CS3933 Winter 2007 – Solo Assignment 5
Due Thursday April 12, 2007, at 5pm.
Sample Solution

1. (5 marks) Prove that Committee Formation, as described in Group Assignment 5, is in NP by giving a polynomial time certificate-checking algorithm for it.

   Input: P, $S = \{S_i\}$, k
   Certificate: candidate committee P'

```
valid ← true
if P' ⊈ P
  valid ← false
else
  if |P'| < k
    valid ← false
  for all S_i ∈ S
    count ← 0
    for all x ∈ S_i
        if x ∈ P'
          count ← count + 1
    if count > 1
        valid ← false
return valid
```

   Time analysis: checking $P' \not\subseteq P$ can be done in $O(|P|^2)$ time. Measuring $|P'|$ (subsequently) will take $O(|P|)$ time. After these steps, the nested for loops together iterate through every element in every set (linear in the input size n). Each iteration includes a membership test, which can be computed in $O(|P|)$ time. All other steps take constant time.
   Therefore, the overall running time is in $O(|P|^2 + n \cdot |P|)$, which is polynomial in the input size.

   Thus Committee Formation $\in NP$.

2. Consider the following set problem, known as Hitting Set.
   Given: a collection $\mathcal{C} = \{C_i\}$ where each C_i is a set; and an integer k.
   Question: Is there a set S, where $|S| \leq k$, such that S intersects all members of $\mathcal{C}$?

$$\forall C_i \in \mathcal{C} \quad S \cap C_i \neq \emptyset$$

Prove that HITTING SET is NP-complete, using the following steps:

(a) (5 marks) Show that HITTING SET is in NP.

```
Input: 𝒞 = {C_i}, k
Certificate: candidate set S
    valid ← true
    if |S| > k
      valid ← false
    else
      for all C_i ∈ 𝒞
          found ← false
          for all x ∈ C_i
              if x ∈ S
                found ← true
          if not found
              valid ← false
    return valid
```

Time analysis: Testing to determine if $|S| > k$ can be done in $O(k)$ time.
The nested for loops together iterate through all elements of all sets in $\mathcal{C}$, so they iterate n times (where n is the sum of the C_i set sizes). Each iteration contains a membership test $x \in S$, which can be computed in $O(k)$ time (since $|S| \leq k$). All other steps take constant time, so the overall running time is $\in O(nk)$, which is polynomial in the input size.

Therefore HITTING SET $\in NP$.

(b) (7 marks) Show that the graph problem VERTEX COVER $\leq_p$ HITTING SET.

Reduction: Let $(G = (V, E), k)$ be any instance of VERTEX COVER.

Construct $(\mathcal{C} = \{C_i\}, k_{HS})$, an instance of HITTING SET, by:

Let $\mathcal{C} = \{\{u, v\} \mid (u, v) \in E\}$ (so essentially $\mathcal{C} = E$).

Let $k_{HS} = k$.

Time Analysis:
This construction simply copies data from one instance to the other, so it is linear (and therefore polynomial) in the input size.

We need to show that $(G = (V, E), k)$ is a yes-instance of VERTEX COVER if and only if $(\mathcal{C} = \{C_i\}, k_{HS})$ is a yes-instance of HITTING SET.

$\Rightarrow$: Let $(G = (V, E), k)$ be any yes-instance of VERTEX COVER. Then there is a $V' \subseteq V$ with $|V'| \leq k$, such that every edge has at least one endpoint in V'.
Since each $C_i \in \mathcal{C}$ is the set of endpoints for an edge, then each $C_i \in \mathcal{C}$ has at least one element in V'. $|V'| \leq k = k_{HS}$, so V' is a hitting set of size at most k_{HS} in $\mathcal{C}$.
Therefore $(\mathcal{C}, k_{HS})$ is a yes-instance of HITTING SET.

$\Leftarrow$: Let $(\mathcal{C} = \{C_i\}, k_{HS})$ be any constructed instance of HITTING SET. Then there is some S with $|S| \leq k_{HS}$ such that every $C_i \in \mathcal{C}$ contains at least one element of S. Since each edge from $G = (V, E)$ has a C_i that equals the set of its endpoints, each edge has an endpoint in S. Therefore S is a vertex cover in G of size at most $k_{HS} = k$, so (G, k) is a yes-instance of VERTEX COVER.

Therefore VERTEX COVER $\leq_p$ HITTING SET.

3. (8 marks) Show that HITTING SET $\leq_p$ SET COVER.

Reduction: Let $(\mathcal{C} = \{C_i\}, k)$ be any instance of HITTING SET.

Construct $(S, \mathcal{S} = \{S_j\}, k_{SC})$, an instance of SET COVER, by:

Let $S = \{i \mid C_i \in \mathcal{C}\}$.

Let $C = \cup_{C_i \in \mathcal{C}} C_i$.

Let $\mathcal{S} = \{S_j \mid j \in C\}$, where each $S_j = \{i \mid j \in C_i\}$.

Let $k_{SC} = k$.

Time analysis:
S is constructed by counting $|\mathcal{C}|$, which is linear, as is copying k to k_{SC}. C is constructed by finding the union of all $C_i \in \mathcal{C}$, which can be done in $O(n \log n)$ time (where n is the total input size) by appending all C_i sets, sorting the result, and removing duplicates. $\mathcal{S}$ is then constructed by, for each C_i, adding i to S_j for each $j \in C_i$; this part of the construction would take at worst $O(n^2)$ time. The overall construction time is thus in $O(n^2)$, which is polynomial in the input size.

We need to show that $(\mathcal{C}, k)$ is a yes-instance of HITTING SET if and only if $(S, \mathcal{S}, k_{SC})$ is a yes-instance of SET COVER.

$\Rightarrow$: Let $(\mathcal{C}, k)$ be any yes-instance of HITTING SET. Then there is a set X, $|X| \leq k$, such that all $C_i \in \mathcal{C}$ contain at least one element of X. Since $\mathcal{S}$ is constructed to include one S_j for each possible element of any C_i, we can consider $\mathcal{S}' = \{S_j | j \in X\}$, where $|\mathcal{S}'| \leq k = k_{SC}$.

$$\cup_{j \in X} S_j \;=\; \{i \mid \exists j \in X \text{ where } i \in S_j \;\},$$

which, since X has an element of every C_i, equals $\{i | C_i \in \mathcal{C}\} = S$. Therefore $\mathcal{S}'$ covers all elements of S, so $(S, \mathcal{S}, k_{SC})$ is a yes-instance of SET COVER.

$\Leftarrow$: Let $(S, \mathcal{S}, k_{SC})$ be any constructed yes-instance of SET COVER. Then there is a collection $\mathcal{S}' \subseteq \mathcal{S}$, $|\mathcal{S}'| \leq k_{SC}$, where $\cup_{S_j \in \mathcal{S}'} S_j = S$. Since each S_j corresponds to some element $j \in C$, in selecting at most k_{SC} subsets for $\mathcal{S}'$ we are selecting the corresponding $k = k_{SC}$ elements $C' \subseteq C$, with $j \in C'$ if and only if $S_j \in \mathcal{S}'$.
Furthermore, since $S = \{i | C_i \in \mathcal{C}\}$ and $S_j = \{i | j \in C_i\}$,

$$\cup_{j \in C'} \{i | j \in C_i\} \;=\; \{i | C_i \in \mathcal{C}\}$$

which means that C' is a hitting set of size k for $\mathcal{C}$.
Therefore, $(\mathcal{C}, k)$ is a yes-instance of HITTING SET.

Thus HITTING SET $\leq_p$ SET COVER.