CS314: Algorithms
Homework 7

1. Consider the following problem: Given as input two graphs H and G, is H isomorphic to to some subgraph of G? Prove that this problem is NP-Complete.

2. Consider a set $A = \{a_1, \dots, a_n\}$ and a collection $B_1, \dots, B_m$ of subsets of A (so that each $B_i \subseteq A$).

   We say that H is a *hitting set* for the collection $B_1, \dots, B_m$ if H contains at least one element from each B_i, so that $H \cap B_i$ is not empty for each i. In other words, H "hits" each of the sets B_i.

   We now ask the following: Given a set A and a collection of subsets B_i as described above, and a number k, is there a hitting set $H \subseteq A$ of size at most k? Prove that this problem is NP-Complete.

3. A store trying to analyze the behavior of its customers will often maintain a 2-dimensional array A, where the rows correspond to its customers and the columns corespond to the products it sells. The entry $A[i, j]$ specifies the quantity of product j that has been purchased by customer i.

   Here's a tiny example of such an array A:

	detergent	beer	diapers	cat litter
Alice	0	6	0	3
Bob	2	3	0	0
Eve	0	0	1	3

   One thing the store might want to do with this data is the following. We say that a subset S of the customers is *diverse* if no two of the customers have ever bought the same produce (i.e. for each product, at most one person in S has ever bought it). A diverse set of customers can be useful, for example, as a target pool for market research.

   We can now define the diverse subset problem as follows: Given an $m \times n$ array A defined as above, and a number $k \leq m$, is there a subset of at least k customers which is diverse?

   Show that the diverse subset problem is NP-Hard.

4. **Extra Credit:** A boolean formula is in disjunctive normal form (or DNF) if it consists of a disjunction (OR) or several terms, each of which is the conjunction (AND) of one or more literals. For example, the formula:

$$(\overline{x} \wedge y \wedge \overline{z}) \vee (y \wedge z) \vee (x \wedge \overline{y} \wedge \overline{z})$$

   is in disjunctive normal form. DNF-SAT asks, given a boolean formula in disjunctive normal form, whether that formula is satisable.

   (a) Describe a polynomial-time algorithm to solve DNF-SAT.

(b) What is the error in the following argument that P=NP?

Suppose we are given a boolean formula in conjunctive normal form with at most three literals per clause, and we want to know if it is satisable. We can use the distributive law to construct an equivalent formula in disjunctive normal form. For example,

$$(x \vee y \vee \overline{z}) \wedge (\overline{x} \vee \overline{y}) \iff (x \wedge \overline{y}) \vee (y \wedge \overline{x}) \vee (\overline{z} \wedge \overline{x}) \vee (\overline{z} \wedge \overline{y})$$

Now we can use the algorithm from part (a) to determine, in polynomial time, whether the resulting DNF formula is satisable. We have just solved 3SAT in polynomial time. Since 3SAT is NP-hard, we must conclude that P=NP!