Università di Padova - DEI
Corso di Fondamenti di Informatica II, Anno Accademico 2002-2003

SUBSET-SUM

Innanzitutto, una definizione informale. Dato un insieme finito S di numeri interi e un ulteriore intero $t \in \mathbb{N}$, SUBSET-SUM è il problema di rispondere alla domanda: "Esiste un sottoinsieme $S' \subseteq S$ la cui somma è pari a t?".

Esempio: se $S = \{1, 27, 58, 116, 532, 1274\}$ e $t = 175$ allora la risposta è si; una possibile soluzione è $S' = \{1, 58, 116\}$.

Una definizione formale del problema è la seguente.

SUBSET-SUM
$I = \{< S, t > | S \subseteq \mathbb{N}$ finito, $t \in \mathbb{N}\}$
$S = \{$si,no$\}$
D: esiste $S' \subseteq S$ tale che $\Sigma_{s \in S'} s = t$?

Vogliamo ora provare che SUBSET-SUM è NP-completo. Come usuale, la dimostrazione prevede due punti.

1. Provare che SUBSET-SUM è in NP: questa prova è semplice ed è lasciata come esercizio (suggerimento: si utilizzi S' come certificato polinomiale).

2. Provare che SUBSET-SUM è NP-hard: questo punto è invece non banale e verrà ora illustrato in dettaglio.

La prova che SUBSET-SUM è NP-hard si basa su una riduzione da VERTEX-COVER. Si noti che per provare VERTEX-COVER$<_P$SUBSET-SUM è necessario "trasformare un grafo in un insieme di interi"! Come si può fare? L'idea è di associare ad ogni nodo ed arco del grafo un numero opportuno in modo che, nel loro complesso, i numeri incorporino la descrizione del grafo fornita dalla *matrice di adiacenza nodi-archi* B:

$$B = \begin{pmatrix} B_{0,|E|-1} & B_{0,|E|-2} & \cdots & B_{0,0} \\ B_{1,|E|-1} & B_{1,|E|-2} & \cdots & B_{1,0} \\ \vdots & \vdots & \ddots & \vdots \\ B_{|V|-1,|E|-1} & B_{|V|-1,|E|-2} & \cdots & B_{|V|-1,0} \end{pmatrix}.$$

Dato $G = (V, E)$, $B_{i,j} = 1 \iff v_i \in V$ è estremo di $e_j \in E$. Procediamo ora all'illustrazione formale della riduzione definendo la funzione $f : \{0,1\}^* \to \{0,1\}^*$ che trasforma una istanza di VERTEX-COVER in una istanza di SUBSET-SUM.

$$\begin{array}{lll} x =< G, k > & \to & f(x) =< S, t > \\ G = (V, E) & & S = \{x_0, x_1, \ldots, x_{|V|-1}, y_0, y_1, \ldots, y_{|E|-1}\} \\ V = \{v_0, v_1, \ldots, v_{|V|-1}\} & & x_i = 4^{|E|} + \Sigma_{j=0}^{|E|-1} B_{i,j} 4^j \\ E = \{e_0, e_1, \ldots, e_{|E|-1}\} & & y_i = 4^i \\ & & t = k4^{|E|} + 2\Sigma_{j=0}^{|E|-1} 4^j \end{array}$$

In altre parole, l'insieme S contiene un numero per ciascun nodo in V e un numero per ciascun arco in E, per un totale di $|V| + |E|$ elementi. La funzione f è computabile in tempo $O(|V| \cdot |E|) = O\left(|E|^2\right)$, quindi polinomiale nella taglia dell'input.

Per completare la riduzione rimane da provare che

$$x \in \text{VERTEX-COVER} \iff f(x) \in \text{SUBSET-SUM}.$$

Per far ciò, determiniamo le proprietà dei numeri x_i e y_i introducendone una particolare rappresentazione "in due pezzi". La rappresentazione è descritta dalla funzione $\rho_E(i)$ che segue.

$$i \in \mathbb{N} \to \rho_E(i) = (a, b) = \left(i\text{DIV}4^{|E|}, (i\text{MOD}4^{|E|})_4\right).$$

Con $(\cdot)_4$ si vuole indicare la rappresentazione in base 4 di un numero. Si noti che a costituisce il modulo della divisione di i per $4^{|E|}$, e b costituisce il resto della medesima divisione: di conseguenza, si ha che $i = a4^{|E|} + b$.

Esempio: $\rho_E(6515) = (101, (51)_4) = (101, 303)$.

Proprietà 1. $i_1 = i_2 \iff \rho_e(i_1) = \rho_e(i_2)\ \forall i_1, i_2 \in \mathbb{N}$.

Dimostrazione. L'implicazione "⇒" è banale. Per quanto riguarda l'implicazione "⇐", se $\rho_E(i_1) = \rho_E(i_2)$ allora $a_1 = a_2$ e $b_1 = b_2$, pertanto

$$a_1 4^{|E|} + b_1 = a_2 4^{|E|} + b_2 \Rightarrow i_1 = i_2.$$

□

Vediamo ora quali sono le rappresentazioni "in due pezzi" dei numeri x_i, y_i e t ottenuti mediante la funzione $f(\cdot)$.

$$\rho_E(x_i) = (1, B_{i,|E|-1} B_{i,|E|-2} \cdots B_{i,0}), 0 \le i \le |V| - 1$$
$$\rho_E(y_i) = (0, 00 \cdots 01 \underbrace{0 \cdots 0}_{i-1 \text{ volte}}), 0 \le i \le |E| - 1$$
$$\rho_E(t) = (k, 22 \cdots 22)$$

Sulla base di queste rappresentazione le seguenti proprietà possono essere facilmente provate.

1. La prima componente della rappresentazione degli x_i è 1, e la prima componente della rappresentazione degli y_i è 0: è quindi possibile distinguere i numeri associati a nodi da quelli associati ad archi.

2. La rappresentazione dei numeri associati ai nodi "contiene" la matrice di adiacenza nodi-archi B: più precisamente, la rappresentazione di x_i contiene l'i-esima riga di B.

3. Nella cifra/colonna j della componente b della rappresentazione compaiono al più tre "1" (due dovuti ai nodi che sono estremi di e_j e uno dovuto a y_j): di conseguenza, sommando le rappresentazioni di più numeri non si può generare riporto in *alcuna* cifra. La rappresentazione della somma di più numeri si ottiene perciò sommando cifra per cifra le rappresentazioni degli addendi.

Servendosi della funzione ρ_E e delle sue proprietà, è ora possibile provare che la funzione f consente la riduzione voluta da VERTEX-COVER a SUBSET-SUM.

Teorema 1. $x =< G = (V, E), k >\in$ *VERTEX-COVER* $\iff f(x) =< S, t >\in$ *SUBSET-SUM.*

Dimostrazione. [⇒] Sia $V' = \{v_{i_1}, \ldots, v_{i_k}\}$ un vertex cover di taglia k per G. Ebbene, l'insieme S' che risolve SUBSET-SUM è

$$S' = \{x_{i_1}, \ldots, x_{i_k}\} \cup \{y_j | e_j \text{ha un solo nodo di } V' \text{ come vertice}\}.$$

Proveremo infatti che

$$\rho_E\left(\Sigma_{s \in S'} s\right) = \rho_E(t) = (k, 22 \cdots 2),$$

il che implica immediatamente $\Sigma_{s\in S'} s = t$. Si considerino separatamente le due componenti della rappresentazione di $\rho_E(\Sigma_{s\in S'} s)$. Per quanto riguarda la prima componente, essa vale k come desiderato in quanto vengono sommati k contributi pari a 1 dovuti a $x_{i_1}, \ldots, x_{i_k}$. Per quanto riguarda la seconda componente, si consideri la generica cifra j, $0 \leq j \leq |E| - 1$. È necessario distinguere due casi.

1. Se y_j appartiene all'insieme S' allora l'arco e_j ad esso corrispondente incide su un solo vertice $v_l \in V'$: di conseguenza, tra i numeri in S' solo x_l e y_j hanno la j-esima componente diversa da zero, e la j-esima cifra della somma è pari a 2.

2. Se y_j non appartiene all'insieme S' allora l'arco e_j ad esso corrispondente incide su due nodi in V' (non può accadere che non incida su alcun vertice perché V' è un vertex cover): i due numeri associati a questi nodi fanno sì che la j-esima cifra della somma sia ancora pari a 2.

Riassumendo, tutte le cifre della seconda componente di $\rho_E(\Sigma_{s\in S'} s)$ sono uguali a 2 come desiderato. Questo conclude la prova dell'implicazione "⇒".

[⇐] Sia S' l'insieme di numeri che risolve SUBSET-SUM: allora

$$\rho_E\left(\Sigma_{s\in S'} s\right) = \rho_E(t) = (k, 22\cdots 2).$$

In particolare, poiché la prima componente di $\rho_E(\Sigma_{s\in S'} s)$ è k l'insieme S' contiene *esattamente* k numeri $\{x_{i_1}, \ldots, x_{i_k}\}$ corrispondenti a vertici. Ebbene, l'insieme

$$V' = \{v_{i_1}, \ldots, v_{i_k}\}$$

è un vertex cover di taglia k per G. Ciò è garantito dal fatto che tutte le cifre della seconda componente di $\rho_E(\Sigma_{s\in S'} s)$ sono uguali a 2: poiché gli y_i possono contribuire al più un "1" a ciascuna cifra, l'altro "1" è dovuto ad un numero in $\{x_{i_1}, \ldots, x_{i_k}\}$ e quindi tutti gli archi sono "coperti" da almeno un vertice di V'. $\square$