Subset Sum

"Given a finite set S of positive integers and an integer t, is there a subset of S that sums to t?"

- Clearly this is in NP
- We will prove this is NP-Hard by giving a reduction from VC. The text uses 3SAT, showing the reduction from VC for variety.
- Consider an undirected graph $G = (V, E)$ for which we want to find a vertex cover of size $\leq k$. We generate an instance of subset sum as follows
  - Arbitrarily number the m edges in E as $e_0, e_1, \ldots, e_{m-1}$. For each edge e_i define $b_i = 4^i$

- For each vertex $v \in V$, let $\Delta(v)$ be the set of edges that have v as an endpoint

For each v define

$$a_i = 4^m + \sum_{i \in \Delta(v)} 4^i$$

- Let set $S = \{a_v \mid v \in V\} \cup \{b_i \mid i \in E\}$

- Let the target sum be

$$t = k \cdot 4^m + \sum_{i=0}^{m-1} 2 \cdot 4^i$$

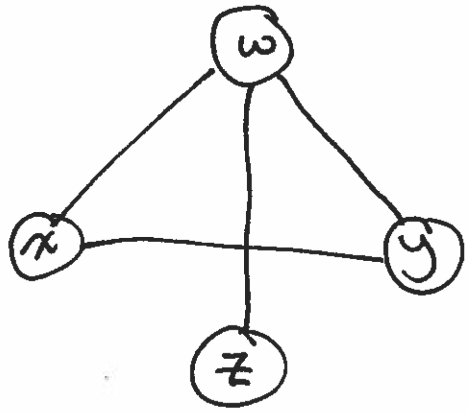
It is useful to think of the elements of S as $(m+1)$-digit numbers written in base 4.

The reason for using base 4 will become clear later.

For each $i < m$, the i^{th} digit represents the preseace of absence of edge i

The m^{th} digit is 1 if the number is an a value and 0 if it is a b value.

Example:



$k = 2$

$V = \{w, x, y, z\}$

$E = \{ e_0 : (w, z)$
$e_1 : (w, y)$
$e_2 : (y, x)$
$e_3 : (x, w)$

The set S would have

$a_w = 11011$ $\quad$ $b_0 = 00001$

$a_x = 11100$ $\quad$ $b_1 = 00010$

$a_y = 10110$ $\quad$ $b_2 = 00100$

$a_z = 10001$ $\quad$ $b_3 = 01000$

$t = 22222$

all in base 4

To prove that this is really a reduction

- Suppose there is a vertex cover V' of size $\leq k$ in G
- We can add vertices to the cover to make its size exactly K.
- Now to pick a subset $S_c \subseteq S$ that sums to t.
    - For every $v \in V'$, include a_v in S_c. Consider the resulting sum written in base 4. It will have a 1 or a 2 in each of the low order m digits depending on whether a given edge is covered by one or two vertices in V'
    - To obtain a two in each of these digits we also include b_i in S_c for each edge i that has exactly one incident vertex in V'

- Finally since each a_v value in S_c contributes a 1 in the $(m+1)$ position, and the b_i values are always 0 in that position, the sum contains k in the $(m+1)^{st}$ digit

of course if $k > 3$ its representation will spill over into higher digits.

Conversely, suppose there is a subset $S' \subseteq S$ summing to t. That is, there exists subsets $V' \subseteq V$ and $E' \subseteq E$ such that

$$\sum_{v \in V'} a_v + \sum_{i \in E'} b_i = t$$

- If we sum these base-4 numbers there are no carries from the low order m digits because for each i there are at most three numbers in S whose i^{th} digit is 1. Namely b_i and the a_v for the vertice incident on edge i

- In ~~t~~ each of these digits is a two
  - Since b_i can only contribute a single 1 at least 1 must come from an a_v incident on edge i.
  - In other words, for each edge, at least one of its endpoints must be in V', i.e. V' is a VC
- Since only the a_v can contribute to the $(m+1)^{st}$ digit (and higher), and these equal k, so V' has k elements.