, k] > t) then
              t = c[h, k]
          c[i, j] = 1 + t
        if c[i, j] > max then
          max = c[i, j]
  return c, max

```

Il tempo di esecuzione dell'algoritmo é  $O(m^2n^2)$ , dove  $m$  é la lunghezza della stringa  $X$  e  $n$  é la lunghezza della stringa  $Y$ . Si noti infatti come l'algoritmo sia costituito da due cicli innestati, dove il primo esegue il secondo  $m$  volte ed il secondo esegue potenzialmente  $n$  volte a sua volta una coppia di cicli innestati, dove il primo esegue il secondo al piú  $m$  volte ed il secondo esegue al piú  $n$  volte un blocco di istruzioni avente tempo di esecuzione immediato. Si ha quindi  $O(m \cdot n \cdot (m \cdot n)) = O(m^2n^2)$ .

#### 1.5.5 Programmazione dinamica: ricostruzione di una soluzione

#### 1.5.6 Osservazioni

Il problema LICS non vincolato può essere ridotto al problema LICS vincolato. Infatti, la soluzione del problema LICS vincolato non é altro che la maggior soluzione parziale del problema LICS vincolato.

### 1.6 Sottovettore di peso massimo

### 1.6.1 Definizione del problema

Sia  $V = (v_1, v_2, \dots, v_n)$  un vettore composto da numeri interi (positivi e/o negativi). Il peso di un vettore é definito come la somma dei suoi elementi; per convenzione, il peso del vettore nullo si assume essere zero.

$$V(A) = \begin{cases} \sum_{i=1}^{|A|} v_i & \text{se } A \neq \emptyset \\ 0 & \text{se } A = \emptyset \end{cases}$$

Si dice **sottovettore** di  $V$  un qualunque vettore  $V' = (v_h, v_{h+1}, \dots, v_k)$  con  $1 \leq h \leq k \leq n$ . Sia  $SV(V)$  l'insieme che contiene tutti i sottovettori costruibili a partire da  $V$ .

$$SV(V) = \{(v_h, v_{h+1}, \dots, v_k) \mid 1 \leq h \leq k \leq n\}$$

Il **problema del sottovettore di peso massimo (HSV)** richiede di trovare, per un vettore  $V$ , il sottovettore  $V^*$  che, fra tutti i sottovettori di  $SV(V)$ , é quello avente peso complessivo massimo.

Dato il vettore  $V = (3, -5, 10, 2, -3, 1, 4, -8, 7, -6, 1)$ , il sottovettore di peso massimo é  $V^* = (10, 2, -3, 1, 4)$ , avente peso 14.

Attaccare il problema HSV direttamente tramite programmazione dinamica é complesso. Un miglior approccio consiste nel restringere i requisiti del problema per considerarne una versione semplificata.

Una versione piú semplice del problema HSV, chiamato **HSV vincolato**, richiede di trovare, fra tutti i sottovettori costruibili a partire da un vettore  $V$ , quello avente peso complessivo massimo e avente l'ultimo elemento coincidente con l'ultimo elemento del vettore originario.

### 1.6.2 Programmazione dinamica: equazione di ricorrenza

Dato un vettore  $V$  avente cardinalità  $n$ , sia  $V_i$  il sottovettore  $(v_1, v_2, \dots, v_i)$ . La soluzione  $S_i$  per l' $i$ -esima istanza del problema corrisponde a trovare la soluzione ottimale per il problema HSV vincolato rispetto al sottovettore  $V_i$ . La soluzione per l'intero vettore  $V$  é la soluzione per l'istanza  $V_n$ . Il peso del sottovettore di peso massimo per la  $i$ -esima istanza del problema viene indicato con  $\text{Opt}(i)$ .

Il caso base dell'equazione di ricorrenza é immediato: se il vettore é il vettore nullo, allora il sottovettore di peso massimo piú lungo vincolato é esso stesso il vettore nullo. D'altro canto, se il vettore é composto da un solo elemento, tale elemento deve essere parte per definizione della soluzione ottima per tale istanza del problema, ed é inoltre l'unico elemento di tale soluzione.

$$S_0 = \varepsilon$$

$$S_1 = (v_1)$$

$$\text{Opt}(0) = |S_0| = 0$$

$$\text{Opt}(1) = |S_1| = v_1$$

Per quanto riguarda il passo ricorsivo, si assuma di avere a disposizione tutte le soluzioni  $S_{i-1}, S_{i-2}, \dots$  e di voler calcolare  $S_i$ . Dato che il problema HSV vincolato richiede di trovare il sottovettore di peso massimo che termini con l'ultimo elemento del vettore, deve per forza aversi  $v_i \in S_i$ .

L'unica scelta da compiere in merito a tale soluzione é se includere o non includere anche la  $i-1$ -esima soluzione ottimale. Se tale soluzione ha un peso negativo, allora accodando  $S_{i-1}$  a  $v_i$  si otterrebbe un vettore con peso inferiore del solo  $v_i$ , e quindi una soluzione peggiore di  $(v_i)$ . Se tale soluzione ha invece un peso positivo, allora accodando  $S_{i-1}$  a  $v_i$  si otterrebbe un vettore con peso superiore del solo  $v_i$ , e quindi una soluzione migliore di  $(v_i)$ .

$$S_i = \begin{cases} v_i & \text{se } \text{Opt}(i-1) < 0 \\ S_{i-1} \cup v_i & \text{se } \text{Opt}(i-1) > 0 \end{cases}$$

$$\text{Opt}(i) = \begin{cases} v_i & \text{se } \text{Opt}(i-1) < 0 \\ \text{Opt}(i-1) + v_i & \text{se } \text{Opt}(i-1) > 0 \end{cases}$$

### 1.6.3 Programmazione dinamica: implementazione bottom-up

L'algoritmo bottom-up viene costruito a partire dall'equazione di ricorrenza sfruttando un vettore  $c$ . In ciascuna cella  $c[i]$  viene riportato il valore della soluzione ottimale per la  $i$ -esima istanza del problema, che viene utilizzata per calcolare i valori ottimali per le istanze successive. L'algoritmo riceve in input il vettore  $V$  e restituisce in output il vettore  $c$ .

```
procedure HSV(V)
  c[0] = 0
  c[1] = V[1]
```

```

for i = 2 to |V| do
  if (c[i - 1] < 0) then
    c[i] = V[i]
  else
    c[i] = c[i - 1] + V[i]

return c

```

É facile notare come il tempo di esecuzione dell'algoritmo sia  $O(n)$ , dove  $n$  é la lunghezza del vettore  $V$ . Questo perché tale algoritmo é composto da un solo ciclo che esegue per  $n$  volte una istruzione con tempo di esecuzione immediato.

#### 1.6.4 Programmazione dinamica: ricostruzione della soluzione

Una volta calcolate le lunghezze degli HSV vincolate, é possibile ricostruire ciascuna di queste a partire il vettore  $c$ . Alla vecchia procedura ne viene aggiunta un'altra, PRINT-HSV. Questa ha in input il vettore  $c$ , il vettore  $V$  ed un indice di posizione  $i$ , mentre in output ha il sottovettore di peso massimo per l' $i$ -esimo prefisso di  $V$ .

```

procedure HSV(V)
  c[0] = 0
  c[1] = V[1]

  for i = 2 to |V| do
    if (c[i - 1] < 0) then
      c[i] = V[i]
    else
      c[i] = c[i - 1] + V[i]

  return c

procedure PRINT-HSV(V, c, i)
  if c[i] != V[i] then
    PRINT-HSV(V, c, i - 1)
  print V[i]

```

Sebbene la procedura sia ricorsiva, il suo tempo di esecuzione é comunque proporzionale alle dimensioni del vettore  $c$ , perché la ricorsione é una tail-recursion.

#### 1.6.5 Osservazioni

Il problema HSV non vincolato può essere ridotto al problema HSV vincolato. Infatti, la soluzione del problema HSV non vincolato non é altro che la maggior soluzione parziale del problema HSV vincolato.

## 1.7 Knapsack 0/1

### 1.7.1 Definizione del problema

Sia dato un insieme  $X$  costituito da "oggetti". A ciascun oggetto  $i \in X$  é associato un valore  $v_i$  ed un ingombro  $w_i$ .

La nozioni di valore e di ingombro vengono poi estese dai singoli elementi di  $X$  a sottoinsiemi  $A$  di  $X$ . Il valore di un insieme  $A \subseteq X$  é dato dalla somma dei valori dei singoli oggetti di cui é costituito; allo stesso modo, l'ingombro di un insieme  $A \subseteq X$  é dato dalla somma dei pesi dei singoli oggetti di cui é costituito. Naturalmente, si assume che un insieme di oggetti nullo ha associato sia peso sia valore pari a zero.

$$V(A) = \begin{cases} \sum_{i=1}^{|A|} v_{a_i} & \text{se } A \neq \emptyset \\ 0 & \text{se } A = \emptyset \end{cases}$$

$$W(A) = \begin{cases} \sum_{i=1}^{|A|} w_{a_i} & \text{se } A \neq \emptyset \\ 0 & \text{se } A = \emptyset \end{cases}$$

Il problema **Knapsack 0/1** prevede di individuare il sottoinsieme di oggetti avente massimo valore complessivo e peso complessivo inferiore alla capacità dello zaino. Il numero totale di oggetti del sottoinsieme ed il suo peso totale (fintanto che questo é inferiore alla capacità dello zaino) non sono rilevanti ai fini della soluzione.



A	V	W
$A_1$	1	7
$A_2$	1	4
$A_3$	1	5
$A_4$	1	1
$A_5$	1	1

Si consideri uno zaino con capacità totale pari a 10. Il sottoinsieme  $\{A, B, C\}$  ha ingombro totale pari a  $7 + 4 + 5 = 16$ , pertanto non può essere una soluzione al problema. Il sottoinsieme  $\{A, E\}$  ha ingombro totale pari a  $7 + 1 = 8$ , pertanto è una possibile soluzione, per quanto non sia quella ottimale. Questa è infatti data dal sottoinsieme  $\{B, C, D\}$ , che ha ingombro totale pari a  $4 + 5 + 1 = 10$  e valore complessivo pari a  $1 + 1 + 1 = 3$ .

### 1.7.2 Programmazione dinamica: sottostruttura ottima

Si consideri l'insieme  $X_n$  composto da  $n$  oggetti. Data una soluzione ottimale  $S$  per tale insieme, vi sono due possibilità:

- $n$  è parte della soluzione, ovvero  $n \in S$  e  $w_n < C$ . Allora l'insieme  $S' = S \setminus \{n\}$  è la soluzione ottimale per l'istanza data dall'insieme di oggetti  $X \setminus \{n\} = X_{n-1}$  e da uno zaino di capacità  $C' = C - w_n$ .
- $n$  non è parte della soluzione, ovvero  $n \notin S$ . Allora l'insieme  $S$  è la soluzione ottimale per l'istanza data dall'insieme di oggetti  $X \setminus \{n\} = X_{n-1}$  e da uno zaino di capacità  $C$ .

**Dimostrazione.** La dimostrazione procede per casi:

- Innanzitutto,  $S'$  è compatibile con la capacità  $C - w_n$ . Infatti, avendosi  $W(S') = W(S) - w_n$ :

$$W(S) \leq C \Rightarrow W(S) - w_n \leq C - w_n \Rightarrow W(S') \leq C - w_n$$

Inoltre,  $S'$  è l'insieme compatibile avente valore totale massimo. Si assuma infatti per assurdo che esista un insieme  $S''$  tale che  $V(S'') > V(S')$  e  $W(S'') \leq C - w_n$ . Allora  $S'' \cup \{n\}$  è la soluzione ottimale per l'istanza  $X$  e  $C$  di valore totale maggiore di  $V(S)$ , e questo fatto è in contrasto con l'ipotesi che la soluzione ottimale per  $X$  sia  $S$ .

- Si assuma per assurdo che esista  $S'$ , una soluzione ottimale per l'istanza data dall'insieme  $X_{n-1}$  e capacità  $C$  tale per cui  $V(S') > V(S)$ . Ma allora  $S'$  è una soluzione ottimale per l'istanza data dall'insieme  $X$  e dalla capacità  $C$ , che è in contrasto con l'ipotesi che la soluzione ottimale per  $X$  sia  $S$ .

### 1.7.3 Programmazione dinamica: equazione di ricorrenza

Dato un insieme di oggetti  $X$  avente cardinalità  $n$ , ordinati per peso, sia  $X_i$  l'insieme costituito dalle prime  $i$  attività di  $X$ . Sia poi  $C$  un valore intero positivo, che rappresenta la capacità dello zaino. La soluzione  $S_{i,c}$  per la  $i$ -esima istanza del problema corrisponde a trovare la soluzione ottimale per il problema Knapsack 0/1 rispetto al sottoinsieme  $X_i$  e alla capacità massima  $c$ . La soluzione per l'intero insieme  $X$  è la soluzione per l'istanza  $X_n$ . Il valore totale della soluzione ottimale per la  $i$ -esima istanza del problema con capacità  $c$  viene indicata con  $\text{Opt}(i, c)$ .

Il caso base è semplice da determinare; se non vi è alcun oggetto da dover mettere nello zaino o se lo zaino non può contenere alcun oggetto, la soluzione ottimale è l'insieme vuoto.

$$S_{i,c} = \emptyset \text{ se } i = 0 \vee c = 0$$

$$\text{Opt}(i, c) = 0 \text{ se } i = 0 \vee c = 0$$

Per quanto riguarda il passo ricorsivo, si assuma di avere a disposizione tutte le soluzioni nella forma  $S_{a,b}$ , con  $1 \leq a < i$  e  $1 \leq b < c$ , e di voler calcolare  $S_{i,c}$ . Possono verificarsi solamente due situazioni, mutualmente esclusive: l' $i$ -esimo oggetto fa parte della soluzione ottimale  $S_{i,c}$  oppure non ne fa parte.

Se  $i$  non fa parte di  $S_{i,c}$ , allora la soluzione per la  $i$ -esima istanza è la medesima per l'istanza  $i-1$ -esima. Se invece  $i$  fa parte di  $S_{i,c}$ , allora tale soluzione è costruita a partire da  $S_{i-1,c-1}$  a cui viene aggiunto l' $i$ -esimo elemento e la cui capacità  $c$  è data dalla capacità della  $i-1$ -esima soluzione che viene diminuita del peso dell' $i$ -esimo oggetto.

$$S_{i,c} = \begin{cases} S_{i-1,c} & \text{se } i \notin S_{i,c} \\ S_{i-1,c-w_i} \cup \{i\} & \text{se } i \in S_{i,c} \end{cases} \quad \text{Opt}(i,c) = \begin{cases} \text{Opt}(i-1,c) & \text{se } i \notin S_{i,c} \\ \text{Opt}(i-1,c-w_i) + v_i & \text{se } i \in S_{i,c} \end{cases}$$

Naturalmente, non é possibile sapere a priori se l' $i$ -esimo elemento di  $X$  faccia oppure non faccia parte di  $S_{i,c}$ , a meno che il peso dell' $i$ -esimo oggetto sia maggiore della capacità  $c$ . In quel caso, si ha la certezza che  $i \notin S_{i,c}$ , e che di conseguenza la soluzione per la  $i$ -esima istanza del problema é la stessa per la  $i-1$ -esima. Se invece il peso dell'oggetto é inferiore o uguale a  $c$ , allora  $i$  può fare parte di  $S_{i,c}$ ; la scelta fra l'includere o l'escludere  $i$  da  $S_{i,c}$  é determinata da quale delle due restituisce un insieme  $S_{i,c}$  avente valore maggiore.

$$S_{i,c} = \begin{cases} S_{i-1,c} & \text{se } c > 0 \wedge w_i > c \\ \max\{S_{i-1,c-w_i} \cup \{i\}, S_{i-1,c}\} & \text{se } c > 0 \wedge w_i \leq c \end{cases} \quad \text{Opt}(i,c) = \begin{cases} \text{Opt}(i-1,c) & \text{se } c > 0 \wedge w_i > c \\ \max\{\text{Opt}(i-1,c-w_i) + v_i, \text{Opt}(i-1,c)\} & \text{se } c > 0 \wedge w_i \leq c \end{cases}$$

#### 1.7.4 Programmazione dinamica: implementazione bottom-up

L'algoritmo bottom-up viene costruito a partire dall'equazione di ricorrenza sfruttando una tabella  $d$ . In ciascuna cella  $c[i, j]$  viene riportato il valore della soluzione ottimale per la  $i, j$ -esima istanza del problema, che viene utilizzata per calcolare i valori ottimali per le istanze successive. L'algoritmo riceve in input l'insieme  $V$  dei valori degli oggetti, l'insieme  $W$  dei pesi degli oggetti e la capacità totale  $C$  e restituisce in output la tabella  $d$ .

```

procedure KNAPSACK(V, W, C)
  for i = 0 to |V| do
    for c = 0 to C do
      d[i, c] = 0

  for i = 1 to |V| do
    for c = 1 to C do
      d[i, c] = d[i - 1, c]
      temp = d[i - 1, c - W[i]] + V[i]
      if (c >= W[i]) and (temp > d[i, c]) then
        d[i, c] = temp

  return d

```

A	V	W
$A_1$	1	7
$A_2$	1	4
$A_3$	1	5
$A_4$	1	1
$A_5$	1	1

0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1	1	1	1
0	0	0	0	1	1	1	1	1	1	1
0	0	0	0	1	1	1	1	1	2	2
0	1	1	1	1	2	2	2	2	2	3
0	1	2	2	2	2	3	3	3	3	3

Il tempo di esecuzione dell'algoritmo é  $O(nC)$ , dove  $n$  é il numero di oggetti nell'insieme  $X$ . Infatti, l'algoritmo é costituito da una coppia di cicli for innestati, dove il ciclo piú esterno esegue il ciclo interno esattamente  $n$  volte, mentre il ciclo interno esegue una serie di operazioni aventi tempo di esecuzione unitario.

#### 1.7.5 Programmazione dinamica: ricostruzione di una soluzione

Una volta calcolato il valore dell'insieme ottimale di oggetti, é possibile individuarne uno ripercorrendo la tabella a ritroso. In particolare, questo viene fatto introducendo una nuova procedura, PRINT-KNAPSACK, che ha in input la tabella  $D$ , l'insieme dei pesi  $W$ , una capacità  $c$  ed un indice  $i$ , mentre in output ha uno dei possibili insiemi di oggetti aventi massimo valore totale. La prima chiamata alla procedura ha  $|X|$  come valore per  $i$  e  $C$  come valore per  $c$  in modo da ottenere la soluzione per l'istanza  $|X|, c$ .

La procedura ripercorre il vettore  $c$  dalla cella  $i$  verso la cella 0. Ogni volta che la capacità é maggiore del peso dell' $i$ -esimo oggetto e la  $i, c$ -esima cella di  $D$  non ha lo stesso valore della  $i-1, c$ -esima, allora l' $i$ -esimo oggetto é un elemento della soluzione.

```

procedure PRINT-KNAPSACK(D, W, i, c)
  if i > 0 and c > 0 then
    if c >= W[i] then

```

```

    if D[i, c] == D[i - 1, c] then
        PRINT-KNAPSACK(D, W, i - 1, c)
    else
        PRINT-KNAPSACK(D, W, i - 1, c - W[i])
        print i
else
    PRINT-KNAPSACK(D, W, i - 1, c)

```

Sebbene la procedura sia ricorsiva, il suo tempo di esecuzione é comunque proporzionale al numero di oggetti e alla capacità massima, perché la ricorsione é una tail-recursion.

## 1.8 Chiusura palindroma

### 1.8.1 Definizione del problema

Una stringa  $X = \{x_1, x_2, \dots, x_n\}$  é detta **palindroma** se é uguale alla stringa  $X' = \{x_n, \dots, x_2, x_1\}$ , che é composta dagli stessi caratteri di  $X$  ma in ordine inverso. Il **problema della chiusura palindroma** chiede di trovare, per una generica stringa  $X$ , il numero minimo di caratteri da inserirvi per renderla palindroma.

### 1.8.2 Programmazione dinamica: equazione di ricorrenza

Data una stringa  $X$  di lunghezza  $n$ , sia  $X_{i,j} = X[i : j]$  la sottostringa di  $X$  che ha il primo carattere in posizione  $i$  e l'ultimo carattere in posizione  $j$ . La soluzione  $S_{i,j}$  per l' $i, j$ -esima istanza del problema corrisponde a trovare la soluzione ottimale per il problema della chiusura palindroma rispetto alla sottostringa  $X_{i,j}$ . La soluzione per l'intera stringa  $X$  é la soluzione per l'istanza  $X_{1,n}$ .

Il caso base é immediato. Se la sottostringa  $X_{i,j}$  é la stringa vuota, ovvero se  $i > j$ , non vi é bisogno di aggiungere alcun carattere a  $X_{i,j}$  per renderla palindroma, perché la stringa vuota é per definizione palindroma. Similmente, se la sottostringa é formata da un solo carattere, ovvero se  $i = j$ , non vi é bisogno di aggiungere alcun carattere a  $X_{i,j}$  per renderla palindroma, perché una stringa di lunghezza unitaria é per definizione palindroma.

$$S_{i,j} = 0 \text{ se } i \geq j$$

Per quanto riguarda il passo ricorsivo, si assuma di avere a disposizione tutte le soluzioni nella forma  $S_{a,b}$ , con  $1 \leq a < i$  e  $1 \leq b < j$ , e di voler calcolare  $S_{i,j}$ . Possono verificarsi solamente due situazioni: il primo ed ultimo carattere di  $X_{i,j}$  sono uguali oppure sono diversi.

Se sono uguali, allora tale stringa é palindroma solo se lo é anche la sottostringa ottenuta eliminandovi tali caratteri, ovvero se lo é anche  $X_{i+1,j-1}$ . Se invece sono distinti, allora tale stringa é palindroma solo se lo é anche la sottostringa ottenuta eliminandovi il primo oppure l'ultimo carattere.

$$S_{i,j} = \begin{cases} S_{i+1,j-1} & \text{se } x_i = x_j \\ 1 + \min\{S_{i+1,j}, S_{i,j-1}\} & \text{se } x_i \neq x_j \end{cases}$$

### 1.8.3 Programmazione dinamica: implementazione bottom-up

L'algoritmo bottom-up viene costruito a partire dall'equazione di ricorrenza sfruttando una tabella  $c$ . In ciascuna cella  $c[i, j]$  viene riportato il valore della soluzione ottimale per la  $i, j$ -esima istanza del problema, che viene utilizzata per calcolare i valori ottimali per le istanze successive. L'algoritmo riceve in input la stringa  $X$  e restituisce in output  $c[1, |X|]$ , la cella che contiene la soluzione del problema.

```

procedure PAL-CLOSURE(X)
    for i = 0 to |X| do
        for j = 0 to |X| do
            c[i, j] = 0

    for i = |X| down to 1 do
        for j = 1 to |X| do
            if (i < j) then
                if (X[i] == X[j]) then
                    c[i, j] = c[i + 1, j - 1]
                else
                    c[i, j] = 1 + min(c[i + 1, j], c[i, j - 1])

```

```
return c[1, |X|]
```

1.9 Distanza di modifica

1.9.1 Definizione del problema

Il grado di somiglianza fra due sequenze viene genericamente chiamato **distanza**. Sebbene vi siano diverse formulazioni della distanza fra due sequenze, tutte possiedono le stesse proprietà di base. Matematicamente, la distanza é una metrica definita come la funzione  $d : S \times S \mapsto R^+$ , che assegna un valore reale ad ogni coppia di sequenze. La funzione distanza possiede le seguenti tre proprietà:

- 1.  $d(x, z) \leq d(x, y) + d(y, z)$ , la disuguaglianza triangolare;
- 2.  $d(x, y) = 0$  se e soltanto se  $x = y$ ;
- 3.  $d(x, y) = d(y, x)$ , la simmetricità.

La formulazione della distanza fra due sequenze più semplice e più utilizzata é la cosiddetta **distanza di modifica (edit distance)** <sup>1</sup>. Date due sequenze  $S_1$  e  $S_2$ , la distanza di modifica fra le due é data dal numero di operazioni atomiche sui singoli caratteri necessario a trasformare  $S_1$  in  $S_2$ . Le operazioni ammesse sono quattro:

- 1. *Inserire* un carattere della seconda sequenza nella prima;
- 2. *Rimuovere* un carattere dalla prima sequenza;
- 3. *Sostituire* un carattere della prima sequenza con uno della seconda;
- 4. non fare nulla (o equivalentemente, sostituire un carattere della prima sequenza con sé stesso).

Siano le quattro operazioni sopra citate abbreviate rispettivamente con  $I$ ,  $D$  e  $R$  e  $M$ . Una concatenazione di queste operazioni può essere espressa come una sequenza sull'alfabeto  $\{I, D, M, R\}$ . Ovvero, una sequenza del tipo  $op_1op_2op_3...op_n$ , costruita usando i quattro caratteri dell'alfabeto sopra descritto, é una sequenza dove ciascun carattere  $op_i$  indica l'operazione applicata all' $i$ -esimo carattere di  $S_1$  per poterla trasformare in  $S_2$ . Una sequenza che esprime la sequenza di operazioni necessarie a trasformare una sequenza in un'altra viene chiamata **trascrizione** delle due sequenze.

La distanza di modifica fra due sequenze può essere allora vista anche come la lunghezza della trascrizione più corta possibile per tale coppia di sequenze escludendo le operazioni  $M$ , ovvero il minimo numero di operazioni "rilevanti" (tutte tranne quelle nulle) necessarie a trasformare una sequenza in un'altra. La trascrizione (o le trascrizioni) che permette di trasformare una sequenza in un'altra nella maniera più efficiente possibile é quella che contiene il minimo numero di operazioni rilevanti.

R	I	M	D	M	D	M	M	I
v		i	n	t	n	e	r	
w	r	i		t		e	r	s

La sequenza  $S_1 = \text{vintner}$  può essere trasformata nella sequenza  $S_2 = \text{writers}$  mediante 5 operazioni: sostituire  $v$  con  $w$ , inserire  $r$ , eliminare  $n$ , eliminare  $n$  (di nuovo), inserire  $s$ . Le altre 4 operazioni sono no-op, perché quei caratteri in quelle posizioni sono già corretti. Pertanto, si ha che  $d(S_1, S_2) = 5$ .

Il **Problema della distanza di modifica** richiede di trovare, prese due sequenze, la lunghezza di una qualsiasi trascrizione ottimale che permette di trasformare la prima nella seconda.

1. Anche chiamata **distanza di Levenshtein**.

### 1.9.2 Programmazione dinamica: equazione di ricorrenza

Date due sequenze  $X$  e  $Y$  rispettivamente di lunghezza  $m$  e  $n$ , siano  $X_i = X[1 : i]$  e  $Y_j = Y[1 : j]$  i prefissi di lunghezza  $i$  e  $j$  delle rispettive sequenze. La soluzione  $S_{i,j}$  per l' $i, j$ -esima istanza del problema corrisponde a trovare la soluzione ottimale per il problema della distanza di modifica rispetto alle sottosequenze  $X_i$  e  $Y_j$ . La soluzione per le intere sequenze  $X$  e  $Y$  è la soluzione per l'istanza  $X_m, Y_n$ . La lunghezza trascrizione soluzione per la  $i, j$ -esima istanza del problema viene indicata con  $\text{Opt}(i, j)$ .

Il caso base della relazione di ricorrenza è immediato. Il numero di operazioni necessarie a trasformare la sequenza vuota in sé stessa è zero, perché non c'è alcuna operazione da compiere, e di conseguenza il relativo trascritto sarà la sequenza vuota. Similmente, per trasformare una sequenza generica di lunghezza  $i$  nella sequenza vuota sono necessarie  $i$  operazioni di tipo  $D$ , pertanto la lunghezza della relativa trascrizione è  $i$ . Infine, per trasformare la sequenza vuota in una sequenza generica di lunghezza  $j$  sono necessarie  $j$  operazioni di tipo  $I$ , pertanto la lunghezza della relativa trascrizione è  $j$ .

$$\begin{aligned} S_{0,0} &= \varepsilon \\ \text{Opt}(0,0) &= 0 \end{aligned}$$

$$\begin{aligned} S_{i,0} &= DDDDD\ldots \\ \text{Opt}(i,0) &= i \end{aligned}$$

$$\begin{aligned} S_{0,j} &= IIIII\ldots \\ \text{Opt}(0,j) &= j \end{aligned}$$

Per quanto riguarda il passo ricorsivo, si assuma di avere a disposizione tutte le soluzioni nella forma  $S_{a,b}$ , con  $1 \leq a < i$  e  $1 \leq b < j$ , e di voler calcolare  $S_{i,j}$ . Possono verificarsi solamente due situazioni, mutualmente esclusive: l' $i$ -esimo carattere di  $X_i$  è uguale al  $j$ -esimo carattere di  $Y_j$  oppure l' $i$ -esimo carattere di  $X_i$  è diverso dal  $j$ -esimo carattere di  $Y_j$ .

Nel primo caso, la  $i, j$ -esima soluzione coincide con la  $i-1, j-1$ -esima. Questo perché è l'unica situazione in cui è possibile non aggiungere alcun carattere alla trascrizione, e dato che questa può solo crescere di lunghezza o rimanere invariata, non aggiungere alcun carattere è la scelta che per definizione ne minimizza la lunghezza.

Nel secondo caso, la soluzione ottimale per la  $i, j$ -esima istanza deve necessariamente essere più lunga di tutte le soluzioni per istanze precedenti. Le uniche soluzioni candidate per tale istanza sono  $S_{i-1,j}$ ,  $S_{i,j-1}$  e  $S_{i-1,j-1}$ , alle quali viene accodato uno dei caratteri  $I$ ,  $D$  o  $R$ . Dato che sono tutte valide, il criterio di scelta consiste nello scegliere la soluzione con lunghezza minore.

$$S_{i,j} = \begin{cases} S_{i-1,j-1} & \text{se } X[i] = Y[j] \\ \min \begin{cases} S_{i-1,j} \cup \{D\} \\ S_{i,j-1} \cup \{I\} \\ S_{i-1,j-1} \cup \{R\} \end{cases} & \text{altrimenti} \end{cases}$$

$$\text{Opt}(i,j) = \begin{cases} \text{Opt}(i-1,j-1) & \text{se } X[i] = Y[j] \\ 1 + \min \begin{cases} \text{Opt}(i-1,j) \\ \text{Opt}(i,j-1) \\ \text{Opt}(i-1,j-1) \end{cases} & \text{altrimenti} \end{cases}$$

### 1.9.3 Programmazione dinamica: implementazione bottom-up

L'algoritmo bottom-up viene costruito a partire dall'equazione di ricorrenza sfruttando una tabella  $c$ . In ciascuna cella  $c[i,j]$  viene riportato il valore della soluzione ottimale per la  $i, j$ -esima istanza del problema, che viene utilizzata per calcolare i valori ottimali per le istanze successive. L'algoritmo riceve in input le due sequenze  $X$  e  $Y$  e restituisce in output la tabella  $c$ .

Si noti come le celle che hanno 0 come uno dei due indici possano venire riempite immediatamente con il valore dell'altro indice, come da caso base dell'equazione di ricorrenza.

```
procedure ED(X, Y)
  for i = 0 to |X| do
    c[i][0] = i
  for j = 0 to |Y| do
    c[0][j] = j

  for i = 1 to |X| do
    for j = 1 to |Y| do
      if X[i] == Y[j] then
        c[i][j] = c[i-1][j-1]
      else
        c[i][j] = 1 + min{c[i-1][j], c[i][j-1], c[i-1][j-1]}

  return c
```

È facile notare come il tempo di esecuzione dell'algoritmo sia  $O(nm)$ , dove  $n$  e  $m$  sono le lunghezze di rispettivamente la stringa  $X$  e la stringa  $Y$ . I primi due cicli eseguono una operazione immediata rispettivamente per  $n$  e per  $m$  volte, dopodiché si presenta un doppio ciclo innestato: il ciclo esterno esegue il ciclo interno  $m$  volte, mentre il ciclo interno esegue  $n$  volte un blocco di decisione il cui tempo di esecuzione può considerarsi immediato. Si ha allora che, asintoticamente,  $O(m) + O(n) + O(mn) = O(mn)$ .

		s	u	n	d	a	y
	0	1	2	3	4	5	6
s	1	0	1	2	3	4	5
a	2	1	1	2	3	3	4
t	3	2	2	2	3	4	4
u	4	3	2	3	3	4	5
r	5	4	3	3	4	4	5
d	6	5	4	4	3	4	5
a	7	6	5	5	4	3	4
y	8	7	6	6	5	4	3

#### 1.9.4 Programmazione dinamica: ricostruzione di una soluzione

Una volta calcolata la lunghezza della trascrizione ottima, é possibile individuarne una ripercorrendo la tabella a ritroso. In particolare, questo viene fatto estendendo l'algoritmo per il calcolo della lunghezza introducendo una nuova tabella  $b$  di puntatori. In particolare, nella cella  $(i, j)$  della tabella  $b$  viene inserito:

- Un puntatore ↖ se  $c(i, j) = c(i-1, j-1) + 1$  o se  $c(i, j) = c(i-1, j-1)$
- un puntatore ← se  $c(i, j) = c(i-1, j) + 1$
- Un puntatore ↑ se  $c(i, j) = c(i, j-1) + 1$ ,

Si noti come vi siano celle in cui potrebbero essere inseriti più puntatori, perché più condizioni possono verificarsi contemporaneamente. Per convenzione, é possibile stabilire un ordine di priorità con cui scegliere un puntatore piuttosto che un altro; questo porta a diverse possibili ricostruzioni, ma queste avranno comunque la stessa lunghezza.

Per ricavare una trascrizione ottimale a partire dalla tabella, é sufficiente seguire un qualsiasi percorso definito dai puntatori dalla cella  $(n, m)$  alla cella  $(0, 0)$ . Se ci si sposta orizzontalmente, ovvero da una cella  $(i, j)$  ad una cella  $(i, j-1)$ , allora significa che l'operazione più efficiente per quel passaggio é un inserimento. Se ci si sposta verticalmente, ovvero da una cella  $(i, j)$  ad una cella  $(i-1, j)$  allora significa che l'operazione più efficiente per quel passaggio é una rimozione. Se ci si sposta diagonalmente, ovvero da una cella  $(i, j)$  ad una cella  $(i-1, j-1)$ , allora significa che l'operazione più efficiente per quel passaggio é una sostituzione se  $X[i] \neq Y[j]$  o una no-op altrimenti.

Alla vecchia procedura opportunamente modificata ne viene aggiunta un'altra, PRINT-ED. Questa ha in input la tabella  $b$ , la stringa  $X$ , la stringa  $Y$  e due indici di posizione  $i$  e  $j$ , mentre in output ha una delle possibili trascrizioni ottimali per le stringhe in input. La prima chiamata alla procedura ha  $|X|$  come valore per  $i$  e  $|Y|$  come valore per  $j$ , in modo da ottenere la soluzione per l'istanza  $|X|, |Y|$ . Si noti come la trascrizione restituita dalla procedura abbia l'ordine dei caratteri invertito, perché questi sono restituiti dall'ultimo al primo.

```

procedure ED(X, Y)
  for i = 0 to |X| do
    c[i, 0] = i
    b[i, j] = "←"
  for j = 0 to |Y| do
    c[0, j] = j
    b[i, j] = "↑"

  for i = 1 to |X| do
    for j = 1 to |Y| do
      if X[i] == Y[j] then
        c[i, j] = c[i - 1, j - 1] + 1
        b[i, j] = "↖"
      else
        if (c[i - 1, j - 1] <= c[i, j - 1]) and
           (c[i - 1, j - 1] <= c[i - 1, j]) then
          c[i, j] = c[i - 1, j - 1] + 1
          b[i, j] = "↖"
        else if (c[i - 1, j] <= c[i, j - 1]) and
                 (c[i - 1, j] <= c[i - 1, j - 1]) then
          c[i, j] = c[i - 1, j] + 1
          b[i, j] = "←"
        else
          c[i, j] = c[i, j - 1] + 1
          b[i, j] = "↑"

  return c, b

```

```

procedure PRINT-ED(X, Y, i, j, b)
  if i == 0 or j == 0
    return

  if b[i, j] == "←" then
    print "I"
    PRINT-ED(X, Y, i - 1, j, b)
  else if b[i, j] == "↑" then
    print "D"
    PRINT-ED(X, Y, i, j - 1, b)
  else
    if X[i] != Y[j] then
      print "R"
      PRINT-ED(X, Y, i - 1, j - 1, b)

```

Sebbene la procedura sia ricorsiva, il suo tempo di esecuzione é comunque proporzionale alle dimensioni della tabella  $b$ , perché la ricorsione é una tail-recursion.

		w	r	i	t	e	r	s
	0	1	2	3	4	5	6	7
v	1	1	2	3	4	5	6	7
i	2	2	2	2	3	4	5	6
n	3	3	3	3	3	4	5	6
t	4	4	4	4	3	4	5	6
n	5	5	5	5	4	4	5	6
e	6	6	6	6	5	4	5	6
r	7	7	6	7	6	5	4	5

		w	r	i	t	e	r	s
		←	←	←	←	←	←	←
v	↑	↖	↖	↖	↖	↖	↖	↖
i	↑	↖	↖	↖	←	←	←	←
n	↑	↖	↖	↖	↑	←	←	←
t	↑	↖	↖	↖	↖	↖	↖	↖
n	↑	↖	↖	↖	↑	↖	↖	↖
e	↑	↖	↖	↖	↑	↖	↖	↖
r	↑	↖	↖	↖	↑	↑	↖	←

## 1.10 Problema Interleaving

### 1.10.1 Definizione del problema

Date le sequenze  $X = \langle x_1, \dots, x_m \rangle$  e  $Y = \langle y_1, \dots, y_n \rangle$ , la sequenza  $W = \langle w_1, \dots, w_{m+n} \rangle$  é detta **interleaving** di  $X$  e di  $Y$  se é possibile trovarle entrambe come sottosequenze disgiunte all'interno di  $W$ .

Piú formalmente,  $W$  é interleaving di  $X$  e di  $Y$  se esistono:

$$\{i_1, \dots, i_m\} \subseteq \{1, \dots, m+n\} \text{ con } i_1 < i_2 < \dots < i_m$$

$$\{j_1, \dots, j_n\} \subseteq \{1, \dots, m+n\} \text{ con } j_1 < j_2 < \dots < j_n$$

Tali per cui:

$$\{i_1, \dots, i_m\} \cap \{j_1, \dots, j_n\} = \emptyset$$

$$\{i_1, \dots, i_m\} \cup \{j_1, \dots, j_n\} = \{1, \dots, m+n\}$$

$$\forall k \in \{1, \dots, m\} : x_k = w_{i_k}$$

$$\forall h \in \{1, \dots, n\} : y_h = w_{j_h}$$

Il **problema interleaving** chiede di determinare se, per una tripla di sequenze  $W$ ,  $X$  e  $Y$ , la sequenza  $W$  é interleaving di  $X$  e di  $Y$ .

Date le sequenze  $Y = \langle D, E, E, P \rangle$  e  $X = \langle L, O, R, E \rangle$ , un interleaving di  $X$  e di  $Y$  é la sequenza  $W = \langle D, L, O, E, R, E, P, E \rangle$ .

### 1.10.2 Programmazione dinamica: equazione di ricorrenza

Date tre sequenze  $X$ ,  $Y$  e  $W$ , rispettivamente di lunghezza  $m$ ,  $n$  e  $m+n$ , siano  $X_i = X[1 : i]$  e  $Y_j = Y[1 : j]$  i prefissi di lunghezza  $i$  e  $j$  delle rispettive sequenze. La soluzione  $S_{i,j}$  per l' $i, j$ -esima istanza del problema corrisponde a determinare se  $W_{i+j}$  é interleaving per le sottosequenze  $X_i$  e  $Y_j$ . La soluzione per le intere sequenze  $X$  e  $Y$  é la soluzione per l'istanza  $X_m, Y_n$ .

$$S_{i,j} = T \text{ se } W_{i+j} \text{ é interleaving di } X_i \text{ e } Y_j, \text{ altrimenti } S_{i,j} = F.$$

Approcciare il problema direttamente a partire dal determinare casi base e passi ricorsivi é ostico; é piú semplice trattare il problema in termini di prefissi delle tre stringhe. Si consideri innanzitutto il caso in cui almeno una delle due sequenze é la sequenza vuota:

- Se  $X_i$  e  $Y_j$  sono entrambe la sequenza vuota, allora l'unica sequenza  $W_{i+j}$  costruibile a partire da queste é essa stessa la sequenza vuota, che per definizione é interleaving di  $X_i$  e di  $Y_j$ ;
- Se  $X_i$  é la sequenza vuota ma  $Y_j$  non lo é, occorre distinguere due situazioni diverse:

1. Se  $W[j] = Y[j]$ , allora  $W_j$  é interleaving di  $X_i$  e  $Y_j$  soltanto se  $W_{i+j-1}$  era interleaving di  $X_i$  e  $Y_{j-1}$ ;
  2. Se  $W[j] \neq Y[j]$ , allora  $W_j$  non può essere interleaving di  $X_i$  e  $Y_j$ .
- Se  $Y_j$  é la sequenza vuota ma  $X_i$  non lo é, occorre distinguere due situazioni diverse:
    1. Se  $W[i] = X[i]$ , allora  $W_i$  é interleaving di  $X_i$  e  $Y_j$  soltanto se  $W_{i+j-1}$  era interleaving di  $X_{i-1}$  e  $Y_j$ ;
    2. Se  $W[i] \neq X[i]$ , allora  $W_i$  non può essere interleaving di  $X_i$  e  $Y_j$ .

Riassumendo:

$$S_{i,j} = \begin{cases} T & \text{se } i = 0 \wedge j = 0 \\ S_{i,j-1} & \text{se } i = 0 \wedge j > 0 \wedge W[j] = Y[j] \\ F & \text{se } i = 0 \wedge j > 0 \wedge W[j] \neq Y[j] \\ S_{i-1,j} & \text{se } i > 0 \wedge j = 0 \wedge W[i] = X[i] \\ F & \text{se } i > 0 \wedge j = 0 \wedge W[i] \neq X[i] \end{cases}$$

Si consideri invece il caso in cui entrambe le sequenze non sono la sequenza vuota. In tal senso, si assuma di avere a disposizione tutte le soluzioni nella forma  $S_{a,b}$ , con  $1 \leq a < i$  e  $1 \leq b < j$ , e di voler calcolare  $S_{i,j}$ . Possono presentarsi quattro possibilità mutualmente esclusive:

- $W[i+j] \neq X[i]$  e  $W[i+j] \neq Y[j]$ . Se non é possibile accodare né  $X[i]$  né  $Y[j]$  a  $W[i+j]$  allora  $W_{i+j}$  non può essere interleaving di  $X_i$  e di  $Y_j$ ;
- $W[i+j] = X[i]$  e  $W[i+j] \neq Y[j]$ . Siccome l'elemento  $W[i+j]$  é uguale all'elemento  $X[i]$  e diverso da  $Y[j]$ , allora l'unico carattere valido per la posizione  $i+j$  in  $W$  é proprio  $X[i]$ . Quindi  $W_{i+j}$  é interleaving di  $X_i$  e di  $Y_j$  solo se  $W_{i+j-1}$  era interleaving di  $X_{i-1}$  e  $Y_j$ ;
- $W[i+j] \neq X[i]$  e  $W[i+j] = Y[j]$ . Siccome l'elemento  $W[i+j]$  é uguale all'elemento  $Y[j]$  e diverso da  $X[i]$ , allora l'unico carattere valido per la posizione  $i+j$  in  $W$  é proprio  $Y[j]$ . Quindi  $W_{i+j}$  é interleaving di  $X_i$  e di  $Y_j$  solo se  $W_{i+j-1}$  era interleaving di  $X_i$  e  $Y_{j-1}$ ;
- $W[i+j] = X[i]$  e  $W[i+j] = Y[j]$ . In questo caso, é possibile accodare sia  $X[i]$  che  $Y[j]$  al simbolo in posizione  $i+j$  in  $W$ . Pertanto, la soluzione all' $i, j$ -esima istanza é data dall'OR logico fra la  $i, j-1$ -esima soluzione e la  $i-1, j$ -esima.

Riassumendo:

$$S_{i,j} = \begin{cases} F & \text{se } W[i+j] \neq X[i] \wedge W[i+j] \neq Y[j] \\ S_{i-1,j} & \text{se } W[i+j] = X[i] \wedge W[i+j] \neq Y[j] \\ S_{i,j-1} & \text{se } W[i+j] \neq X[i] \wedge W[i+j] = Y[j] \\ S_{i-1,j} \vee S_{i,j-1} & \text{se } W[i+j] = X[i] \wedge W[i+j] = Y[j] \end{cases}$$

Alla luce di quanto analizzato finora, é possibile mettere a sistema le differenti equazioni e definire un caso base ed un passo ricorsivo:

$$S_{i,j} = \begin{cases} T & \text{se } i = 0 \wedge j = 0 \\ F & \text{se } i = 0 \wedge j > 0 \wedge W[j] \neq Y[j] \\ F & \text{se } i > 0 \wedge j = 0 \wedge W[i] \neq X[i] \\ F & \text{se } W[i+j] \neq X[i] \wedge W[i+j] \neq Y[j] \end{cases} \quad S_{i,j} = \begin{cases} S_{i,j-1} & \text{se } i = 0 \wedge j > 0 \wedge W[j] = Y[j] \\ S_{i-1,j} & \text{se } i > 0 \wedge j = 0 \wedge W[i] = X[i] \\ S_{i-1,j} & \text{se } W[i+j] = X[i] \wedge W[i+j] \neq Y[j] \\ S_{i,j-1} & \text{se } W[i+j] \neq X[i] \wedge W[i+j] = Y[j] \\ S_{i-1,j} \vee S_{i,j-1} & \text{se } W[i+j] = X[i] \wedge W[i+j] = Y[j] \end{cases}$$

### 1.10.3 Programmazione dinamica: implementazione bottom-up

L'algoritmo bottom-up viene costruito a partire dall'equazione di ricorrenza sfruttando una tabella  $c$ . In ciascuna cella  $c[i, j]$  viene riportato il valore della soluzione ottimale per la  $i, j$ -esima istanza del problema, che viene utilizzata per calcolare i valori ottimali per le istanze successive. L'algoritmo riceve in input le due sequenze  $X$  e  $Y$  e restituisce in output il valore  $c[|X|, |Y|]$ , che corrisponde alla soluzione del problema per l'istanza completa.

```
procedure INTERLEAVING(X, Y, W)
  c[0, 0] = T

  for i = 1 to |X| do
    if (W[i] == X[i]) then
      c[i, 0] = c[i - 1, 0]
```



```

else
    c[i, 0] = F

for j = 1 to |Y| do
    if (W[j] == Y[j]) then
        c[0, j] = c[0, j - 1]
    else
        c[0, j] = F

for i = 1 to |X| do
    for j = 1 to |Y| do
        if (W[i + j] != X[i] and W[i + j] != Y[j]) then
            c[i, j] = F
        else if (W[i + j] == X[i] and W[i + j] != Y[j]) then
            c[i, j] = c[i - 1, j]
        else if (W[i + j] != X[i] and W[i + j] == Y[j]) then
            c[i, j] = c[i, j - 1]
        else
            c[i, j] = (c[i - 1, j] or c[i, j - 1])

return c[|X|, |Y|]

```

É facile notare come il tempo di esecuzione dell'algoritmo sia  $O(nm)$ , dove  $m$  e  $n$  sono le lunghezze di rispettivamente la stringa  $X$  e la stringa  $Y$ . I primi due cicli eseguono una operazione immediata rispettivamente per  $m$  e per  $n$  volte, dopodiché si presenta un doppio ciclo innestato: il ciclo esterno esegue il ciclo interno  $m$  volte, mentre il ciclo interno esegue  $n$  volte un blocco di decisione il cui tempo di esecuzione può considerarsi immediato. Si ha allora che, asintoticamente,  $O(m) + O(n) + O(mn) = O(mn)$ .

## 1.11 Cammino su un grafo

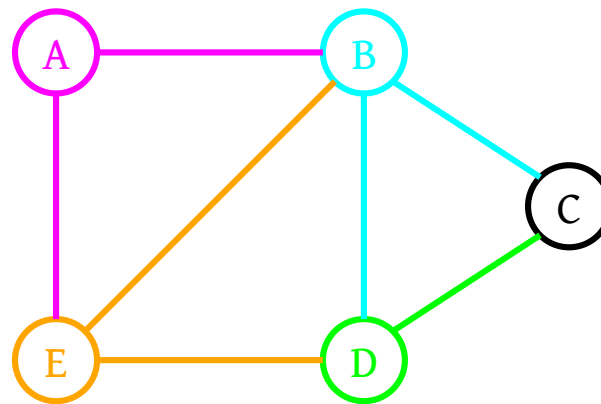
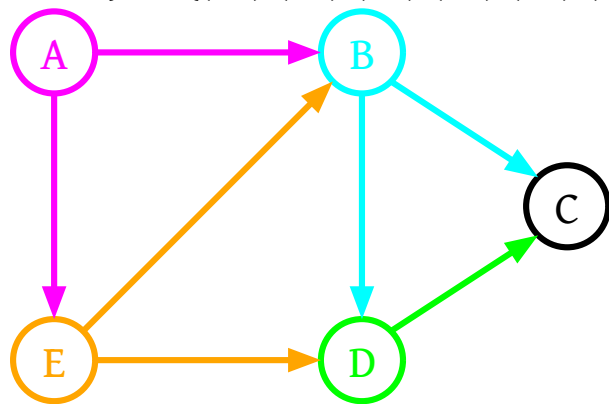
### 1.11.1 Definizione del problema

Si definisce **grafo** una coppia  $G = (V, E)$ , costituita da un insieme  $V$  di **vertici**  $\{v_1, v_2, \dots, v_n\}$  e da un insieme  $E$  di **archi**  $\{e_1, e_2, \dots, e_m\}$ . Il numero di elementi di  $V$  e di  $E$  indicando la **dimensione** del grafo.

Ogni arco  $e_k$  indica l'esistenza di una relazione  $R$  tra i due vertici  $v_i$  e  $v_j$ . Se la relazione  $R$  é simmetrica, ovvero se per qualsiasi coppia  $(v_i, v_j)$  l'esistenza di  $v_i R v_j$  implica l'esistenza di  $v_j R v_i$ , il grafo si dice **non orientato**. Se invece la relazione non é simmetrica, il grafo si dice **orientato**. Un vertice si dice **adiacente** ad un'altro vertice se esiste un arco che ha tali vertici come elementi. Ovvero,  $u$  é adiacente a  $v$  in un grafo  $G = (V, E)$  se  $(u, v) \in E$ . Un arco  $(v_i, v_i)$ , ovvero un arco che unisce un vertice con sé stesso, viene chiamato **cappio**.

Un grafo può venire rappresentato in forma estensionale elencando gli elementi dei due insiemi di cui é costituito. In alternativa, può essere rappresentato graficamente riportando i nodi come cerchi numerati e gli archi come frecce che connettono tali cerchi. Se il grafo é orientato, le frecce hanno una punta, mentre se non é orientato sono piatte.

$V = \{1, 2, 3, 4, 5\}, E = \{(1, 2), (1, 5), (2, 3), (2, 4), (4, 3), (5, 2), (5, 4)\}.$



Esistono principalmente due strutture dati in grado di contenere le informazioni relative ad un grafo, entrambe con punti di forza e punti deboli:

- **Liste di adiacenza:** vettore  $L_v$  di dimensione  $|V|$  dove  $V[i]$  é una lista che contiene tutti i vettori rispetto a cui  $v_i$  é adiacente.

- **Matrice di adiacenza:** matrice  $M_v$  di dimensione  $n \times n$  dove ciascuna cella  $(i, j)$  ha valore 1 se i vertici  $v_i$  e  $v_j$  sono adiacenti, e 0 altrimenti.

Le liste di adiacenza riportano solamente le informazioni relative a quali nodi e quali archi esistono all'interno del grafo, ed infatti il numero di elementi che contiene è  $|V| + |E|$ . D'altro canto, per sapere se una certa coppia di vertici del grafo che rappresenta sono adiacenti è una operazione eseguibile in tempo lineare, perché occorre "scorrere" le liste relative ai due vertici fino a trovare (se esistono) una corrispondenza. Pertanto, le liste di adiacenza sono una codifica vantaggiosa se il numero di archi è, rispetto al numero di vertici, piccolo.

La matrice di adiacenza deve riportare esplicitamente informazioni in merito all'esistenza o alla non esistenza di ogni arco, ed infatti il numero di elementi che contiene è  $|V|^2$ . D'altro canto, sapere se una certa coppia di vertici del grafo che rappresenta sono adiacenti è una operazione eseguibile in tempo costante, perché è sufficiente leggere il valore della cella che ha tale vertici per indici. Pertanto, la matrice di adiacenza è una codifica vantaggiosa se il numero di archi è, rispetto al numero di vertici, grande.

$$V = \{1, 2, 3, 4, 5\}, E = \{(1, 2), (1, 5), (2, 3), (2, 4), (4, 3), (5, 2), (5, 4)\}$$

$A \Rightarrow B \Rightarrow E$   
 $B \Rightarrow C \Rightarrow D$   
 $C$   
 $D \Rightarrow C$   
 $E \Rightarrow B \Rightarrow D$

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

$A \Rightarrow B \Rightarrow E$   
 $B \Rightarrow A \Rightarrow C \Rightarrow D \Rightarrow E$   
 $C \Rightarrow B \Rightarrow D$   
 $D \Rightarrow B \Rightarrow C \Rightarrow E$   
 $E \Rightarrow A \Rightarrow B \Rightarrow D$

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

Dato un grafo  $G = (V, E)$ , prende il nome di **cammino** una qualsiasi sequenza  $P = \langle v_{i1}, v_{i2}, \dots, v_{ik-1}, v_{ik} \rangle$  tale per cui un generico  $v_{ij}$  appartiene a  $V$  per ogni  $1 \leq j \leq k$  e  $(v_{ij}, v_{i,j+1})$  appartiene ad  $E$  per  $1 \leq j < k$ . La **lunghezza** del cammino è  $k-1$ , dove  $k$  è il numero di archi.

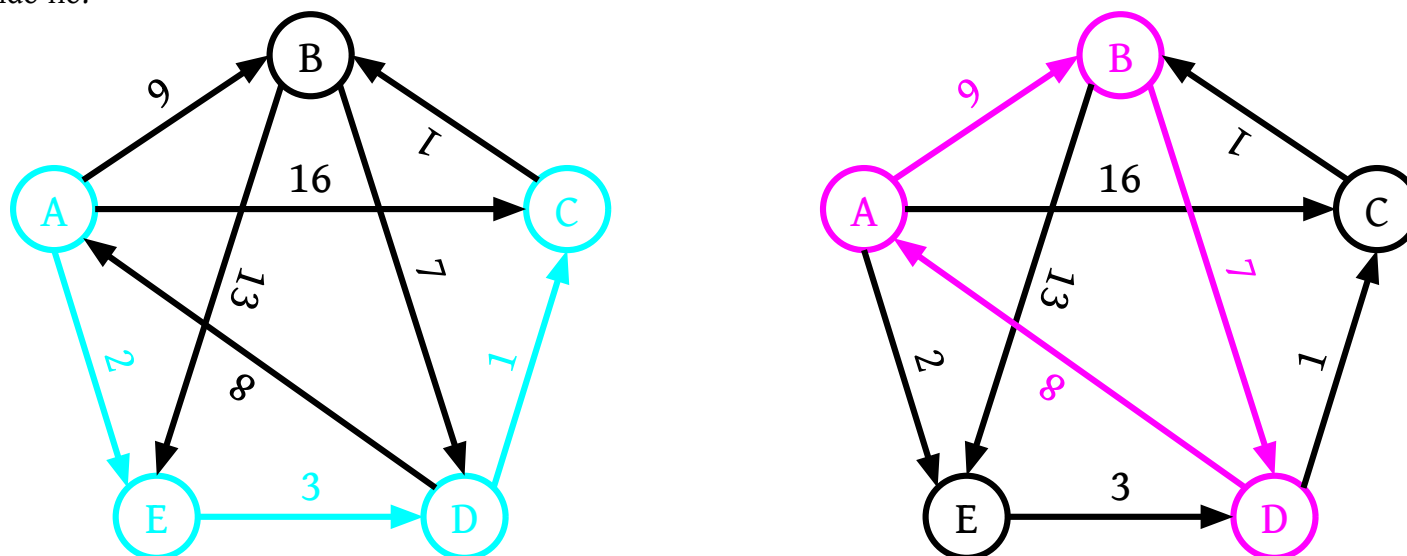
Un cammino  $P$  in cui si ha che  $v_{i1}$  coincide con  $v_{ik}$  viene detto **ciclo**. Un cammino in cui ogni suo vertice compare una sola volta, ovvero che non contiene alcun ciclo, prende il nome di **cammino semplice**. Preso un qualsiasi vertice  $v_{ik} \in P$ , viene detto **predecessore** di  $v_{ik}$  il vertice  $v_{i,k-1}$ . Tutti i vertici in un cammino  $P = \langle v_{i1}, v_{i2}, \dots, v_{ik} \rangle$  diversi da  $v_{i1}$  e da  $v_{ik}$  sono detti **vertici intermedi**.

Un **grafo orientato pesato** è costituito da una tripla  $G = (V, E, W)$  dove, oltre all'insieme  $V = \{v_1, \dots, v_n\}$  di vertici e all'insieme  $E = \{e_1, \dots, e_m\}$  di archi, figura la funzione  $W : V \times V \mapsto \mathbb{R}^+$ . Tale funzione restituisce, per ciascuna coppia di vertici in  $V \times V$ , un numero positivo  $w_{ij}$  chiamato **peso** dell'arco  $(v_i, v_j)$ . Per convenzione, se l'arco  $(v_i, v_j)$  non è presente nel grafo, si ha  $W(v_i, v_j) = \infty$ , mentre se  $v_i = v_j$  si ha  $W(v_i, v_j) = 0$ . Il peso di un cammino è dato dalla somma dei pesi degli archi che costituiscono il cammino.

$$W(v_i, v_j) = \begin{cases} 0 & \text{se } v_i = v_j \\ \infty & \text{se } v_i \neq v_j \wedge (v_i, v_j) \notin E \\ w_{ij} & \text{altrimenti} \end{cases}$$

$$W(\langle v_{i1}, v_{i2}, \dots, v_{ik} \rangle) = \sum_{j=1}^{k-1} W(v_{ij}, v_{i,j+1})$$

Il cammino blu ha peso  $2 + 3 + 1 = 6$ , mentre il cammino rosa ha peso  $8 + 9 + 7 = 24$ . Il primo è un cammino semplice, mentre il secondo no.



Dato un grafo  $G = (V, E, W)$  senza cappi, orientato e pesato, il **problema dei cammini minimi** richiede, per ogni coppia di vertici  $(v_i, v_j) \in V \times V$ , di trovare il cammino che ha inizio in  $v_i$  e fine in  $v_j$  avente il minimo peso. Il problema può essere formulato come un problema di ottimizzazione di minimo:

- Per ciascuna coppia di vertici  $(v_i, v_j)$ , l'insieme delle soluzioni possibili è dato da tutti i possibili cammini che hanno  $v_i$  come primo vertice e  $v_j$  come ultimo;
- La funzione obiettivo è il peso del cammino;
- Il valore ottimo per  $v_i$  e  $v_j$  è il peso del cammino minimo da  $v_i$  a  $v_j$ ;
- La soluzione ottimale è data da uno qualsiasi dei cammini di peso minimo fra i vertici  $v_i$  e  $v_j$ ;

### 1.11.2 Programmazione dinamica: sottostruttura ottima

**Proprietà della sottostruttura ottima per il problema dei cammini minimi.** Sia dato un grafo orientato e pesato  $G = (V, E, W)$ . Sia  $P_{0k}$  un cammino minimo su  $G$  per i vertici  $v_0$  e  $v_k$ . Si considerino due indici  $i, j$  qualsiasi tali per cui  $0 \leq i \leq j \leq k$ : il sottocammino  $P_{ij}$  di  $P_{0k}$  è un cammino minimo per i vertici  $v_i$  e  $v_j$ .

**Dimostrazione.** Il cammino  $P_{0k}$  può essere separato in tre sottocammini: il cammino  $P_{0i}$  da  $v_0$  a  $v_i$ , il cammino  $P_{ij}$  da  $v_i$  a  $v_j$  ed il cammino  $P_{jk}$  da  $v_j$  a  $v_k$ .

$$P_{0k} = \langle v_0, v_1, \dots, v_i, \dots, v_j, \dots, v_k \rangle = P_{0i} \rightsquigarrow P_{ij} \rightsquigarrow P_{jk}$$

$$W(P_{0k}) = W(P_{0i}) + W(P_{ij}) + W(P_{jk})$$

Si supponga per assurdo che esista un cammino  $P'_{ij}$  tale che  $W(P'_{ij}) < W(P_{ij})$ . Allora  $P_{0i} \rightsquigarrow P'_{ij} \rightsquigarrow P_{jk}$  è un cammino da  $v_0$  a  $v_k$  il cui peso è inferiore a  $P_{0k}$ , ma questo va contro l'ipotesi secondo la quale  $P_{0k}$  è un cammino minimo da  $v_0$  a  $v_k$ .

### 1.11.3 Programmazione dinamica: equazione di ricorrenza

Dato un grafo orientato e pesato  $G = (V, E, W)$ , se ne numerino i vertici in maniera univoca. Supponendo che i vertici di  $G$  siano  $V = \{v_1, v_2, \dots, v_n\}$ , se ne consideri un sottoinsieme  $K = \{v_1, v_2, \dots, v_k\}$  per un  $k$  generico. Presi due vertici  $v_i, v_j \in V$ , si indichi con  $P_{ij}^k$  un cammino minimo da  $v_i$  a  $v_j$  i cui vertici intermedi sono stati estratti dall'insieme  $K$ .

Parametrizzando il problema rispetto ai  $k$  vertici intermedi si ottiene un algoritmo di programmazione dinamica chiamato **Algoritmo di Floyd-Warshall**. L'algoritmo può essere descritto in maniera informale come segue:

1. Per tutte le possibili coppie di vertici  $(v_i, v_j)$  si calcoli  $P_{ij}^0$ , il cammino minimo da  $v_i$  a  $v_j$  che non ha alcun vertice intermedio. Tali cammini possono essere costruiti immediatamente a partire dai dati del problema;

2. Si utilizzi tale informazione per calcolare, per tutte le possibili coppie di vertici  $(v_i, v_j)$ ,  $P_{ij}^k$ , il cammino minimo da  $v_i$  a  $v_j$  i cui vertici intermedi sono estratti dall'insieme  $\{v_1, v_2, \dots, v_k\}$ . Si noti come l'insieme dei vertici intermedi di tale cammino minimo non deve necessariamente coincidere con l'intero  $\{v_1, v_2, \dots, v_k\}$ , ma ne é certamente un sottoinsieme;
3. L'algoritmo termina quando viene calcolato, per tutte le possibili coppie di vertici  $(v_i, v_j)$ ,  $P_{ij}^n$ , il cammino minimo da  $v_i$  a  $v_j$  i cui vertici intermedi sono estratti dall'intero insieme  $V$ ;

Il caso base dell'equazione di ricorrenza si ha con  $P_{ij}^0$ , il cammino minimo da  $v_i$  a  $v_j$  che non ha alcun vertice intermedio, l'arco che connette direttamente  $v_i$  e  $v_j$ . Se i due vertici sono coincidenti, ovvero se  $v_i = v_j$ , l'unico cammino possibile é il cammino degenero che va da  $v_i$  a sé stesso. Tale cammino non solo effettivamente non possiede alcun vertice intermedio, rispettando la definizione, ma essendo l'unico cammino possibile é anche certamente quello di peso minimo. Se i due vertici sono distinti ma non é presente nel grafo un arco fra i due, si ha per convenzione  $P_{ij}^0 = \text{NULL}$ .

$$P_{ij}^0 = \begin{cases} \text{NULL} & \text{se } v_i \neq v_j \wedge \langle v_i, v_j \rangle \notin E \\ \langle v_i \rangle & \text{se } v_i = v_j \\ \langle v_i, v_j \rangle & \text{altrimenti} \end{cases} \quad \text{Opt}_0(i, j) = \begin{cases} \infty & \text{se } v_i \neq v_j \wedge \langle v_i, v_j \rangle \notin E \\ 0 & \text{se } v_i = v_j \\ W(v_i, v_j) & \text{altrimenti} \end{cases}$$

Per quanto riguarda la relazione di ricorrenza, si consideri  $P_{ij}^k$ , il cammino minimo da  $v_i$  a  $v_j$  i cui vertici intermedi sono estratti dall'insieme  $\{v_1, v_2, \dots, v_k\}$ . Si indichi con  $\text{Opt}_k(i, j)$  il peso totale del cammino minimo  $P_{ij}^k$ . Possono verificarsi due situazioni:  $v_k$  é oppure non é uno dei vertici intermedi del cammino.

Se  $v_k$  non é uno dei vertici del cammino, allora questo equivale a dire che i vertici intermedi di  $P_{ij}^k$  sono estratti dall'insieme  $\{v_1, v_2, \dots, v_{k-1}\}$ , che é lo stesso insieme da cui sono stati estratti i vertici intermedi di  $P_{ij}^{k-1}$ . Questo significa che un cammino minimo che va da  $i$  a  $j$  i cui vertici intermedi sono stati estratti da  $\{v_1, v_2, \dots, v_{k-1}\}$  é anche un cammino minimo che va da  $i$  a  $j$  i cui vertici intermedi sono stati estratti da  $\{v_1, v_2, \dots, v_k\}$ .

Se invece  $v_k$  é uno dei vertici intermedi di  $P_{ij}^k$ , allora tale cammino può essere certamente diviso in due parti:  $v_i \rightsquigarrow v_k$  e  $v_k \rightsquigarrow v_j$ . Il primo sottocammino é a sua volta un cammino minimo che ha inizio in  $v_i$  e fine in  $v_k$ , mentre il secondo sottocammino é un cammino minimo che ha inizio in  $v_k$  e fine in  $v_j$ .

Dato che entrambi i sottocammini provengono da un cammino i cui vertici intermedi sono stati estratti dall'insieme  $\{v_1, v_2, \dots, v_k\}$ , anche questi avranno i loro vertici intermedi estratti da tale insieme. Dato che però in nessuno dei due sottocammini figura  $v_k$  come vertice intermedio, essendo sempre agli estremi, allora é possibile affermare con certezza che i vertici intermedi di entrambi i sottocammini sono estratti dall'insieme  $\{v_1, v_2, \dots, v_{k-1}\}$ .

$$P_{ij}^k = \begin{cases} P_{ij}^{k-1} & \text{se } v_k \notin P_{ij}^k \\ P_{ik}^{k-1} \rightsquigarrow P_{kj}^{k-1} & \text{se } v_k \in P_{ij}^k \end{cases} \quad \text{Opt}_k(i, j) = \begin{cases} \text{Opt}_{k-1}(i, j) & \text{se } v_k \notin P_{ij}^k \\ \text{Opt}_{k-1}(i, k) + \text{Opt}_{k-1}(k, j) & \text{se } v_k \in P_{ij}^k \end{cases}$$

Naturalmente non é possibile sapere, se non a posteriori, se  $v_k$  fa o non fa parte della  $k$ -esima soluzione ottimale. La scelta migliore fa l'includere o il non includere  $v_k$  nella soluzione dipende da quale rende ottimale, ovvero minimo, il peso del cammino risultante.

$$P_{ij}^k = \min\{P_{ij}^{k-1}, P_{ik}^{k-1} \rightsquigarrow P_{kj}^{k-1}\} \quad \text{Opt}_k(i, j) = \min\{\text{Opt}_{k-1}(i, j), \text{Opt}_{k-1}(i, k) + \text{Opt}_{k-1}(k, j)\}$$

Per la restituzione di una soluzione ottima é utile definire, oltre ai cammini minimi  $P_{ij}^k$ , anche  $\Pi_{ij}^k$ . Questi valori rappresentano il predecessore nel cammino minimo che ha inizio nel vertice  $v_i$  e ha fine nel vertice  $v_j$ , dove i vertici intermedi sono stati estratti dall'insieme  $\{v_1, v_2, \dots, v_k\}$ . Tali valori sono calcolati, come  $P_{ij}^k$ , a partire da un'equazione di ricorrenza.

Il caso base dell'equazione di ricorrenza si ha con  $\pi_{ij}^0$ , il predecessore di  $v_j$  in  $P_{ij}^k$ , ovvero nel cammino minimo da  $v_i$  a  $v_j$  che non ha alcun vertice intermedio. Se i due vertici sono coincidenti, ovvero se  $v_i = v_j$ , il vertice  $v_j$  non ha un predecessore in  $P_{ij}^0$ , perché si sposta da sé stesso in sé stesso. Allo stesso modo, se  $P_{ij}^0$  non esiste, il vertice  $v_j$  non ha un predecessore in tale cammino. Se i due vertici sono distinti e  $P_{ij}^0$  esiste, allora deve aversi che  $\pi_{ij}^0 = v_i$ , perché  $P_{ij}^0$  é costituito da un arco che connette i due vertici direttamente.

$$\pi_{ij}^0 = \begin{cases} \text{NULL} & \text{se } (v_i \neq v_j \wedge \langle v_i, v_j \rangle \notin E) \vee v_i = v_j \\ v_i & \text{altrimenti} \end{cases}$$

Per quanto riguarda il passo ricorsivo, si assuma di avere a disposizione tutti i valori  $\pi_{ij}^{k-1}, \pi_{ij}^{k-2}, \dots$  e di voler calcolare  $\pi_{ij}^k$ .

Se  $\text{opt}^k(i, j) = \text{opt}^{k-1}(i, j)$ , ovvero se il cammino minimo da  $v_i$  a  $v_j$  avente i vertici intermedi estratti dall'insieme  $\{v_1, v_2, \dots, v_{k-1}\}$  ha la stessa lunghezza di quello avente i vertici intermedi estratti dall'insieme  $\{v_1, v_2, \dots, v_k\}$ , allora é possibile inferire che il cammino

minimo  $P_{ij}^k$  ed il cammino minimo  $P_{ij}^{k-1}$  abbiano lo stesso penultimo vertice, ovvero che  $\pi_{ij}^k = \pi_{ij}^{k-1}$ . Questo perché, se i due cammini hanno la stessa lunghezza, allora significa che esiste una soluzione ottimale (non necessariamente l'unica) comune alla  $k$ -esima e alla  $k-1$ -esima istanza del problema.

Altrimenti si ha  $\pi_{ij}^k = \pi_{kj}^{k-1}$ , ovvero il predecessore del vertice  $v_j$  nel cammino minimo avente i vertici intermedi estratti dall'insieme  $\{v_1, v_2, \dots, v_{k-1}\}$  che va da  $v_k$  a  $v_j$ .

$$\pi_{ij}^k = \begin{cases} \pi_{ij}^{k-1} & \text{se } \text{opt}_k(i, j) = \text{opt}_{k-1}(i, j) \\ \pi_{kj}^{k-1} & \text{se } \text{opt}_k(i, j) \neq \text{opt}_{k-1}(i, j) \end{cases}$$

#### 1.11.4 Programmazione dinamica: implementazione bottom-up

L'algoritmo bottom-up viene costruito a partire dall'equazione di ricorrenza sfruttando tante matrici  $d$  quanti sono i vertici del grafo. All'interno di  $d^k[i, j]$ , la cella  $[i, j]$  della  $k$ -esima matrice, viene  $\text{opt}_k(i, j)$ , la lunghezza del cammino minimo fra i vertici  $v_i$  e  $v_j$  i cui vertici intermedi sono stati estratti dall'insieme  $v_1, v_2, \dots, v_k$ . L'algoritmo riceve la matrice  $W$  dei pesi degli archi e restituisce in output la matrice  $d^n$ , quella per la soluzione all' $n$ -esima istanza.

Si noti come la matrice  $d^0$  coincida esattamente con  $W$ , la matrice passata in input, perché entrambe riportano i cammini (unici, se esistono) costituiti da un solo arco <sup>2</sup>.

```
procedure FLOYD-WARSHALL(W)
  n = W.rows
  d0 = W
  for k = 1 to n do
    dk nuova matrice n x n
    for i = 1 to n do
      for j = 1 to n do
        dk[i, j] = min(dk-1[i, j], dk-1[i, k] + dk-1[k, j])
  return dn
```

È facile notare come il tempo di esecuzione dell'algoritmo sia  $O(|V|^3)$ , perché sono presenti tre cicli dove in quello più interno viene eseguita una operazione con costo unitario.

#### 1.11.5 Programmazione dinamica: ricostruzione di una soluzione

Restituire una soluzione al problema dei cammini minimi può venire fatto estendendo l'algoritmo per il calcolo della matrice dei pesi per costruire, allo stesso tempo, la matrice dei predecessori. Sia allora  $\Pi^k$  la matrice che contiene i predecessori di tutti i vertici del grafo dove i vertici dei cammini sono estratti dall'insieme  $\{v_1, v_2, \dots, v_k\}$ .

Ciascuna riga  $i$  della matrice  $\Pi$  definisce un **sottografo dei predecessori**  $G = (V_{i\pi}, E_{i\pi})$ .  $V_{i\pi}$  è definito come l'insieme dei vertici di  $G$  con predecessori diversi da NULL a cui viene aggiunto  $i$ . L'insieme degli archi orientati  $E_{i\pi}$  è l'insieme degli archi indotto dai valori  $\pi$  per i vertici in  $V_{i\pi}$ :

$$V_{i\pi} = \{j \in V \mid \pi_{ij} \neq \text{NULL}\} \cup \{i\}$$

$$E_{i\pi} = \{(\pi_{ij}, j) \in E \mid v \in (V_{i\pi} \setminus \{i\})\}$$

```
procedure FLOYD-WARSHALL(W)
  n = W.rows
  d0 = W
  for i = 1 to n do
    for j = 1 to n do
      if d0[i][j] != infinity and i != j
        pi0 = i
      else
        pi0 = NULL

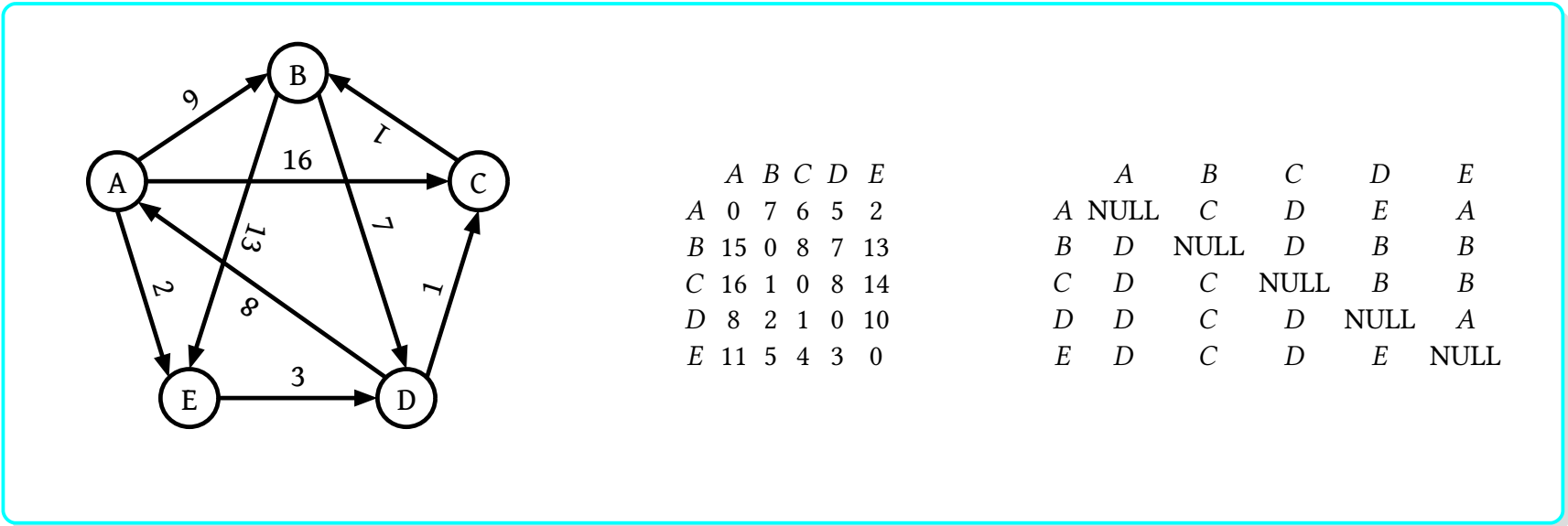
  for k = 1 to n do
    dk nuova matrice n x n
    pik nuova matrice n x n
    for i = 1 to n do
      for j = 1 to n do
        if (dk-1[i, j] <= dk-1[i, k] + dk-1[k, j])
          dk[i, j] = dk-1[i, j]
          pik[i, j] = pik-1[i, j]
        else
```

2. Si noti inoltre come i valori della  $i$ -esima matrice dipendano esclusivamente da quelli della  $i-1$ -esima: questo significa che sarebbe possibile ottimizzare ulteriormente il costo spaziale dell'algoritmo tenendo traccia, in ciascuna  $i$ -esima iterazione, solamente dei valori della  $i-1$ -esima matrice e non di tutte le precedenti.

```

    dk[i, j] = dk-1[i, k] + dk-1[k, j]
    pik[i, j] = pik-1[k, j]
return dn, pin

```



1.11.6 Osservazioni

L'algoritmo Floyd-Warshall può essere utilizzato anche per calcolare la chiusura transitiva di un grafo. La **chiusura transitiva** di un grafo  $G = (V, E)$  è definita come il grafo  $G^* = (V, E^*)$ , dove

$$E^* = \{(i, j): \text{esiste un cammino dal vertice } i \text{ al vertice } j \text{ in } G\}$$

Il metodo consiste nell'assegnare un peso 1 ad ogni arco di  $E$  e nell'eseguire l'algoritmo di Floyd-Warshall. Se esiste un cammino dal vertice  $i$  al vertice  $j$ , si ha  $d_{ij} < n$ , altrimenti  $d_{ij} = \infty$ .

# Capitolo 2

## Tecnica greedy

### 2.1 Introduzione alla tecnica greedy

La **tecnica greedy** é un tipo di approccio alla costruzione di algoritmi che risolvono problemi di ottimizzazione. La tecnica greedy permette di calcolare la soluzione ottima (di una istanza) di un problema attraverso una sequenza di scelte localmente ottime, dove ogni scelta compiuto non dipende da quelle successive. Le differenze con la tecnica di programmazione dinamica sono riassunte di seguito:

Tecnica di programmazione dinamica	Tecnica greedy
Restituisce il valore ottimo, a partire dal quale si ricava una soluzione ottima	Restituisce direttamente la soluzione ottima
Bottom-up	Top-down
Il problema é scomposto in molti sottoproblemi	Il problema é scomposto in pochi sottoproblemi
L'algoritmo é complesso e, in genere, non troppo efficiente	L'algoritmo é semplice e, in genere, efficiente
Permette di risolvere molti problemi	Permette di risolvere pochi problemi

L'approccio greedy alla risoluzione di un problema é possibile solamente se viene dimostrata matematicamente la **proprietá della scelta greedy**, ovvero se la scelta che viene compiuta ad ogni iterazione appartiene ad una soluzione ottima del sottoproblema in esame. A sua volta, affinche tale proprietá sia vera, é necessario dimostrare che il problema possiede la proprietá della sottostruttura ottima (come per la programmazione dinamica).

### 2.2 Interval scheduling

Sia dato un insieme  $A = \{a_1, a_2, ..., a_n\}$  costituito da "attività". Ciascuna attività  $a_i$  é definita come una coppia  $(s_i, f_i)$ , dove i due valori indicano rispettivamente il tempo di inizio dell'attività ed il tempo di fine.

Due attività  $a_i$  e  $a_j$  distinte si dicono *compatibili* se  $a_i$  termina prima che  $a_j$  inizi, ovvero se  $f_i \leq s_j$ . Più in generale, un sottoinsieme  $X$  di  $A$  é detto *compatibile* se é costituito da attività fra di loro mutualmente compatibili, o equivalentemente se non contiene alcuna coppia di elementi di  $A$  fra di loro non compatibili.

$$\text{Comp}(i, j) = \begin{cases} T & \text{se } f_i \leq s_j \\ F & \text{se } f_i > s_j \end{cases}$$

$$\text{Comp}(X) = \begin{cases} T & \text{se } \nexists i, j \in A \text{ t.c. } f_i > s_j \\ F & \text{se } \exists i, j \in A \text{ t.c. } f_i > s_j \end{cases}$$

Il problema **Interval Scheduling** chiede di trovare il sottoinsieme  $X$  di  $A$  di cardinalitá massima composto da attività mutualmente compatibili <sup>1</sup>.

Il problema Interval Scheduling é risolvibile mediante programmazione dinamica. Dato un insieme  $A = \{a_1, a_2, ..., a_n\}$  di attività, si ordinino tali attività per tempo di fine non decrescente. Affinché un algoritmo di programmazione dinamica possa essere applicabile é necessario aggiungere due attività "slack",  $a_0$  e  $a_{n+1}$ , rispettivamente in prima ed in ultima posizione. Sia allora  $A' = A \cup \{a_0, a_{n+1}\}$ . Sia  $X_{i,j}$  il sottoinsieme di  $A'$  composto da tutte le attività che vengono dopo  $a_i$  e prima di  $a_j$ , ovvero  $X = \langle a_{i+1}, a_{i+2}, ..., a_{j-2}, a_{j-1} \rangle$ . La soluzione  $S_{i,j}$  per la  $i, j$ -esima istanza del problema corrisponde a trovare il sottoinsieme di attività mutualmente compatibili di cardinalitá massima rispetto al sottoinsieme  $X_{i,j}$ . La cardinalitá del sottoinsieme di attività mutualmente compatibili di cardinalitá massima per la  $i, j$ -esima istanza del problema viene indicata con  $\text{Opt}(i, j)$ .

La soluzione ottimale per il problema principale é data da  $S_{0,n+1}$ . Infatti, tale insieme é la soluzione per l'insieme che contiene gli elementi di  $A'$  che vengono dopo  $a_0$  e prima di  $a_{n+1}$ , che corrisponde esattamente a  $A = \langle a_1, a_2, ..., a_n \rangle$ .

Il caso base si ha per l'insieme  $X_{i,i+1}$ , ovvero quando  $i + 1 = j$ . Infatti, l'insieme costituito da tutte le attività che vengono dopo  $a_i$  e prima di  $a_{i+1}$  é, per definizione, l'insieme vuoto.

1. Si noti la differenza con il problema Wait Interval Scheduling: tale problema chiede di trovare il sottoinsieme di valore totale massimo, mentre il problema Interval Scheduling chiede di trovare quello di cardinalitá massima.

$$S_{i,i+1} = \emptyset$$

$$\text{Opt}(i, i+1) = |\emptyset| = 0$$

Per quanto riguarda il passo ricorsivo, si consideri una generica soluzione  $S_{i,j}$ . Tale soluzione può essere suddivisa come  $S_{i,k} \cup \{a_k\} \cup S_{k,j}$ , dove  $S_{i,k}$  e  $S_{k,j}$  sono le soluzioni rispettivamente delle istanze  $X_{i,k}$  e  $X_{k,j}$  e dove  $a_k$  è il valore che permette di restituire tali soluzioni ottimali.

$$S_{i,j} = \max\{S_{i,k} \cup \{a_k\} \cup S_{k,j} \mid i < k < j\}$$

$$\text{Opt}(i, j) = \max\{\text{Opt}(i, k) + 1 + \text{Opt}(k, j) \mid i < k < j\}$$

È facile verificare che un algoritmo di programmazione dinamica basato su tale equazione avrebbe un tempo di esecuzione (almeno) quadratico. Questo perché per ottenere la soluzione ottimale per una intera istanza è necessario calcolare le soluzioni ottimali per tutte le istanze  $X_{a,b}$  con  $a, b \in [1, n]$ .

Il problema Interval Scheduling può essere risolto in maniera più semplice e più efficiente applicando la tecnica greedy. Dato un insieme  $A$  di attività, l'algoritmo greedy procede come segue:

```

procedure INTERVAL-SCHEDULING(A)
1  A = A ordinato per tempo di fine non decrescente
2  X = {a1}
3  k = 1

5  for i from 2 to |A| do
6    if si ≥ ek then
7      X = {as} ∪ X
8      k = i

9  return X

```

1. Gli elementi di  $A$  vengono ordinati per tempo di fine non decrescente;
2. Viene creato l'insieme soluzione  $X$ , che inizialmente contiene soltanto  $a_1$ ;
3. Sia  $x'$  l'ultimo elemento dell'insieme  $X$ . Viene aggiunta ad  $X$  l'attività  $a_i \in A$  compatibile con  $x'$  che viene per prima (quella che ha tempo di inizio minore);
4. Se tutte le attività di  $A$  sono state considerate l'algoritmo termina, altrimenti si riprende dal punto precedente.

Tale algoritmo è effettivamente un algoritmo greedy, perché la scelta localmente ottima viene considerata anche globalmente ottima. Tuttavia, affinché questo possa considerarsi corretto, occorre dimostrare la validità della proprietà della scelta greedy, ovvero che l'attività che viene per prima in una certa iterazione (soluzione ottimale locale) è effettivamente membro della soluzione ottimale globale. A dire il vero, dato che ad ogni iterazione l'insieme  $A$  viene ridotto e l'attività che viene considerata è sempre la prima, per dimostrare la validità della proprietà della scelta greedy è sufficiente dimostrare che l'attività  $a_1$ , quella che nell'ordine temporale viene prima di tutte, è sempre parte della soluzione ottima.

**Proprietà della scelta greedy per il problema Interval Schedule.** Sia dato un insieme di attività  $A = \langle a_1, a_2, \dots, a_n \rangle$  ordinate per tempo di fine non decrescente. L'attività  $a_1$  è sempre parte della soluzione ottima.

**Dimostrazione.** Sia  $X$  la soluzione ottima per l'insieme  $A$ . Sia poi  $a'_1$  l'attività in  $X$  avente minor tempo di fine. Naturalmente, se  $a'_1$  coincide con  $a_1$ , allora la dimostrazione è terminata; altrimenti, si sostituisca in  $X$  l'attività  $a'_1$  con  $a_1$ . Il nuovo insieme  $X$  ha mantenuto la stessa cardinalità ed è ancora costituito da attività mutualmente compatibili, pertanto il nuovo  $X$  è un sottoinsieme massimo di attività compatibili di  $A$  che include ora  $a_1$ .

Si noti come l'algoritmo che risolve il problema Interval Scheduling applicando la tecnica greedy è nettamente più veloce di quello che lo risolve applicando la programmazione dinamica. Infatti, indicando con  $n$  la cardinalità di  $A$ , si ha che il loop principale (righe 5-8) esegue esattamente  $|n|$  volte un blocco di codice con tempo di esecuzione unitario, mentre l'ordinamento di  $A$  avviene (assumendo di usare un algoritmo di ordinamento efficiente) in tempo  $O(n \log(n))$ . Pertanto, asintoticamente, il tempo di esecuzione complessivo viene ad essere  $O(n \log(n)) + O(n) = O(n \log(n))$ .

## 2.3 Zaino frazionario

Siano dati  $n$  oggetti  $X = \langle 1, 2, \dots, n \rangle$ , un valore intero  $C$  e due funzioni  $V$  e  $W$ . Tali funzioni associano a ciascun oggetto un numero intero, rispettivamente un valore ed un peso. La definizione di peso e di valore viene poi generalizzata ad un sottoinsieme di oggetti come la somma dei pesi e dei valori di ciascun oggetto di cui è costituito.

$$V : X \mapsto \mathbb{N}$$

$$V(i) = v_i$$

$$W : X \mapsto \mathbb{N}$$

$$W(i) = w_i$$



Il **problema dello Zaino Frazionario** chiede di trovare una sequenza di quantità di ciascun oggetto che permette di massimizzare il valore complessivo di tale sequenza ed al contempo essere inferiore alla capacità dello zaino.

Più formalmente, indicando con  $P = \langle p_1, p_2, \dots, p_n \rangle$  con  $p_i \in [0, 1]$  per ogni  $1 \leq i \leq n$  la sequenza di percentuali con cui ciascun  $i$ -esimo oggetto viene aggiunto ad una certa soluzione, il problema chiede di trovare la sequenza  $P^*$  tale per cui:

$$\sum_{i=1}^n p_i^* w_i \leq C$$

$$P^* = \left\{ P \mid \max \left\{ \sum_{i=1}^n p_i v_i \right\} \right\}$$

Sia dato l'insieme di tre oggetti  $X = \langle (10, 20), (9, 8), (8, 5) \rangle$  ed una capacità  $C = 20$ . Una soluzione ottimale al problema dello Zaino Frazionario è data dalla sequenza  $P^* = \langle 0.35, 1, 1 \rangle$ , avente valore complessivo  $V(P^*) = 0.35 \cdot 10 + 1 \cdot 9 + 1 \cdot 8 = 20.5$  e peso complessivo  $W(P^*) = 0.35 \cdot 20 + 1 \cdot 8 + 1 \cdot 5 = 20$ .

Il problema può essere risolto applicando la tecnica greedy. Innanzitutto, dato un certo insieme di oggetti  $X = \langle 1, 2, \dots, n \rangle$ , per ciascun  $i$ -esimo oggetto sia  $s_i = v_i / w_i$  il *valore specifico*<sup>2</sup> di tale oggetto, ovvero il valore che ha una unità di peso dell' $i$ -esimo oggetto.

Si ordini l'insieme  $X$  in ordine non crescente rispetto ai valori specifici degli oggetti di cui è costituito. Ad ogni iterazione dell'algoritmo greedy viene selezionata la quantità dell' $i$ -esimo oggetto che viene scelta (non superiore ad 1) e la capacità dello zaino viene ridotta di tanto quanto è il peso di tale quantità. L'algoritmo prevede di prendere, per ciascun oggetto, la massima quantità compatibile con la capacità residua dello zaino, fermandosi quando è stata raggiunta la capacità massima.

```

procedure FRACTIONAL-KNAPSACK(n, V, W, C)
1   for i = 1 to n do
2       S[i] = V[i] / W[i]

3   S = S ordinato per valore specifico non crescente
4   i = 1
5   c = C

6   while i ≤ n and c > 0 do
7       P[i] = min{c / W[i], 1}
8       c = c - P[i] * W[i]
9       i = i + 1

10  return P

```

Tale algoritmo è effettivamente un algoritmo greedy, perché la scelta localmente ottima viene considerata anche globalmente ottima. Tuttavia, affinché questo possa considerarsi corretto, occorre dimostrare la validità della proprietà della scelta greedy, ovvero che la percentuale di oggetto che viene scelta per ciascuna iterazione (soluzione ottimale locale) è effettivamente membro della soluzione ottimale globale. A dire il vero, dato che ad ogni iterazione viene sempre considerato il primo oggetto e la capacità dello zaino diminuisce sempre di una quantità relativa al peso di quest'ultimo, per dimostrare la validità della proprietà della scelta greedy è sufficiente dimostrare che la massima quantità compatibile con l'intera capacità  $C$  del primo oggetto, quello che ha il valore specifico maggiore di tutti, è sempre parte della soluzione ottima.

**Proprietà della scelta greedy per il problema dello Zaino Frazionario.** Sia dato un insieme di oggetti  $X = \langle 1, 2, \dots, n \rangle$  ordinati per valore specifico non crescente ed una capacità  $C$ . La massima percentuale del primo oggetto compatibile  $C$  che può essere aggiunta alla soluzione ottima è sempre parte di quest'ultima.

**Dimostrazione.** La percentuale che viene scelta per il primo oggetto è  $p_1 = \min(C / w_1, 1)$ , la massima possibile. Sia  $P$  una soluzione ottima: dato che l'algoritmo sceglie sempre la massima percentuale possibile, la percentuale  $p'_1$  che si trova in  $P$  può essere esclusivamente uguale oppure inferiore a  $p_1$ .

Se  $p'_1$  e  $p_1$  sono uguali, la dimostrazione è terminata. Se  $p'_1$  è inferiore a  $p_1$ , è possibile costruire una nuova soluzione in cui la percentuale del primo oggetto è uguale a  $p_1$ . La proporzione di uno o più oggetti aggiunti successivamente sarà ridotta di conseguenza. Dato che il valore specifico del primo oggetto è superiore o uguale a quello dei successivi non è possibile ottenere una soluzione avente valore inferiore. Si conclude quindi che  $p_1$  è sempre parte di una soluzione ottima.

2. Il nome "valore specifico" è dato in analogia con il concetto di **peso specifico**, ovvero il peso di un metro cubo di una certa sostanza.

## 2.4 Matroidi e sistemi di indipendenza

Siano dati un insieme finito  $S$  ed un insieme non vuoto  $F$ , sottoinsieme dell'insieme potenza di  $S$ . La coppia  $(S, F)$  é detta **sistema di indipendenza** se, per ciascun elemento di  $F$ , anche tutti i sottoinsiemi di tale elemento appartengono ad  $F$ .

$$(S, F \subseteq \text{Pow}(S)) \text{ con } A \in F \wedge B \subset A \Rightarrow B \in F \quad \forall A \in F, \forall B \subset A$$

Ciascun elemento dell'insieme  $F$  é detto **sottoinsieme indipendente** di  $S$ . Si noti come, affinché una coppia  $(S, F)$  possa essere un sistema di indipendenza, l'insieme  $F$  deve almeno contenere l'insieme vuoto, perché l'insieme vuoto é sottoinsieme di ogni insieme.

Si considerino gli insiemi  $S$  e  $F$  così definiti:

$$\begin{cases} S = \{1, 2, 3\} \\ F \subset \text{Pow}(S) = \{\emptyset, \{1\}, \{3\}, \{1, 3\}\} \end{cases} \quad \text{Pow}(S) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

La coppia  $(S, F)$  costituisce un sistema di indipendenza. Infatti, tutti i sottoinsiemi dell'insieme vuoto sono in  $F$  (sé stesso), tutti i sottoinsiemi dell'insieme  $\{1\}$  sono in  $F$  (l'insieme vuoto), tutti i sottoinsiemi dell'insieme  $\{3\}$  sono in  $F$  (l'insieme vuoto) e tutti i sottoinsiemi dell'insieme  $\{1, 3\}$  sono in  $F$  ( $\{1\}$ ,  $\{3\}$  e l'insieme vuoto).

Un sistema di indipendenza  $(S, F)$  é detto **matroide** se possiede la **proprietá di scambio**, ovvero per qualsiasi coppia di sottoinsiemi di  $F$ , dove il primo ha un elemento in piú del secondo, esiste almeno un elemento del secondo insieme non presente nel primo tale che aggiunto al primo insieme restituisce un insieme che appartiene ancora ad  $F$ .

$$(S, F \subseteq \text{Pow}(S)) \text{ con } \forall A, B \in F \text{ t.c. } |B| = |A| + 1 \exists b \in B - A \text{ t.c. } \{b\} \cup A \in F$$

Siano  $S$  un insieme finito generico e  $F$  il sottoinsieme dell'insieme potenza di  $S$  che contiene tutti gli elementi di  $\text{Pow}(S)$  con cardinalitá inferiore o uguale a  $k$ , con  $k$  numero naturale fissato.

La coppia  $(S, F)$  costituisce un sistema di indipendenza perché, preso un qualsiasi insieme  $f \in F$ , tutti i suoi sottoinsiemi hanno cardinalitá inferiore a  $f$ , che a sua volta ha cardinalitá inferiore o uguale a  $k$ . Pertanto, anche tutti i sottoinsiemi di  $f$  fanno parte di  $F$ .

Inoltre, la coppia  $(S, F)$  costituisce un matroide. Siano infatti due insiemi qualsiasi  $A, B \in F$  tale che  $|B| = |A| + 1$ . Avendo  $A$  un elemento in meno di  $B$ , si ha  $|A| < k$ . Aggiungendo un qualsiasi elemento  $b \in B - A$  all'insieme  $A$  si ottiene l'insieme  $\{b\} \cup A$  che ha cardinalitá  $|A| + 1 = |B|$ , che per ipotesi é un valore inferiore o uguale a  $k$ , e che quindi appartiene ad  $F$ .

Sia  $M = (S, F)$  un matroide. Un elemento  $s \in S$  é detto **estensione** dell'insieme  $A \in F$  se  $s \notin A$  e  $A \cup \{s\} \in F$ , ovvero aggiungendo  $s$  ad  $A$  si ottiene ancora un sottoinsieme indipendente di  $S$ . Se non esiste alcun elemento  $s$  che possa essere aggiunto ad  $A$  tale che  $A \cup \{s\} \in F$ , si dice che  $A$  é **massimale**.

Tutti i sottoinsiemi indipendenti massimali in un matroide hanno la stessa cardinalitá.

**Dimostrazione.** Sia  $M = (S, F)$  un matroide. Si supponga per assurdo che vi siano (almeno) due sottoinsiemi indipendenti massimali in  $M$ ,  $A$  e  $B$ , tali che  $|A| < |B|$ .

Essendo  $M$  un matroide ed essendo  $|A| \neq |B|$ , deve essere a questi applicabile la proprietá di scambio. Ovvero, deve esistere un elemento  $b \in B - A$  tale che  $\{b\} \cup A \in F$ , tuttavia questo contraddice l'ipotesi che  $A$  sia massimale e che quindi non possa esistere alcun  $b$ . Si deduce quindi che due (o piú) sottoinsiemi  $A$  e  $B$  indipendenti massimali in uno stesso matroide aventi cardinalitá diversa non possano esistere.

Sia  $(S, F)$  un sistema di indipendenza e sia  $C$  un sottoinsieme di  $S$ . La coppia  $(C, F_C)$  con  $F_C = \{A \in F \text{ t.c. } A \subseteq C\}$  é il **sottosistema di indipendenza indotto** da  $C$ .

Un qualsiasi sottosistema indotto  $(C, F_C)$  di un matroide  $(S, F)$  é esso stesso un matroide se i suoi insiemi massimali hanno la stessa cardinalità.

Un sistema di indipendenza  $(S, F)$  é (anche) un matroide se e solo se per ogni  $C \subseteq S$  gli insiemi massimali di  $(C, F_C)$  hanno la stessa cardinalità.

Dato un grafo  $G = (V, E)$  non orientato e connesso, si dice **matroide grafico** la coppia di insiemi  $M_G = (S, F)$ , dove  $S$  coincide con l'insieme degli archi  $E$  e  $F$  é costituito da tutti i sottoinsiemi di  $S$  che non presentano cicli, ovvero da tutti ed i soli sottoinsiemi di  $S$  che definiscono dei sottografi che formano una **foresta**.

Un matroide grafico é un matroide.

**Dimostrazione.** Dato un grafo non orientato e connesso  $G = (V, E)$ , sia  $M_G = (S, F)$  il matroide grafico a questo associato. Si dimostri innanzitutto che  $M_G$  é un sistema di indipendenza: preso un qualsiasi  $A \in F$ , sia  $B$  un suo sottoinsieme. Tale sottoinsieme viene costruito eliminando da  $A$  uno o piú archi; essendo però  $A$  privo di cicli per ipotesi, anche  $B$  dovrà esserlo, e quindi anche  $B$  é un elemento di  $F$ . Questo significa che qualsiasi sottoinsieme di un elemento di  $F$  fa a sua volta parte di  $F$ , pertanto  $M_G$  é un sistema di indipendenza.

Siano poi  $A, B \in F$  due sottoinsiemi tali per cui  $|B| = |A| + 1$ , i quali definiscono rispettivamente i sottografi  $G_A = (V, A)$  e  $G_B = (V, B)$ . Per definizione tali sottografi sono delle foreste, costituite rispettivamente da  $|V| - |A|$  e da  $|V| - |B|$  alberi. La foresta  $G_B$  contiene esattamente un albero in meno della foresta  $G_A$ , infatti:

$$|V| - |A| - (|V| - |B|) = |V| - |A| - |V| + |B| = -|A| + |A| + 1 = 1$$

Questo implica che, essendo  $G_A$  e  $G_B$  costituite dagli stessi vertici, esiste in  $G_B$  un arco  $(u, v)$  che connette due vertici  $u$  e  $v$  che in  $G_B$  si trovano in un solo albero ma che in  $G_A$  si trovano in due alberi diversi. Allora tale arco può essere aggiunto a  $G_A$  senza creare un ciclo, ottenendo quindi un nuovo sottografo i cui vertici formano ancora un insieme che appartiene ad  $F$ . Questa é però precisamente la definizione di proprietà di scambio, e quindi é possibile affermare che  $M_G$  sia un matroide.



Si consideri il grafo  $G = (V, E)$  sopra presentato. A partire dall'insieme  $F$ , che contiene tutti i sottoinsiemi di  $E$  che non hanno cicli. Da  $F$  é estratto il sottoinsieme di archi  $A = \{(a, b), (b, c), (c, d), (d, e), (i, g), (f, g), (g, h)\}$ . Il sottografo  $G_A = (V, A)$  definisce una foresta di  $|V| - |A| = 2$  alberi, il primo formato dagli archi  $\{(a, b), (b, c), (c, d), (d, e)\}$  ed il secondo dagli archi  $\{(i, g), (f, g), (g, h)\}$

Dato il matroide grafico  $M_G = (S, F)$ , l'insieme  $A \in F$  é massimale se non é possibile aggiungere un arco ad  $A$  (che non faccia già parte di  $A$ ) senza formare un ciclo. Questo equivale a dire che il sottografo  $G_A$  indotto da  $A$  é una foresta formata da un solo albero, avente  $|V| - 1$  archi che connette tutti i vertici del grafo. Un albero con queste caratteristiche viene chiamato **albero di copertura**, o **Spanning Tree (ST)**.



L'insieme di archi sopra evidenziato é uno spanning tree per il grafo dell'esempio precedente.

Un **matroide pesato** é un matroide  $M = (S, F)$  a cui viene aggiunta una funzione peso  $W$ , la quale associa ad ogni elemento dell'insieme  $S$  un numero reale strettamente positivo. Tale funzione viene estesa ad un generico insieme  $A$  membro di  $F$  come la somma dei pesi di tutti gli elementi contenuti in  $A$ .

$$W : S \mapsto \mathbb{R}^+$$

$$W(s) = w_s \quad s \in S$$

$$W(A) = \sum_{a \in A} w_a \quad A \in F$$

## 2.5 Algoritmo Greedy Standard

Il legame fra matroidi e problemi risolvibili mediante tecnica greedy é molto forte. Infatti, molti problemi per i quali un metodo greedy fornisce soluzioni ottime possono essere riformulati come la ricerca di un sottoinsieme indipendente di peso massimo in un matroide pesato. Tale sottoinsieme, che é indipendente ed ha assegnato il peso piú grande possibile, é detto **sottoinsieme ottimo** del matroide. Poiché la funzione peso é stata definita per restituire solo valori strettamente positivi, un sottoinsieme ottimo é sempre un sottoinsieme indipendente massimale.

L'algoritmo che permette di ricavare il sottoinsieme ottimo di un matroide pesato, presentato di seguito, prende il nome di **algoritmo greedy standard**. Tale algoritmo riceve in input un matroide pesato  $M = (S, F, W)$  e restituisce un sottoinsieme ottimo  $A$ . L'algoritmo é effettivamente un algoritmo greedy perché esamina uno dopo l'altro in ordine di peso ogni elemento  $x \in S$  e lo aggiunge immediatamente all'insieme  $A$  se  $A \cup \{x\}$  é esso stesso indipendente.

```

Procedure GREEDY(S, F, W)
1   A ← ∅
2   S ← S ordinato per peso W decrescente

3   for i from 1 to |S| do
4       if ({si} ∪ A) ∈ F then
5           A ← {si} ∪ A

6   return A

```

L'algoritmo greedy standard restituisce sempre un sottoinsieme indipendente.

**Dimostrazione.** Si osservi lo pseudocodice dell'algoritmo greedy standard, presentato sopra. La riga 4 controlla se, aggiungendo un elemento  $s_i$  ad  $A$ , l'insieme  $A$  resta indipendente. Se  $A$  resta indipendente, la riga 5 aggiunge  $s_i$  ad  $A$ , altrimenti viene scartato. Poiché l'insieme vuoto é per definizione un sottoinsieme indipendente e poiché ogni esecuzione del corpo del ciclo for lascia  $A$  indipendente, il sottoinsieme  $A$  é indipendente per induzione. Di conseguenza, l'algoritmo greedy standard restituisce sempre un sottoinsieme indipendente.

Il tempo di esecuzione dell'algoritmo può essere calcolato osservando che questo può essere diviso in tre parti: l'ordinamento dell'insieme  $S$  (riga 2), il ciclo for (riga 3) e la condizione di appartenenza dell'elemento in esame alla soluzione del problema (righe 4 e 5). La prima ha tempo di esecuzione pari a quella di un normale ordinamento, ovvero  $O(n \log(n))$ , mentre la seconda ha tempo di esecuzione proporzionale alla dimensione dell'input, ovvero  $O(n)$ . La terza non ha un tempo di esecuzione predefinito, ma varia da problema a problema, e viene indicata con un generico  $O(f(n))$ .

La condizione di appartenenza viene esaminata  $n$  volte, pertanto il tempo di esecuzione complessivo della seconda e della terza parte è  $O(n) \times O(f(n)) = O(nf(n))$ . Il tempo di esecuzione complessivo si ottiene sommando questa quantità al tempo di esecuzione necessario per l'ordinamento, ovvero  $O(n \log(n)) + O(nf(n))$ .

**Teorema di Rado (I).** Un sistema di indipendenza è (anche) un matroide se, per qualsiasi funzione peso, l'algoritmo greedy standard fornisce il sottoinsieme ottimo per tale sistema.

**Dimostrazione.** Siano  $(S, F)$  un matroide e  $W$  una qualsiasi funzione peso. Sia poi  $X = \{s_1, s_2, \dots, s_p\}$  la soluzione fornita dall'algoritmo greedy standard per tale matroide, con gli elementi disposti in ordine decrescente di peso, ovvero  $W(s_1) \geq W(s_2) \geq \dots \geq W(s_p)$ . Naturalmente, essendo il risultato dell'applicazione dell'algoritmo greedy standard, l'insieme  $X$  è massimale per  $(S, F)$ .

Sia allora  $X' = \{s'_1, s'_2, \dots, s'_p\}$  un altro insieme massimale, distinto da  $X$ , avente anch'esso gli elementi ordinati in ordine decrescente di peso. Avendo  $X$  e  $X'$  la stessa cardinalità, è possibile mettere in corrispondenza biunivoca l' $i$ -esimo elemento di  $X$  con l' $i$ -esimo elemento di  $X'$ . Il teorema è dimostrato se, per un qualsiasi insieme  $X'$  massimale per  $(S, F)$ , il peso  $W(X)$  è almeno pari al peso  $W(X')$ .

Se ciascun elemento di  $X$  ha peso maggiore del corrispondente elemento di  $X'$  allora certamente  $W(X) \geq W(X')$ . Ma questo significa che  $X$ , l'insieme restituito dall'algoritmo greedy standard, ha peso almeno pari a quello di un generico insieme massimale  $X'$ , ed il teorema è dimostrato.

Si consideri il caso più generico in cui non tutti gli elementi di  $X$  hanno peso maggiore dei loro corrispondenti in  $X'$ . Ricordando che gli elementi di  $X$  e di  $X'$  sono disposti in ordine di peso, deve esistere almeno un elemento con indice  $k$  tale per cui tutti gli elementi di  $X$  con indice inferiore a  $k$  hanno peso maggiore dei rispettivi elementi di  $X'$  mentre il  $k$ -esimo elemento di  $X$  ha peso inferiore al  $k$ -esimo elemento di  $X'$ .

In particolare, deve certamente aversi  $W(s_k) \leq W(s_{k-1})$  e  $W(s'_k) \leq W(s'_{k-1})$ . Ricordando però che vale  $W(s_k) < W(s'_k)$ , si ha allora  $W(s_k) < W(s'_k) \leq W(s'_{k-1}) \leq W(s_{k-1})$ . Si consideri allora  $C = \{s \in S \text{ t.c. } W(s) \geq W(s'_k)\}$ , il sottoinsieme di  $S$  che contiene tutti gli elementi con peso maggiore o uguale al peso di  $s'_k$ . La coppia  $(C, F_C)$  forma un sottosistema indotto per  $C$ . L'elemento  $s'_k$  appartiene certamente a  $C$ , perché evidentemente  $W(s'_k) \geq W(s'_k)$ . Ma allora  $X' \cap C = \{s'_1, s'_2, \dots, s'_{k-1}, s'_k\}$ . D'altro canto, l'elemento  $s_k$  non può appartenere a  $C$ , perché per ipotesi è stato assunto che  $W(s_k) < W(s'_k)$ . Ma allora  $X \cap C = \{s_1, s_2, \dots, s_{k-1}\}$ . Questo significa che  $X' \cap C$  ha un elemento in più di  $X \cap C$ , e quindi  $|X' \cap C| > |X \cap C|$ . Questo contraddice però l'ipotesi che  $X$  e  $X'$  siano insiemi massimali per  $(S, F)$ , ed implica che non possa esistere un  $k$  tale per cui  $W(s_k) < W(s'_k)$ .

Occorre allora concludere che ciascun elemento di  $X$  deve per forza avere peso maggiore o uguale al peso del rispettivo elemento di  $X'$ . Ma questa situazione coincide con la situazione precedente, ed il teorema è dimostrato.

**Teorema di Rado (II).** Se l'algoritmo greedy standard fornisce, per qualsiasi funzione peso, il sottoinsieme ottimo per un sistema di indipendenza, allora tale sistema di indipendenza è (anche) un matroide.

**Dimostrazione.** Sia  $(S, F)$  un sistema di indipendenza che non è un matroide. Allora esiste un insieme  $C \subseteq S$  e due insiemi massimali  $A, B \in (C, F_C)$  con cardinalità diversa. Sia  $|B| = p$  e  $|A| > |B|$ . La seconda parte del teorema è dimostrata se è possibile costruire almeno una funzione peso per la quale l'algoritmo greedy standard applicato a  $(S, F)$  restituisce un insieme non ottimo. Si consideri a tal proposito la funzione peso  $W$  così costruita:

$$W(s) = \begin{cases} p+2 & \text{se } s \in B \\ p+1 & \text{se } s \in (A-B) \\ 1 & \text{altrimenti} \end{cases}$$

L'algoritmo greedy standard restituirebbe l'insieme  $B$ , perché la funzione peso attribuisce peso maggiore agli elementi di  $B$ .

L'insieme  $A$  ha almeno  $p+1$  elementi, perché per ipotesi  $|A| > |B|$  e  $|B| = p$ . Inoltre, ciascun elemento presente in  $A$  ma non in  $B$  (almeno uno certamente esiste) ha peso  $p+1$ . Pertanto, il peso complessivo di  $A$  è certamente almeno pari a  $(p+1)(p+1)$ .

L'insieme  $B$  ha, per ipotesi,  $p$  elementi ciascuno di peso  $p+2$ , pertanto il peso complessivo di  $B$  è  $p+2$ . Si noti però come:

$$W(A) \geq (p+1)(p+1) = (p+1)^2 \geq p^2 + 2p + 1 > p^2 + 2p = p(p+2) = W(B)$$

Ovvero,  $W(A) > W(B)$ . Questo significa che, per questa funzione peso, l'algoritmo greedy standard non restituisce l'insieme di peso maggiore per il sistema di indipendenza  $(S, F)$ , ed il teorema è provato.

Il Teorema di Rado implica che se un problema può essere formulato come un matroide pesato allora esiste un algoritmo greedy che lo risolve per qualsiasi funzione peso. Se invece non è possibile, potrebbe comunque esistere un algoritmo greedy che risolve il problema, ma solo per alcune specifiche funzioni peso, non tutte.

Sia  $V = v_1, v_2, \dots, v_n$  un insieme di vettori  $m$ -dimensionali. Ciascun vettore  $v_i$  è costituito da  $m-1$  componenti e da un peso, assegnato da una funzione  $W$ . Si richiede di determinare il sottoinsieme di vettori linearmente indipendenti che ha peso massimo, applicando l'algoritmo greedy standard.

Sia  $S$  l'insieme finito di vettori, e sia poi  $F$  la famiglia dei sottoinsiemi di  $S$  composti da vettori linearmente indipendenti. Affinché l'algoritmo greedy standard sia applicabile, la tripla  $(S, F, W)$  deve essere codificabile come un matroide, che a sua volta richiede di dimostrare che la tripla è un sistema di indipendenza.

Sia  $A \in F$  un qualsiasi insieme di vettori linearmente indipendenti estratti da  $F$ . Se da  $A$  vengono eliminati uno o più vettori, si ottiene un insieme  $A'$  ancora costituito da vettori linearmente indipendenti, pertanto anche  $A'$  appartiene ad  $F$ . Pertanto,  $(S, F, W)$  è un sistema di indipendenza.

Siano  $A, B \in F$  due insiemi di vettori linearmente indipendenti estratti da  $F$ , dove  $|A| = |B| + 1$ . È sempre possibile scegliere un elemento  $b_i \in B$  tale per cui  $A \cup \{b_i\} \in F$ , perché se non esistesse tale vettore allora l'insieme  $B$  non sarebbe costituito da vettori linearmente indipendenti.

È quindi possibile risolvere il problema utilizzando l'algoritmo greedy standard; è sufficiente utilizzare l'appartenenza ad un insieme di vettori linearmente indipendenti come condizione di appartenenza.

```
Procedura greedy_ind_vectors( $V = \{v_1, \dots, v_n\}$ )
   $X \leftarrow \emptyset$ 
   $V$  viene ordinato per peso  $W$  decrescente

  for  $i$  from 1 to  $n$  do
    if  $(\{v_i\} \cup X)$  contiene vettori linearmente indipendenti then
       $X \leftarrow \{v_i\} \cup X$ 

  return  $X$ 
```

Nonostante l'algoritmo greedy standard restituisca sempre il sottoinsieme ottimo di peso massimo, è comunque possibile utilizzarlo per risolvere problemi di ottimizzazione di minimo (ammesso che questo sia risolvibile mediante tecnica greedy). Una volta convertito il problema in esame in un matroide pesato  $M = (S, F, W)$  equivalente, viene selezionato il peso  $w_0 \in W$  di valore massimo. A questo punto la funzione peso  $W$  viene sostituita con un'altra funzione  $W'$ , il cui elemento  $i$ -esimo  $w'_i$  è dato dalla differenza fra  $w_0$  e  $w_i$ , l' $i$ -esimo elemento di  $W$ .

In questo modo, ordinare gli elementi di  $S$  rispetto a  $W'$  in ordine decrescente equivale ad ordinarlo rispetto a  $W$  in ordine crescente, e di conseguenza restituire il sottoinsieme ottimo massimale rispetto a  $W'$  equivale a restituire l'insieme ottimo minimale rispetto a  $W$ .

## 2.6 Strutture dati per insiemi disgiunti

Una **struttura dati per insiemi disgiunti** mantiene una collezione  $\{S_1, S_2, \dots, S_k\}$  di insiemi dinamici disgiunti, ovvero insiemi di cardinalità non fissata i cui elementi non possono trovarsi in più di un insieme contemporaneamente. Ciascun insieme è identificato da un **rappresentante**, che è un elemento dell'insieme. Quale elemento questo debba essere è lasciato all'implementazione: può sia essere un elemento qualunque, sia un elemento che possiede una certa caratteristica all'interno dell'insieme (l'elemento più piccolo, l'elemento che ha associato il valore maggiore, ecc...).

Indicando con  $x$  un generico elemento di uno degli insiemi, si vuole costruire una struttura dati che supporti le seguenti tre operazioni:

- **MAKE-SET( $x$ )** crea un nuovo insieme il cui unico elemento e rappresentante è  $x$ . Poiché gli insiemi sono disgiunti,  $x$  non può trovarsi in un'altro insieme.
- **UNION( $x, y$ )** unisce gli insiemi (univoci) che contengono gli elementi  $x$  e  $y$ , siano questi  $S_x$  e  $S_y$ , in un unico insieme, che è l'unione di questi due, che viene poi aggiunto alla collezione. Naturalmente, si assume che  $S_x$  e  $S_y$  siano disgiunti prima di applicare l'operazione. Il rappresentante di  $S_x \cup S_y$  diviene uno qualsiasi degli elementi di  $S_x$  o di  $S_y$ <sup>3</sup>. Al fine di mantenere la

3. In alcune implementazioni si richiede che il rappresentante dell'insieme unione sia sempre un elemento del primo insieme o sempre un elemento del secondo insieme.

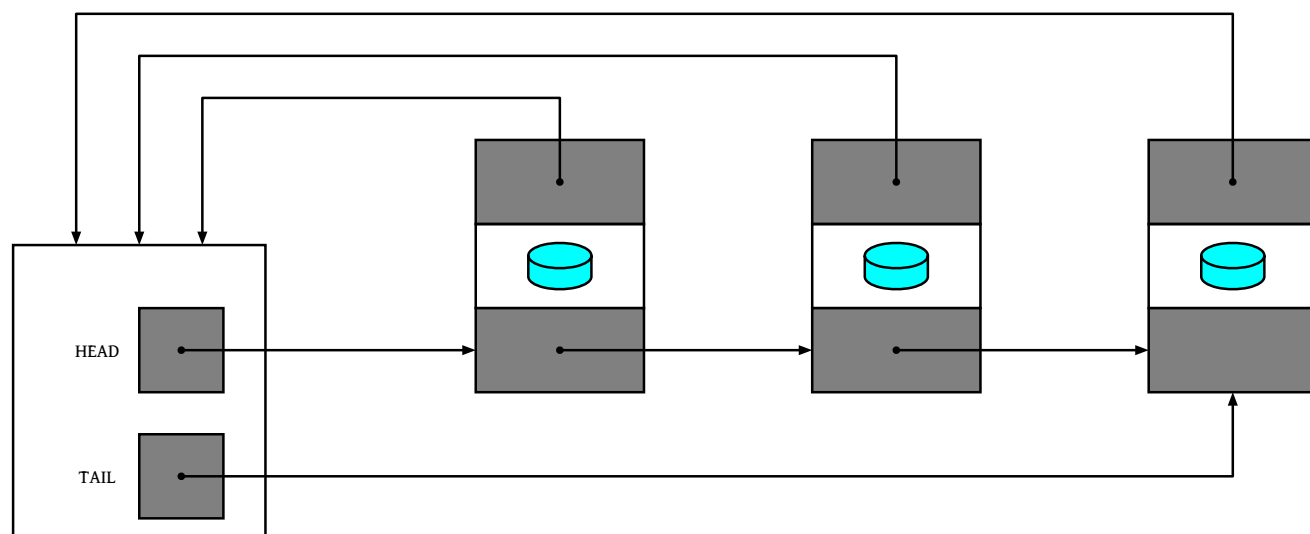


propriet  di disgiunzione degli insiemi della collezione, occorre eliminare dalla collezione  $S_x$  e  $S_y$ .

- $\text{FIND-SET}(x)$  restituisce un puntatore al rappresentante dell'insieme (unico) che contiene  $x$ .

Una prima implementazione di una struttura dati per insiemi disgiunti prevede di rappresentare ciascun insieme mediante una lista concatenata. L'oggetto di ciascun insieme ha gli attributi `head`, che punta al primo oggetto della lista, e `tail`, che punta all'ultimo oggetto. Ogni oggetto nella lista contiene un elemento dell'insieme, un puntatore al successivo oggetto della lista e un puntatore che ritorna all'oggetto dell'insieme. All'interno di ciascuna lista concatenata, gli oggetti possono apparire in qualsiasi ordine. Il rappresentante   l'elemento dell'insieme nel primo oggetto della lista.

Con questa rappresentazione, entrambe le operazioni  $\text{MAKE-SET}$  e  $\text{FIND-SET}$  possono essere implementate per avere tempo di esecuzione unitario. Per implementare  $\text{MAKE-SET}$    sufficiente creare una nuova lista concatenata il cui unico oggetto    $x$ . Per implementare  $\text{FIND-SET}$  occorre seguire il puntatore da  $x$  per arrivare all'oggetto del suo insieme e poi ritornare all'elemento nell'oggetto cui punta `head`.



La pi  semplice implementazione dell'operazione  $\text{UNION}(x, y)$  richiede un tempo di esecuzione nettamente superiore a quello di  $\text{MAKE-SET}(x)$  e di  $\text{FIND-SET}(x)$ . Per eseguire  $\text{UNION}(x, y)$  viene aggiunta la lista di  $y$  alla fine della lista di  $x$ , ed il rappresentante di  $x$  diviene rappresentante dell'insieme risultante. Questo richiede di aggiornare il puntatore all'oggetto dell'insieme per ogni oggetto che si trovava in  $y$ , e questa operazione richiede tempo lineare nella lunghezza della lista di  $y$ .

Una delle tante applicazioni delle strutture dati per insiemi disgiunti consiste nel determinare le componenti connesse di un grafo non orientato. La procedura  $\text{CONNECTED-COMPONENTS}$  usa le operazioni degli insiemi disgiunti per calcolare le componenti connesse di un grafo. Una volta che  $\text{CONNECTED-COMPONENTS}$  ha preprocessato il grafo, la procedura  $\text{SAME-COMPONENT}$    in grado di determinare se due vertici sono nella stessa componente connessa.

```

procedure CONNECTED-COMPONENTS(V, E)
1  foreach v ∈ V
2    MAKE-SET(v)
3  foreach (u, v) ∈ E
4    if (FIND-SET(u) ≠ FIND-SET(v)) then
5      UNION(u, v)

```

```

procedure SAME-COMPONENT(u, v)
6  if (FIND-SET(u) == FIND-SET(v))
7    return TRUE
8  else
9    return FALSE

```

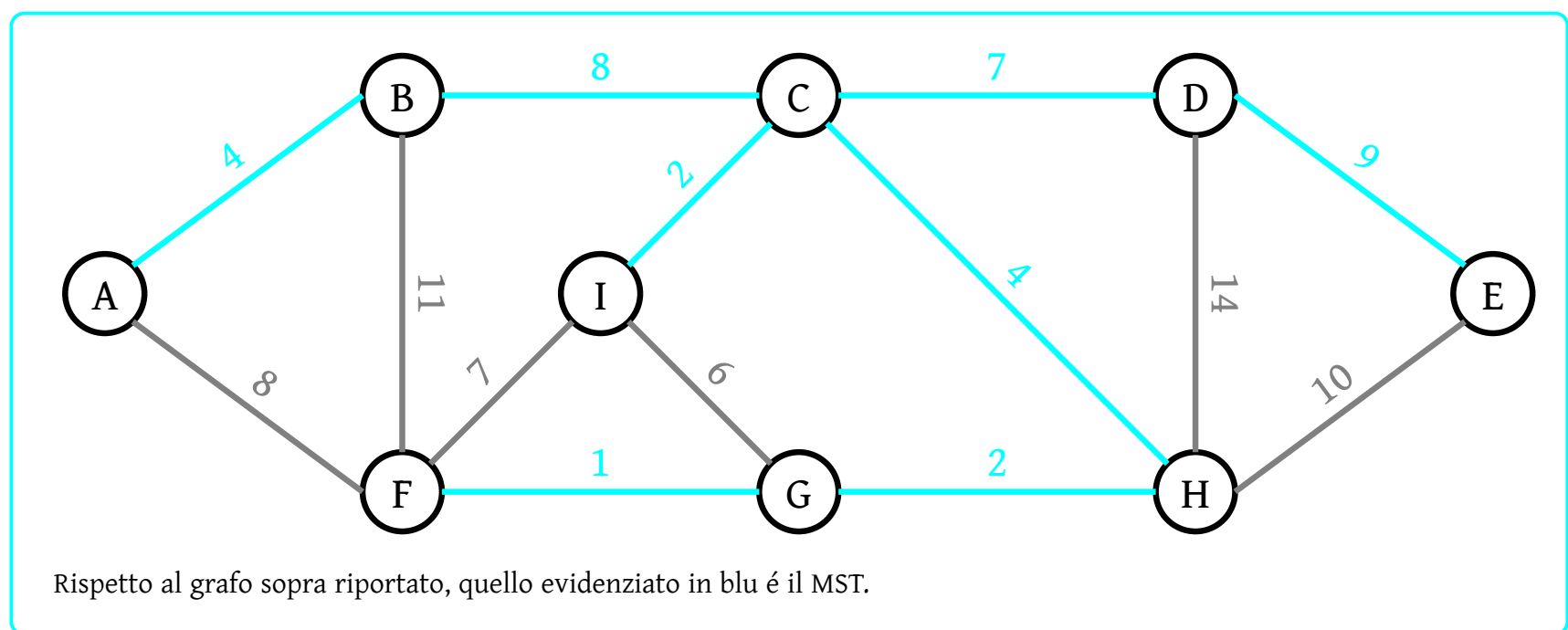
La procedura  $\text{CONNECTED-COMPONENTS}$  costruisce un insieme per ciascun vertice. Dopodich , unisce tutti gli insiemi i cui rappresentanti hanno un arco nel grafo che li unisce. In questo modo, vengono a crearsi tanti insiemi quante sono le componenti connesse di un grafo. A questo punto   possibile utilizzare la procedura  $\text{SAME-COMPONENT}$  per sapere se due vertici si trovano nella stessa componente.

## 2.7 Minimum Spanning Tree

Sia dato un grafo connesso non orientato pesato  $G = (V, E)$ , con  $W$  funzione che associa un intero positivo a ciascuna coppia di archi di  $G$ . Il **problema del minimum spanning tree (MST problem)** richiede di trovare l'albero  $T$ , sottoinsieme di  $E$ , tale per cui  $G_T = (V, T)$    lo spanning tree per il grafo  $G$  che ha peso complessivo minimo. Il peso di uno spanning tree   dato dalla somma dei pesi di tutti gli archi di cui   composto.

$$\forall v \in V, \exists (u, v) \in T$$

$$W(T) = \sum_{(u,v) \in T} W(u, v) \text{ minimo}$$



É evidente come il problema possa essere risolto mediante tecnica greedy, perché un insieme massimale di un matroide grafico é uno spanning tree. Si ha quindi che il minimum spanning tree é dato dall'insieme massimale del matroide grafico con peso minimo. In termini molto generici, l'algoritmo può essere descritto come segue:

1. Viene inizializzato un insieme  $A$  vuoto;
2. Viene scelto l'arco  $(u, v)$  di modo che  $A \cup (u, v)$  sia un sottoinsieme dell'insieme  $T$  degli archi del MST. Un arco che aggiunto ad  $A$  restituisce un insieme di archi ancora sottoinsieme dell'MST é detto **arco sicuro** per  $A$ ;
3. Se  $A = T$ , ovvero se  $G_A = (V, T)$  é il MST, l'algoritmo termina, altrimenti si torna al punto precedente.

```

procedure GENERIC-MST(G, W)
1  A ← ∅

2  while  $G_A = (V, A) \neq \text{MST}$  do
3      trova arco  $(u, v)$  sicuro per A
4      A ← A ∪ {(u, v)}

5  return A

```

L'algoritmo é effettivamente un algoritmo greedy perché ad ogni iterazione viene operata una scelta localmente ottima assumendo che lo sia anche complessivamente.

Si noti come la condizione di terminazione dell'algoritmo, ovvero valutare se il sottografo  $G_A = (V, A)$  é il MST, sia piuttosto vaga, dato che non é direttamente computabile. A tal proposito, si presti attenzione all'insieme  $A$ : in una qualsiasi iterazione, tale insieme definisce una foresta  $G_A = (V, A)$  per un grafo  $G = (V, E)$  avente  $|V| - |A|$  alberi.

Dato che per definizione lo Spanning Tree si ha quando la foresta é costituita da un solo albero, la condizione di terminazione dell'algoritmo può essere formulata come un controllo sulla differenza fra il numero di vertici del grafo ed il numero di alberi della foresta. In particolare, tale condizione é verificata quando la differenza fra i due é pari ad 1, perché questo significa che é rimasto un solo albero.

```

procedure GENERIC-MST(G, W)
1  A ← ∅

2  while  $|V| - |A| > 1$  do
3      trova arco  $(u, v)$  sicuro per A
4      A ← A ∪ {(u, v)}

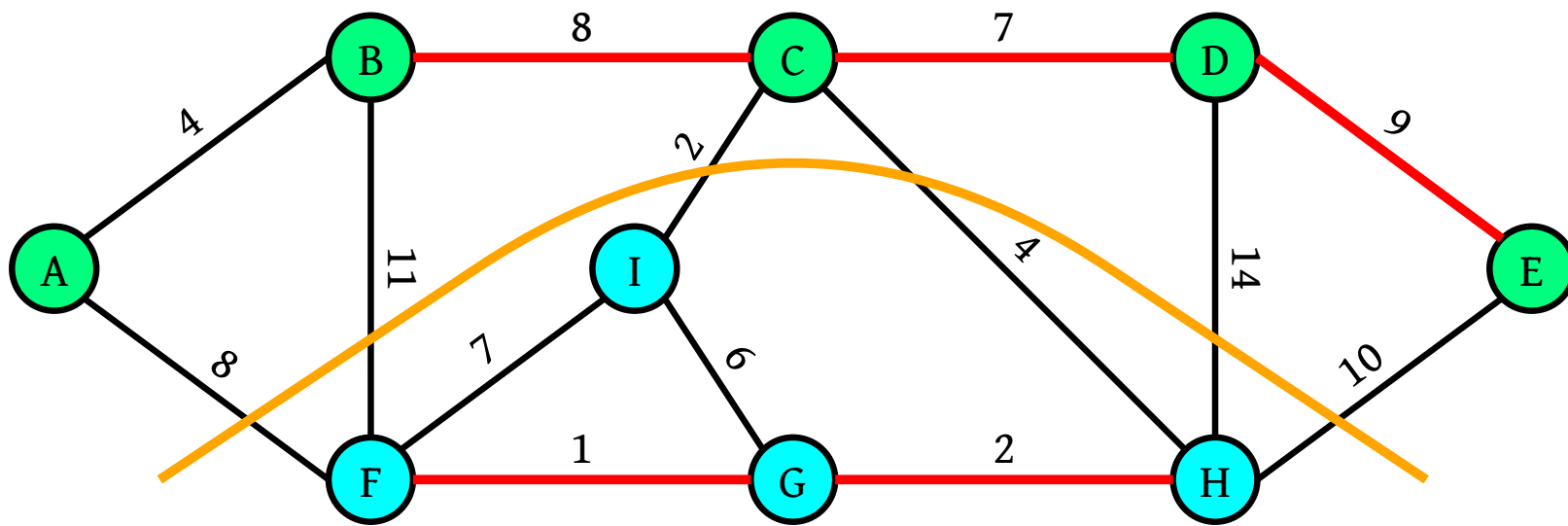
5  return A

```

Per rendere l'algoritmo effettivamente utilizzabile é infine necessario capire come compiere la scelta greedy, ovvero definire un metodo che discrimini quando un arco é un arco sicuro.

Dato un grafo  $G = (V, E)$ , viene detto **taglio** una qualsiasi partizione di  $V$  in due insiemi,  $V'$  e  $V - V'$ . Un arco  $(u, v) \in E$  si dice che **attraversa il taglio** se il vertice  $u$  appartiene a  $V'$  e  $v$  appartiene a  $V - V'$ . L'arco che attraversa il taglio avente peso minimo é detto **arco leggero**. Dato un sottoinsieme  $A \subseteq E$ , si dice che un taglio **rispetta** l'insieme  $A$  se nessun arco di  $A$  attraversa tale taglio.





Si consideri il grafo sopra presentato. Il taglio evidenziato partiziona i vertici del grafo in due sottoinsiemi,  $V = \{A, B, C, D, E\}$  e  $V - V' = \{F, G, H, I\}$ . Tale taglio rispetta il sottoinsieme di archi  $A = \{(B, C), (C, D), (D, E), (F, G), (G, H)\}$ . Gli archi che attraversano il taglio sono  $(A, F), (B, F), (C, I), (C, H), (D, H), (E, H)$ ; fra questi, l'arco  $(C, I)$  è l'arco leggero per il taglio, essendo quello fra questi di peso minimo.

**Teorema dell'arco sicuro.** Siano dati un grafo connesso non orientato e pesato  $G = (V, E, W)$ , un sottoinsieme  $A$  dell'insieme  $T$  di archi del MST e un qualsiasi taglio che rispetti  $A$ . L'arco leggero  $(u, v)$  del taglio è certamente un arco sicuro per  $A$ , ovvero  $A \cup \{(u, v)\} \subseteq T$ .

**Dimostrazione.** Se  $(u, v)$  è l'unico arco che attraversa il taglio il teorema è dimostrato, dato che è l'unico arco in grado di connettere le due componenti distinte nel MST.

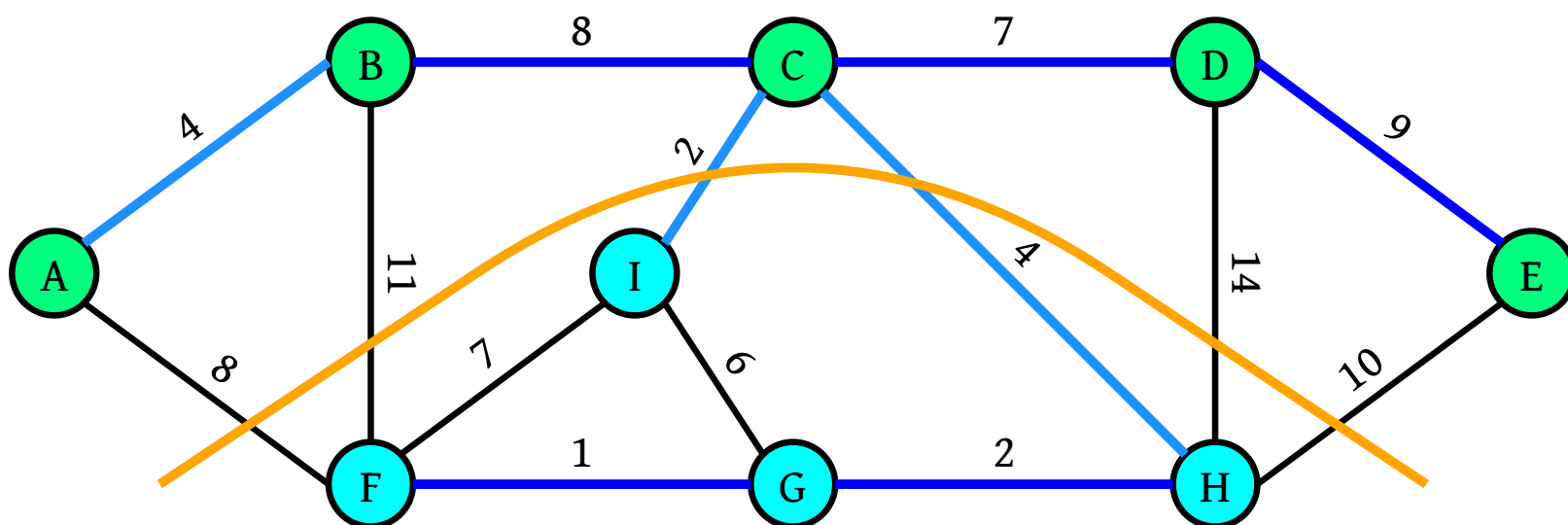
Si consideri invece il caso in cui esista almeno un altro arco che attraversa il taglio. Si supponga per assurdo che l'arco leggero  $(u, v)$  del taglio non sia un arco sicuro per  $A$ , ovvero che aggiungendo  $(u, v)$  all'insieme  $A$  non si ottiene un sottoinsieme di  $T$ . Deve allora esistere un'altro arco che attraversa il taglio e che appartiene a  $T$ , sia questo  $(x, y)$ .

L'arco  $(x, y)$  non può appartenere ad  $A$ , perché per definizione il taglio rispetta  $A$ . Se da  $T$  viene rimosso l'arco  $(x, y)$  si ottengono due alberi separate; se viene poi aggiunto  $(u, v)$  i due alberi si ricongiungono per ottenere l'albero di connessione, non necessariamente minimo,  $T' = T - \{(x, y)\} \cup \{(u, v)\}$ . Poiché per ipotesi  $(u, v)$  è l'arco leggero per il taglio, il peso di  $(u, v)$  deve essere inferiore al peso di  $(x, y)$ . Allora:

$$W(T') = W(T) - W(x, y) + W(u, v) \leq W(T)$$

Essendo però  $T$  un albero di connessione minimo, deve aversi  $W(T) \leq W(T')$ ; si ha quindi  $W(T') \leq W(T) \wedge W(T) \leq W(T')$ , ma questo significa  $W(T) = W(T')$ , e quindi anche  $T'$  è un albero di connessione minimo.

Dato che per ipotesi  $A \subseteq T$  e  $(x, y) \notin A$ , allora  $A \subseteq T'$ . Questo significa che  $A \cup \{(u, v)\} \subseteq T'$ . Ma questa è la definizione di arco sicuro, ed avendo appena mostrato che  $T'$  è un MST deve aversi che  $(u, v)$  è un arco sicuro per  $A$ .



Il sottoinsieme di archi  $A = \{(B, C), (C, D), (D, E), (F, G), (G, H)\}$  è parte degli archi del MST. Per questo motivo, il teorema precedente stabilisce che l'arco leggero del taglio che rispetta  $A$  è un arco sicuro per  $A$ , e infatti l'arco  $(C, I)$  è uno degli archi del MST.

**Corollario del teorema dell'arco sicuro.** Siano dati un grafo connesso non orientato e pesato  $G = (V, E)$  ed un sottoinsieme  $A$  dell'insieme  $T$  di archi del MST che definisce una foresta  $G_A = (V, A)$ . Sia  $C = (V_C, A_C)$ , con  $V_C \subseteq V$  e  $A_C \subseteq A$  uno qualsiasi degli alberi di  $G_A$ . Il taglio che divide i vertici di  $G_A$  nei due sottoinsiemi  $V_C$  e  $V - V_C$  rispetta l'insieme  $A$ . Inoltre, un arco leggero per tale taglio è anche arco sicuro per l'insieme  $A$ .

**Dimostrazione.** Il taglio  $(V_C, V - V_C)$  rispetta  $A$  e  $(u, v)$  è un arco leggero per questo taglio, quindi  $(u, v)$  è sicuro per  $A$ .

Sia  $A$  il sottoinsieme degli archi del MST di una generica iterazione dell'algoritmo, e sia  $G_A$  la foresta indotta dall'insieme  $A$ . Il corollario appena mostrato stabilisce che per trovare un arco sicuro per  $A$ , e che può essere aggiunto a quest'ultimo, è sufficiente scegliere uno qualsiasi degli alberi di  $G_A$  e trovare l'arco di peso minimo che collega un vertice che appartiene a tale albero con un vertice che non vi appartiene.

Ci si chiede allora come determinare in una qualsiasi iterazione (ed in maniera efficiente) quale sia l'arco di peso minimo da aggiungere all'insieme di archi del MST in costruzione. Per farlo vi sono diversi approcci, e ciascuno di questi è associato ad un algoritmo diverso.

### 2.7.1 Algoritmo di Kruskal

L'**Algoritmo di Kruskal** è un algoritmo greedy che permette di ricavare il MST di un grafo a partire dall'algoritmo greedy standard. Innanzitutto, sia  $G = (V, E, W)$  un grafo connesso, non orientato e pesato; a questo è possibile associare un matroide grafico  $M_G = (S, F)$ . Come già detto, un qualsiasi insieme massimale di archi  $M_G$  è uno spanning tree, pertanto il MST è quell'insieme massimale avente peso minimo.

L'algoritmo greedy standard risolve problemi di ottimizzazione di massimo, ma la ricerca di un MST (come dice il nome) è un problema di ottimizzazione di minimo. A tale scopo occorre, come già visto in precedenza, sostituire alla funzione peso  $W$  la funzione peso  $W'$ , dove ciascun valore  $i$ -esimo è dato dalla differenza fra il massimo valore in  $W$  e l' $i$ -esimo valore di  $W$ .

Rifacendosi all'algoritmo greedy standard, l'insieme che sarà passato in input all'algoritmo (e che viene poi ordinato) sarà l'insieme degli archi del grafo. Si ricordi che tale insieme va ordinato, secondo ordine decrescente, rispetto a  $W'$ , e non rispetto a  $W$ . In realtà questo equivale ad ordinare gli elementi rispetto a  $W$  ma in ordine inverso (in ordine crescente), pertanto è possibile applicare tale semplificazione.

```

KRUSKAL-MST(E)
  A ← ∅
  E = <e1, ..., en> ordinati per peso crescente

  for i from 1 to n do
    if ei è arco sicuro then
      A ← A ∪ {ei}

  return A

```

Per sapere se l' $i$ -esimo arco è un arco sicuro occorre applicare il corollario al teorema dell'arco sicuro: un arco è sicuro se connette due alberi distinti di  $G_A$ .

```

KRUSKAL-MST(V, E, W)
  A ← ∅
  E = <e1, ..., en> ordinati per peso crescente

  for i from 1 to n do
    (u, v) ← ei
    if u e v ∉ stesso albero in GA = (V, A) then
      A ← A ∪ {ei}

  return A

```

L'algoritmo funziona perché, ordinando gli elementi in ordine crescente, è garantito che fra tutti i possibili archi che connettono due alberi distinti verrà sempre scelto quello di peso minimo.

L'ultimo passo consiste nel determinare che struttura dati usare per salvare le informazioni relative agli alberi. Questo può essere fatto in maniera efficiente mediante una union-find e le relative primitive: per ogni vertice viene creato un insieme che contiene solo quel vertice, ed i confronti vengono fatti su questi insiemi. In particolare, se vengono trovati due insiemi disgiunti significa che non vi è alcun vertice in comune ai due insiemi, e che quindi è stato trovato un arco che unisce due alberi distinti. A questo punto è sufficiente sostituire i due insiemi con la loro unione.

```

KRUSKAL-MST(V, E, W)
1  A ← ∅
2  E = <e1, ..., en> ordinati per peso crescente
3  foreach v ∈ V do
4      MAKE_SET(v)

5  for i from 1 to n do
6      (u, v) ← ei
7      if FIND_SET(u) ≠ FIND_SET(v) then
8          A ← A ∪ {ei}
9          UNION(u, v)

10 return A

```

Per quanto riguarda il tempo di esecuzione dell'algoritmo, è possibile distinguere tre sezioni dello stesso: l'ordinamento degli archi rispetto al peso (riga 2), la costruzione dei set (righe 3 e 4) ed il loop principale (righe da 5 a 9):

1. L'ordinamento avviene con il consueto tempo logaritmico,  $O(n \log(n))$ . Dato che, nello specifico, l'ordinamento è fatto sugli archi di un grafo  $G = (V, E, W)$ , e che il numero di tali archi è dato dalla cardinalità di  $E$ , si ha che l'ordinamento degli archi ha tempo di esecuzione  $O(|E| \log(|E|))$ .
2. È possibile assumere che la costruzione dei set avvenga in tempo lineare per le proprietà della struttura dati union-find. Dato che viene costruito un set per ogni vertice, il tempo di esecuzione di tale costruzione è  $O(|V|)$ .
3. Il tempo di esecuzione del loop principale dipende interamente dal tempo di esecuzione delle primitive della union-find. Il confronto fra due set e l'unione di due set avviene in tempo  $\alpha$ , con  $\alpha$  costante fissata. Dato che il confronto avviene tante volte quanti sono gli archi del grafo, e dato che le unioni avvengono al massimo tante volte quanti sono i confronti, il tempo di esecuzione del loop principale è  $O(|E|\alpha)$ .

Il tempo di esecuzione complessivo è dato dalla somma dei tre tempi di esecuzione parziali, ovvero  $O(|E| \log(|E|)) + O(|V|) + O(|E|\alpha)$ . Dato che si sta considerando un grafo connesso, il numero di archi è necessariamente superiore al numero di vertici, pertanto è possibile effettuare una maggiorazione e sostituire  $|V|$  con  $|E|$ . Si ha quindi  $O(|E| \log(|E|)) + O(|E|) + O(|E|\alpha)$ , che asintoticamente equivale a  $O(|E| \log(|E|)) + O(|E|\alpha)$ . A sua volta, la costante  $\alpha$  è certamente inferiore a  $\log(|V|)$ , che è a sua volta certamente inferiore a  $\log(|E|)$ . È quindi possibile maggiorare il tempo di esecuzione come  $O(|E| \log(|E|)) + O(|E| \log(|E|))$ , ovvero  $O(2|E| \log(|E|))$ , ma che asintoticamente equivale semplicemente a  $O(|E| \log(|E|))$ .

### 2.7.2 Algoritmo di Prim

L'**Algoritmo di Prim** è un algoritmo greedy che permette di ricavare il MST di un grafo che non si rifà all'algoritmo greedy standard. In termini molto generali, l'idea dell'algoritmo è la seguente:

1. Dato un grafo  $G = (V, E, W)$ , viene scelto arbitrariamente un vertice  $r$ , e si considera l'albero (degenere)  $C$  formato da questo vertice isolato;
2. Viene trovato l'arco di peso minimo che connette un vertice in  $C$  con un vertice  $v$  non in  $C$  e si aggiunge  $v$  a  $C$ ;
3. L'algoritmo viene ripetuto fino ad esaurire tutti i vertici.

L'algoritmo funziona perché ad ogni iterazione i vertici del grafo vengono divisi in due sottoinsiemi: i vertici dell'albero in costruzione (l'insieme  $V_C$ ) e tutti quelli rimanenti (l'insieme  $V - V_C$ ). Per definizione, tale taglio rispetta l'insieme degli archi di  $C$ , pertanto l'arco di peso minimo che connette un vertice in  $V_C$  con un vertice in  $V - V_C$  è certamente un arco sicuro per l'arco in costruzione.

Per determinare quale sia l'arco corretto da considerare, l'algoritmo associa a ciascun vertice  $v$  del grafo  $G = (V, E, W)$  due campi: un campo chiave ( $v.key$ ) ed un campo predecessore ( $v.\pi$ ). Il campo chiave contiene il peso dell'arco che, rispetto all'iterazione corrente, è il minimo trovato, mentre il campo predecessore contiene il vertice con il quale il vertice corrente ha l'arco con peso più piccolo. Dato un grafo  $G = (V, E, W)$ , si ha allora:

1. I campi chiave di tutti i vertici vengono inizializzati a  $\infty$ , mentre i campi predecessore vengono inizializzati a NULL;
2. Viene scelto un vertice arbitrario, al quale viene assegnato come campo chiave il valore 0. Dopodiché, tutti i vertici vengono inseriti in una coda di min-priority, dove il valore della chiave definisce l'ordine (la priorità) con cui i vertici vengono estratti. In questo modo, il vertice estratto è sempre quello con valore del campo chiave più piccolo;
3. Viene operata un'estrazione ed il vertice  $v$  estratto viene confrontato con i vertici a questo adiacenti che ancora si trovano nella coda. Per ciascuno di questi, se il relativo campo chiave contiene un valore maggiore di  $v.key$  allora come campo predecessore viene posto  $v$  e come campo chiave viene posto il peso dell'arco che connette  $v$  a tale vertice;
4. Se la coda di priorità non è vuota, l'algoritmo riprende dal punto 2, altrimenti si procede oltre, perché è possibile assumere che a tutti i vertici tranne quello di partenza è stato univocamente assegnato un predecessore;
5. L'insieme degli archi del MST viene costruito riportando, per ciascun vertice, l'arco (quale che sia) che ha con il rispettivo predecessore.

```

PRIM-MST(V, E, W, r)
1  foreach v ∈ V do
2      v.key ← ∞
3      v.π ← NULL
4      r.key ← 0

5  Aggiungi tutti i vertici di V alla coda Q

6  while Q ≠ ∅ do
7      u ← POP(Q)
8      foreach v ∈ ADJ(u) do
9          if v ∈ Q and W(u, v) < v.key then
10             v.key ← W(u, v)
11             v.π ← u

12  A ← ∅
13  foreach v ∈ V do
14      A ← A ∪ {(v, v.π)}

15  return A

```

Per quanto riguarda il tempo di esecuzione dell'algoritmo, é possibile distinguere quattro sezioni dello stesso: la costruzione della coda di priorit  (righe da 1 a 5), il loop principale (righe da 6 a 11) e la ricostruzione della soluzione (righe da 12 a 14)

1. Il campo chiave ed il campo predecessore vengono aggiunti a ciascun vertice in tempo lineare. Allo stesso modo, il tempo di esecuzione dell'aggiunta dei vertici alla coda é proporzionale al numero di vertici, perch  tutti i vertici inizialmente hanno la stessa priorit  e quindi possono essere aggiunti in ordine casuale. Il tempo di esecuzione della costruzione della coda é allora  $O(|V|)$ ;
2. Il loop principale viene eseguito una volta per ciascun vertice del grafo. L'estrazione del vertice in cima alla coda é una operazione che richiede tempo di esecuzione logaritmico, perch  per ricavare il vertice con campo chiave pi  piccolo occorre effettuare una ricerca binaria. Allo stesso modo, l'analisi di ciascun vertice adiacente al vertice corrente viene compiuta in tempo logaritmico. Il tempo di esecuzione del loop principale é allora  $O(|V|\log(|V|))$ ;
3. La ricostruzione della soluzione viene eseguita in tempo proporzionale al numero dei vertici, perch  l'aggiunta di un singolo vertice avviene in tempo costante e ciascun vertice viene analizzato esattamente una sola volta. Si ha allora che il tempo di esecuzione della ricostruzione della soluzione é  $O(|V|)$ .

Il tempo di esecuzione complessivo é dato dalla somma dei tre tempi di esecuzione parziali, ovvero  $O(|V|) + O(|V|\log(|V|)) + O(|V|)$ , che asintoticamente equivale a  $O(|V|) + O(|V|\log(|V|))$ . Essendo il tempo logaritmico pi  influente sul tempo complessivo rispetto a quello lineare, si ha che l'equazione equivale asintoticamente a  $|V|\log(|V|)$ . Dato che si sta considerando un grafo connesso, il numero di archi é necessariamente superiore al numero di vertici, pertanto é possibile effettuare una maggiorazione e sostituire  $|V|$  con  $|E|$ . Si ha allora che il tempo di esecuzione dell'algoritmo di Prim é  $|E|\log(|E|)$ .

## 2.8 Cammini minimi da sorgente unica

Sia dato un grafo  $G = (V, E, W)$  non orientato, pesato e connesso. Si consideri una funzione peso  $W$  che restituisce esclusivamente valori strettamente positivi. Sia poi  $s \in V$  un vertice "privilegiato", chiamato **vertice sorgente**. Il **problema dei cammini minimi da sorgente unica** richiede di trovare, rispetto ad un vertice fissato  $s$ , il cammino di minima lunghezza da ciascun vertice del grafo ad  $s$ .

Si consideri un cammino minimo  $P = \langle v_1, v_2, \dots, v_{k-1}, v_k \rangle$ . Come gi  dimostrato per l'algoritmo Floyd-Warshall, sono a loro volta minimi anche tutti i sottocammini  $P_{ij} = \langle v_i, v_{i+1}, \dots, v_j \rangle$  con  $1 \leq i < j \leq k$ . In particolare, é minimo il sottocammino  $\langle v_1, v_2, \dots, v_{k-1} \rangle$  dove  $v_{k-1}$  é il **predecessore** di  $v_k$  nel cammino minimo  $P$ .

Sia  $\delta(s, v)$  il peso complessivo del cammino minimo dalla sorgente  $s$  ad un qualsiasi vertice  $v$ , ottenuto sommando i pesi di tutti gli archi  $(i, j)$  che compongono tale cammino. Sia  $u$  il predecessore di  $v$  su tale cammino; il peso del cammino pu  essere scomposto come somma fra il peso del sottocammino minimo da  $s$  a  $u$  ed il peso dell'arco che connette  $u$  a  $v$ . Per indicare che  $v_i$  é il predecessore di  $v_j$  rispetto ad un certo cammino si usa la notazione  $v_i = \pi(v_j)$ .

$$\delta(s, v) = \sum_{(i,j) \in P, i=\pi(j)} W(i, j) = \delta(s, u) + W(u, v)$$

Nel caso in cui  $u$  non sia il predecessore di  $v$  sul cammino minimo da  $s$  a  $v$ , ma comunque  $W(u, v)$  esiste, si ha  $\delta(s, v) \leq \delta(s, u) + W(u, v)$ . Questo perch  se si mantiene il medesimo  $\delta(s, u)$ , il valore  $W(u, v)$  non pu  che essere maggiore del peso dell'arco minimo che unisce  $u$  a  $v$ , altrimenti coinciderebbe con tale arco.

Un algoritmo che permette di risolvere il problema dei cammini minimi da sorgente unica é l'**Algoritmo di Dijkstra**. L'algoritmo associa a ciascun vertice  $v$  del grafo in esame un campo  $d$  ed un campo  $\pi$ : il primo contiene un limite superiore per  $\delta(s, v)$ , mentre il secondo riporta uno dei vertici con i quali  $v$  é congiunto da un arco.

Prima dell'esecuzione dell'algoritmo, al vertice sorgente viene assegnato 0 come campo  $d$ , mentre a tutti gli altri vertici viene assegnato  $\infty$ , mentre il campo  $\pi$  viene inizializzato a NULL per tutti i vertici. Alla fine dell'esecuzione, per ciascun vertice il campo  $d$  contiene precisamente il peso del cammino minimo dalla sorgente al vertice, mentre il campo  $\pi$  riporta il predecessore del vertice nel cammino minimo dalla sorgente al vertice.

Per un qualsiasi vertice  $v$ , l'algoritmo di Dijkstra permette al valore di  $v.d$  di passare dal valore iniziale  $\infty$  al valore finale  $\delta(s, v)$  (l'effettivo peso del cammino minimo da  $v$  alla sorgente) applicando una serie di "miglioramenti", detti **rilassamenti**. Presi due vertici  $u$  e  $v$ , se  $v.d$  é maggiore della somma fra  $u.d$  ed il peso dell'arco che unisce  $u$  e  $v$ , allora significa che il cammino che va dalla sorgente a  $v$  passando per  $u$  come predecessore é un cammino con peso inferiore, e quindi migliore, di quello attualmente salvato per  $v$ . Un rilassamento consiste per l'appunto nel valutare se  $v.d > u.d + W(u, v)$  e, in caso affermativo, sostituire  $v.d$  con  $u.d + W(u, v)$  e  $v.\pi$  con  $u$ .

In particolare, l'algoritmo di Dijkstra opera esattamente un solo rilassamento per ciascun vertice. Dato un grafo non orientato, pesato e connesso  $G = (V, E, W)$ , il procedimento é riportato di seguito:

1. A ciascun vertice del grafo viene associato un campo  $d$ , inizializzato a  $\infty$ , ed un campo  $\pi$ , inizializzato a NULL.
2. Viene scelto il noto sorgente, indicato con  $s$ , il cui campo  $d$  viene impostato a 0. Dopodiché, tutti i vertici vengono inseriti in una coda di min-priority, dove la priorità é determinata dal valore del campo  $d$ ;
3. Viene estratto il vertice avente campo  $d$  con valore minore, sia questo  $v$ ;
4. Per ciascun vertice  $u$  adiacente a  $v$  viene operato il rilassamento: se il valore  $v.d$  é maggiore della somma fra  $u.d$  e  $W(u, v)$ , allora  $v.d$  viene sostituito con  $u.d + W(u, v)$  e  $v.\pi$  viene sostituito con  $u$ ;
5. Se la coda non é vuota si riprende l'esecuzione dal punto 3, altrimenti l'algoritmo termina.

```

DIJKSTRA(V, E, W, s)
1  foreach v ∈ V do
2      v.d ← ∞
3      v.π ← NULL
4  s.d ← 0

5  Aggiungi tutti i vertici di V alla coda Q
6  S ← ∅

7  while Q ≠ ∅ do
8      u ← POP(Q)
9      S ← S ∪ {u}
10     foreach v ∈ ADJ(u) do
11         if v.d > u.d + W(u, v) then
12             v.d ← u.d + W(u, v)
13             v.π ← u

```

Dato un grafo connesso, non orientato e pesato  $G = (V, E, W)$ , sia  $s \in V$  un vertice eletto a sorgente. L'algoritmo di Dijkstra restituisce correttamente il cammino minimo dalla sorgente  $s$  a tutti i vertici in  $V$ .

**Dimostrazione.** Sia  $\langle v_1 = s, v_2, \dots, v_n \rangle$  l'ordine con cui i vertici del grafo vengono estratti dalla coda durante l'esecuzione dell'algoritmo di Dijkstra. L'algoritmo é corretto se, quando viene estratto il  $k$ -esimo vertice, si ha  $v_k.d = \delta(s, v_k)$  per tutti i  $k \in \{1, n\}$ . Tale proprietà può essere dimostrata per induzione.

Il caso base é dimostrato se  $s.d = \delta(s, s)$  quando viene estratto  $s$ . Dato che per raggiungere  $s$  da  $s$  non occorre percorrere alcun cammino, la distanza da  $s$  a sé stesso é 0. In effetti l'algoritmo inizializza a 0 il valore di  $s.d$ , quindi quando  $s$  viene estratto il valore di  $s.d$  é effettivamente 0.

Per quanto riguarda il passo ricorsivo, occorre mostrare che per il vertice  $v_k$ , quando viene estratto, vale  $v_k.d = \delta(s, v_k)$  assumendo che la relazione valga per  $v_{k-1}$ . Si supponga per assurdo che questo non sia vero, e che quindi  $v_k.d \neq \delta(s, v_k)$ . Il vertice  $v_k$  non può essere  $s$ , perché é stato appena mostrato che il teorema per  $s$  é valido. Deve allora essere un vertice che si trova dopo di  $s$ , e deve necessariamente esistere un cammino che va da  $s$  a  $v_k$ , dato che altrimenti si avrebbe  $\delta(s, v_k) = \infty$  e questo contraddirebbe l'ipotesi assunta per assurdo  $v_k.d \neq \delta(s, v_k)$ . Esistendo (almeno) un cammino da  $s$  a  $v_k$ , esisterá anche (almeno) un cammino minimo da  $s$  a  $v_k$ .

Sia  $v_y$  il primo vertice lungo il cammino minimo da  $s$  a  $v_k$  ad essere ancora parte della coda di priorità, e sia  $v_x$  il predecessore di  $v_y$ . Il cammino da  $s$  a  $v_k$  può essere allora scomposto come  $s \rightsquigarrow v_x \rightarrow v_y \rightsquigarrow v_k$ . Si noti come  $v_k$  potrebbe coincidere con  $v_y$ , cosí come  $s$  potrebbe coincidere con  $v_x$ , pertanto i sottocammini  $s \rightsquigarrow v_x$  e  $v_y \rightsquigarrow v_k$  potrebbero essere vuoti.

Alla fine dell'iterazione per  $v_x$ , essendo  $v_x$  predecessore di  $v_y$  ed essendo per ipotesi  $v_x$  un vertice già estratto dalla coda, si avrà  $v_y.d = \delta(s, v_y)$ . Poiché  $v_y$  precede  $v_k$  nel cammino minimo da  $s$  a  $v_k$ , deve aversi  $\delta(s, v_y) \leq \delta(s, v_k)$ . Quindi:

$$v_y.d = \delta(s, v_y) \leq \delta(s, v_k) \leq v_k.d$$

Tuttavia, poiché entrambi i vertici  $v_k$  e  $v_y$  si trovavano ancora nella coda quando  $v_k$  viene estratto, si ha  $v_k.d \leq v_y.d$ . Ma allora le disuguaglianze sopra espresse sono di fatto delle uguaglianze:

$$v_y.d = \delta(s, v_y) = \delta(s, v_k) = v_k.d$$

Si ha però  $\delta(s, v_k) = v_k.d$ , che é in contraddizione con l'ipotesi assunta per assurdo. Occorre quindi concludere che effettivamente  $\delta(s, v_k) = v_k.d$ , ed il teorema é provato.

# Capitolo 3

## Visite di grafi

### 3.1 Visita in ampiezza

Una breadth-first search di un grafo  $G = (V, E)$  non orientato e non pesato a partire da una sorgente  $s$  viene indicata come  $\text{BFS}(G, s)$ . Concettualmente, una BFS opera come segue:

1. Viene visitata la sorgente  $s$ ;
2. Vengono visitati uno dopo l'altro tutti i vertici adiacenti ad  $s$ ;
3. Vengono visitati uno dopo l'altro tutti i vertici adiacenti dei vertici adiacenti ad  $s$ ...

In altri termini, vengono prima calcolati i vertici che distano 0 da  $s$  (sé stesso), poi i vertici che distano 1 da  $s$ , poi i vertici che distano 2 da  $s$ , ... dove con **distanza fra due vertici** si intende il numero minimo di archi che compongono un cammino fra i due.

In una visita in ampiezza vengono raggiunti una ed una sola volta tutti e i soli vertici raggiungibili a partire dalla sorgente, e permette di ricavarne la distanza dalla sorgente. Nel fare questo viene a costituirsi una struttura ad albero, con la sorgente come radice e dove al livello  $i$ -esimo dell'albero si trovano tutti i vertici che distano  $i$  dalla sorgente.

L'algoritmo per la costruzione di una BFS prevede di associare tre campi a ciascun vertice:

- Un campo  $c$ , che contiene il *colore* associato al vertice. Tale colore indica lo stato del vertice all'iterazione corrente dell'algoritmo. I colori in questione sono tre:

- ☐ Bianco: il vertice non é stato ancora visitato;
- ☐ Grigio: il vertice é stato visitato ma alcuni vertici a questo adiacenti non sono stati ancora visitati;
- ☐ Nero: il vertice é stato visitato e tutti vertici a questo adiacenti sono stati visitati;

- Un campo  $d$ , che contiene la distanza del vertice dalla sorgente;
- Un campo  $\pi$ , che contiene il predecessore del vertice nell'albero indotto dalla visita.

All'inizio della visita, tutti i vertici hanno assegnato il colore bianco, un valore  $d$  pari a  $\infty$  (ad eccezione della sorgente che ha assegnato 0) ed un valore  $\pi$  pari a NULL. Alla fine della visita, tutti i vertici raggiungibili dalla sorgente avranno assegnato il colore nero ed i relativi valori per  $d$  e per  $\pi$ , mentre i vertici non raggiungibili dalla sorgente avranno ancora assegnato il colore bianco, un valore  $d$  ancora pari a  $\infty$  ed un valore  $\pi$  ancora pari a NULL.

L'algoritmo per la costruzione di una BFS utilizza una coda di priorità, nella quale vengono messi i vertici non appena vengono visitati e nella quale rimangono fino a quando estratti per esplorarne gli adiacenti. L'algoritmo termina quando la coda é vuota, ovvero quando non esistono più vertici grigi. Infatti, un vertice può venire associato il colore grigio solamente durante la visita, mai prima e mai dopo.

Una volta assegnati i valori corretti ai tre campi per ciascun vertice é possibile costruire l'albero  $T = (V_T, E_T)$  indotto dalla visita a partire dai valori  $\pi$ . Tale albero conterrà tutti i vertici che alla fine dell'esecuzione dell'algoritmo hanno bianco come colore e come archi tutti gli archi formati da un vertice di  $V_T$  e dal rispettivo predecessore.

```
BFS(V, E, s)
1  foreach v ∈ (V - {s}) do
2      v.color ← W
3      v.d ← ∞
4      v.π ← NULL

5  s.color ← G
6  s.d ← 0

7  Q ← ∅
8  enqueue(Q, s)

9  while Q ≠ ∅ do
10     v ← head(Q)
11     foreach v ∈ adj(v) do
12         if u.color == W then
13             u.color ← G
14             enqueue(Q, u)
15             u.d ← v.d + 1
16             u.π ← v
17     v.color ← B
18     dequeue(Q)
```

```

19  Et ← ∅
20  Vt ← ∅
21  foreach v ∈ (V - {s}) do
22      if v.π ≠ NULL then
23          Vt ← Vt ∪ {v}
24          Et ← Et ∪ {(v, v.π)}

```

Il costo di inizializzazione dell'algoritmo in termini di tempo di esecuzione é  $O(|V|)$ , perché viene compiuto un numero costante di operazioni per ciascun vertice. Una singola operazione sulla coda può considerarsi eseguita in un tempo costante. Ciascun vertice viene inserito e/o estratto dalla coda non più di una sola volta. Questo significa che il numero massimo di volte che l'algoritmo ispeziona un vertice é  $2E$ , e che quindi il costo del loop principale dell'algoritmo in termini di tempo di esecuzione é  $O(|E|)$ . Il tempo di esecuzione complessivo dell'algoritmo viene allora ad essere  $O(|V| + |E|)$ .

### 3.2 Visita in profondità

Una depth-first search di un grafo  $G = (V, E)$  orientato e non pesato a partire da una sorgente  $s$  viene indicata come  $\text{DFS}(G, s)$ . Concettualmente, una DFS opera come segue:

1. Viene visitata la sorgente  $s$ ;
2. Viene visitato il primo adiacente  $a_1$  di  $s$ , poi viene visitato il primo adiacente  $a_2$  di  $a_1$ , poi visita il primo adiacente  $a_3$  di  $a_2$ , ...
3. Quando viene raggiunto un vertice che non ha adiacenti da visitare si risale al predecessore e la visita riparte (se possibile) da un altro adiacente di tale predecessore;
4. Ogni volta che non ci sono adiacenti da visitare si risale al predecessore;
5. Quando la visita risale alla sorgente e  $s$  non ha più adiacenti da visitare, si sceglie una nuova sorgente e la visita riparte;
6. La visita termina quando non ci sono più vertici disponibili per essere scelti come nuova sorgente.

Dato che potenzialmente ogni vertice può diventare sorgente nella stessa visita in profondità, e dato che la visita termina quando non vi sono più vertici elegibili a sorgente, questo significa che tutti i vertici del grafo verranno sempre raggiunti. In particolare, verrà a crearsi una foresta di alberi dove ogni albero rappresenta una delle componenti connesse del grafo.

L'algoritmo per la costruzione di una DFS prevede di associare quattro campi a ciascun vertice:

- Un campo  $c$ , che contiene il *colore* associato al vertice. Tale colore indica lo stato del vertice all'iterazione corrente dell'algoritmo. I colori in questione sono tre:
  - ☐ Bianco: il vertice non é stato ancora visitato;
  - ☐ Grigio: il vertice é stato visitato ma alcuni vertici a questo adiacenti non sono stati ancora visitati;
  - ☐ Nero: il vertice é stato visitato e tutti vertici a questo adiacenti sono stati visitati;
- Un campo  $\pi$ , che contiene il predecessore del vertice nella foresta indotta dalla visita;
- Un campo  $d$ , che contiene l'istante temporale associato alla "scoperta" del vertice;
- Un campo  $f$ , che contiene l'istante temporale associato alla completa esplorazione del vertice.

Prima della visita, tutti i vertici hanno assegnato il colore bianco; tuttavia, a differenza della visita in ampiezza, alla fine della visita tutti i vertici avranno assegnato il colore nero. Il colore grigio può comparire solamente *durante* la visita, mai prima e mai dopo.

Prima della visita, il campo  $\pi$  é NULL per tutti i vertici. Dopo la visita, il campo  $\pi$  sarà rimasto NULL solamente per quei vertici che sono stati scelti come sorgente, perché per definizione le sorgenti non hanno un predecessore.

L'algoritmo fa uso di un contatore che viene incrementato di uno ogni volta che si avanza di una iterazione. Il valore di tale contatore viene assegnato al campo  $d$  di un vertice quando viene preso in considerazione per la prima volta (ovvero, quando gli viene assegnato il colore grigio), mentre viene assegnato al campo  $f$  quando tutti i vertici adiacenti sono stati esplorati (ovvero, quando gli viene assegnato il colore nero). Naturalmente, il valore del campo  $d$  di un vertice deve essere sempre inferiore al valore del suo campo  $f$ . La differenza fra  $f$  e  $d$  restituisce il tempo effettivo che é stato necessario per esplorare completamente il vertice.



**Teorema delle parentesi.** Dato un grafo  $G = (V, E)$  orientato e non pesato, si operi una visita in profondità su tale grafo. Si consideri una coppia di vertici  $u$  e  $v$  ed i relativi intervalli  $(u.d, u.f)$  e  $(v.d, v.f)$ :

- $(u.d, u.f)$  contiene  $(v.d, v.f)$ . Allora  $u$  e  $v$  si trovano in uno stesso albero della foresta indotta dalla visita, ed in particolare  $u$  è un predecessore di  $v$ ;
- $(v.d, v.f)$  contiene  $(u.d, u.f)$ . Allora  $u$  e  $v$  si trovano in uno stesso albero della foresta indotta dalla visita, ed in particolare  $v$  è un predecessore di  $u$ ;
- $(v.d, v.f)$  e  $(u.d, u.f)$  sono intervalli disgiunti. Allora  $u$  e  $v$  si trovano in due alberi distinti della foresta indotta dalla visita.

**Dimostrazione.** Si consideri la situazione  $u.d < v.d$  (la situazione opposta è del tutto analoga, basta invertire i vertici). Possono presentarsi due casi:

- $v.d < u.f$ . Questo significa che  $v$  è stato scoperto dopo il tempo di scoperta di  $u$  ma prima del tempo di fine esplorazione di  $u$ , il che significa che  $v$  è un discendente di  $u$ . Inoltre, poiché  $v$  è stato scoperto prima di  $u$ , verrà completamente esplorato prima della completa esplorazione di  $u$ . Ne segue che i vari tempi coinvolti sono disposti nell'ordine:

$$u.d < v.d < v.f < u.f$$

Che equivale a dire che l'intervallo  $(v.d, v.f)$  è interamente contenuto nell'intervallo  $(u.d, u.f)$

- $u.f < v.d$ . In questo caso, i due intervalli sono completamente disgiunti, in quanto l'ordine ottenuto è il seguente:

$$u.d < u.f < v.d < v.f$$

Infatti, questo equivale a dire che nessuno dei due vertici è stato scoperto mentre l'altro era grigio. Ne segue che anche  $u$  e  $v$  non sono discendenti l'uno dell'altro.

Un'altra interessante proprietà della visita in profondità è che questa può essere utilizzata per classificare gli archi del grafo in input. Questa può venire poi utilizzata per raccogliere informazioni interessanti sullo stesso. In particolare, una visita in ampiezza assegna ad ogni arco del grafo una etichetta fra quattro:

- **Arco d'albero**, indicato con la lettera **T**. Sono quegli archi  $(u, v)$  tali per cui  $v$  è un nodo di colore bianco quando l'arco viene esplorato. Dato che  $v$  non è stato visitato prima d'ora, il suo colore viene cambiato in grigio, e  $u$  diviene predecessore di  $v$ .
- **Arco all'indietro**, indicato con la lettera **B**. Sono quegli archi  $(u, v)$  tali per cui  $v$  è un nodo di colore grigio quando l'arco viene esplorato. Dato che  $v$  è già stato visitato in precedenza,  $u$  non è predecessore di  $v$  ma  $v$  è un antenato di  $u$  nello stesso albero della visita.
- **Arco in avanti**, indicato con la lettera **F**. Sono quegli archi  $(u, v)$  tali per cui  $v$  è un nodo di colore nero quando l'arco viene esplorato ed il tempo di inizio di  $u$  è inferiore a quello di  $v$ . Dato che  $v$  è già stato completamente esplorato,  $u$  non è predecessore di  $v$  ma  $u$  è un antenato di  $v$  nello stesso albero della visita.
- **Arco trasversale**, indicato con la lettera **C**. Sono quegli archi  $(u, v)$  tali per cui  $v$  è un nodo di colore nero quando l'arco viene esplorato ed il tempo di inizio di  $u$  è superiore a quello di  $v$ . In questo caso, né  $u$  né  $v$  sono l'uno l'antenato dell'altro, pertanto i due potrebbero essere parte di due alberi della visita distinti.

Sono stati presentati tutti gli elementi necessari ad introdurre l'algoritmo vero e proprio per la ricerca in profondità, presentato di seguito:

```

DFS(G)
1  foreach v ∈ V do
2      color[v] = W
3      π[v] = NULL
4      d[v] = 0
5      f[v] = 0
6  time = 0
7  foreach v ∈ V do
8      if color[v] is W then
9          DFS_visit(G, v)

DFS_visit(G, u)
10  time = time + 1
11  d[u] = time
12  color[u] = G
13  foreach v ∈ adj(u) do
14      if color[v] = W then
15          π[v] = u
16          DFS_visit(G, v)
17  color[u] = B
18  time = time + 1
19  f[u] = time

```

Il tempo di esecuzione per l'inizializzazione (righe da 1 a 5) é  $O(|V|)$ , dato che vengono eseguite delle operazioni con tempo di esecuzione unitario una volta per ogni vertice. La procedura `DFS_visit` viene chiamata all'interno di `DFS` una volta per ciascun arco, pertanto il tempo di esecuzione delle righe da 7 a 9 é  $|E|$ . Il tempo di esecuzione complessivo dell'algoritmo é allora  $O(|V|) + O(|E|)$ . L'algoritmo per la visita in profondità può essere esteso facilmente per includere l'etichettatura degli archi:

```
DFS(G)
1  foreach v ∈ V do
2      color[v] = W
3      π[v] = NULL
4      d[v] = 0
5      f[v] = 0
6  time = 0
7  foreach v ∈ V do
8      if color[v] is W then
9          DFS_visit(G, v)
```

```
DFS_visit(G, u)
10  time = time + 1
11  d[u] = time
12  color[u] = G
13  foreach v ∈ adj(u) do
14      if color[v] = W then
15          (u, v) = "arco T"
16          π[v] = u
17          DFS_visit(G, v)
18      else
19          if color[v] = G then
20              (u, v) = "arco B"
21          else
22              if d[u] < d[v] then
23                  (u, v) = "arco F"
24              else
25                  (u, v) = "arco C"
26  color[u] = B
27  time = time + 1
28  f[u] = time
```

L'algoritmo per la visita in profondità può essere esteso, con poche modifiche, anche ai grafi non orientati. Infatti, dato che gli archi di un grafo non orientato sono per definizione simmetrici, occorre specificare che non vanno analizzati i vertici predecessori del vertice in esame, altrimenti si entrerebbe in un loop infinito.

Un grafo non orientato, dopo una visita in profondità, ha solo archi d'albero e archi all'indietro.

```
DFS(G)
1  foreach v ∈ V do
2      color[v] = W
3      π[v] = NULL
4      d[v] = 0
5      f[v] = 0
6  time = 0
7  foreach v ∈ V do
8      if color[v] is W then
9          DFS_visit(G, v)
```

```
DFS_visit(G, u)
10  time = time + 1
11  d[u] = time
12  color[u] = G
13  foreach v ∈ adj(u) - {π[u]} do
14      if color[v] = W then
15          print "arco T"
16          π[v] = u
17          DFS_visit(G, v)
18      else
19          print "arco B"
20  color[u] = B
21  time = time + 1
22  f[u] = time
```

### 3.3 Ordinamento topologico

Si consideri un grafo orientato, diretto e aciclico (*Directed Acyclic Graph*, DAG). Un **ordinamento topologico** del DAG  $G = (V, E)$  é un ordinamento lineare di tutti i suoi vertici tale che, se  $G$  contiene un arco  $(u, v)$ , allora  $u$  compare prima di  $v$  in tale ordinamento.

Un ordinamento topologico viene eseguito estendendo l'algoritmo per la ricerca in profondità. In particolare: l'algoritmo per l'ordinamento topologico può essere descritto a grandi linee come segue:

1. Per ogni vertice viene effettuata una normale ricerca in profondità, in modo da calcolare il tempo di completamento  $v.f$  per ciascun vertice  $v$ ;
2. Una volta completata l'ispezione di un vertice, tale vertice viene inserito in uno stack;
3. Vengono stampati i vertici contenuti nello stack (nell'ordine in cui sono stati inseriti).

```

DFS(G)
1  foreach v ∈ V do
2      color[v] = W
3      π[v] = NULL
4      d[v] = 0
5      f[v] = 0
6  time = 0
7  S = []
8  foreach v ∈ V do
9      if color[v] is W then
10         DFS_visit(G, v)
11  while not isEmpty(S) do
12      pop(S)

DFS_visit(G, u)
13  time = time + 1
14  d[u] = time
15  color[u] = G
16  foreach v ∈ adj(u) do
17      if color[v] = W then
18          π[v] = u
19          DFS_visit(G, v)
20  color[u] = B
21  push(S, u)
22  time = time + 1
23  f[u] = time

```

Il tempo di esecuzione dell'algoritmo per l'ordinamento topologico é  $O(|V| + |E|)$ . L'ordinamento in profondità, anche se viene esteso con l'aggiunta di una operazione di push, richiede comunque tempo  $O(|V| + |E|)$ , perché tale operazione ha tempo di esecuzione immediato. Dato che la stampa degli elementi della coda é una operazione immediata che viene effettuata esattamente  $|V|$  volte, si ha che il tempo di esecuzione complessivo é asintoticamente pari a  $O(|V|) + O(|V| + |E|) = O(|V| + |E|)$ .