A reduction of CLIQUE to MAX 2-SAT

(This note is a lightly edited version of my response to a homework problem in Dr. Lorna Stewart's CMPUT 304 (Algorithms II) course in Fall 2007. I presented this reduction — with considerably more discussion of the thought process leading to it — in class on November 29.)

We wish to show that the following problem is NP-complete:

MAX 2-SAT
Instance: Set U of variables, collection C of clauses over U such that each clause $c \in C$ has $|c| = 2$, and positive integer k.
Question: Is there a truth assignment for U that simultaneously satisfies at least k of the clauses in C?

MAX 2-SAT is in NP for the usual reason with satisfiability problems: a truth assignment such as described serves as a certificate, and it can be verified simply by evaluating the clauses for that truth assignment (and, for this problem, counting how many are true).

To show that MAX 2-SAT is NP-hard, we will reduce from CLIQUE, the following well-known NP-complete problem:

CLIQUE
Instance: Graph G and positive integer k.
Question: Is there a clique in G of size at least k?

(A clique is a set of vertices, every two of which are adjacent.)

Conversion of CLIQUE instances to MAX 2-SAT instances

Consider a graph $G = (V, E)$ and positive integer j, forming an instance of CLIQUE; the problem is to determine whether G contains a clique with at least j vertices. Let $|V| = n$ and $V = \{v_1, v_2, \ldots, v_n\}$. Also let

$$\bar{E} = \{(v_i, v_j) : i < j \text{ and } (v_i, v_j) \notin E\}$$

be, if you will, the set of non-edges in G. Note that $|\bar{E}| \leq \binom{|V|}{2} \leq |V|^2$.

Construct a MAX 2-SAT instance consisting of:

One dummy variable z.
For each vertex v_i, one variable x_i,
and clauses $(x_i \vee z)$
and $(x_i \vee \neg z)$.
For each non-edge $(v_i, v_j) \in \bar{E}$, one clause $(\neg x_i \vee \neg x_j)$.
The number $k = |V| + j + |\bar{E}|$.

(Informal remarks: Think of a game in which you choose some set of vertices in G, then get points for how well you did it. The clauses are the rules of a game: each states a condition, satisfying which earns you a point. We wish to set up the rules so that choosing a large clique maximizes one's score. The "vertex clauses" $(x_i \vee z)$, $(x_i \vee \neg z)$ encourage choosing many vertices; the "non-edge clauses" $(\neg x_i \vee \neg x_j)$ encourage choosing adjacent vertices.)

Efficiency of the conversion

There are $|V|+1$ variables to construct, and constructing each takes constant time.

Computing $\bar{E}$ even by a very naïve method is polynomial time. For example, we can construct all pairs (v_i, v_j) with $i < j$ and then for each one, scan E to determine whether it contains that edge. There are $O(|V|^2)$ pairs to consider, and scanning E takes $O(|E|)$ time, and so the whole process takes $O(|E||V|^2)$ time.

The clauses for the non-edges can be made at the same time $\bar{E}$ is computed.

Clique assignments and how many clauses they satisfy

Before we get to the main results, a wee computation.

Definition: A *clique assignment* is a truth assignment to the variables of our MAX 2-SAT instance such that the set of vertices $\{v_i : x_i = \text{true}\}$ is a clique.

Lemma: A clique assignment giving rise to clique V' satisfies $|V| + |V'| + |\bar{E}|$ clauses.

Proof of lemma: First, note that the set of clauses is, by construction, symmetrical in z and $\neg z$; thus we may assume without loss of generality that the clique assignment has $z = \text{true}$. (If not, exchange z and $\neg z$ in what follows.)

Now, all the clauses $(x_i \vee z)$ are satisfied, since z is true. (That's $|V|$ clauses satisfied.) The clause $(x_i \vee \neg z)$ is satisfied for all $v_i \in V'$. (That's another $|V'|$ clauses satisfied.)

Now suppose $(\neg x_i \vee \neg x_j)$ is not satisfied, that is, $(\neg x_i \vee \neg x_j) = \text{false}$. Equivalently, $(x_i \wedge x_j) = \text{true}$. Thus x_i and x_j are both true. But then v_i and v_j are both in V', which is a clique. Thus $(v_i, v_j) \in E$, and so $(v_i, v_j) \notin \bar{E}$, whence the clause $(\neg x_i \vee \neg x_j)$ is not in our instance of MAX 2-SAT.

By contraposition, every clause of the form $(\neg x_i \vee \neg x_j)$ that *is* in our instance *is* satisfied. (That's another $|\bar{E}|$ clauses satisfied.)

So the clique assignment satisfies $|V| + |V'| + |\bar{E}|$ clauses, as claimed.

If the CLIQUE instance is yes, then the MAX 2-SAT instance is yes

Suppose there is a clique V' of size j in G. Consider the truth assignment given by setting $z = \text{true}$, and setting $x_i = (v_i \in V')$. By the previous section, this truth assignment satisfies $|V| + j + |\bar{E}| = k$ clauses, so the MAX 2-SAT instance is yes.

If the MAX 2-SAT instance is yes, then the CLIQUE instance is yes

Suppose there is a truth assignment (not necessarily a clique assignment) satisfying at least $k = |V| + j + |\bar{E}|$ of the clauses. Let $V' = \{v_i : x_i = \text{true}\}$.

Now we alter the truth assignment so that it becomes a clique assignment satisfying at least as many clauses. If V' is a clique, we are done. If not, there is a pair of vertices in V' with no edge joining them. Let v_i and v_j, with $i < j$, be such a pair.

Change x_i to false. Before this change, both the clauses $(x_i \vee z), (x_i \vee \neg z)$ were satisfied, because x_i was true; after the change, only one of them is. (Which one depends on the assignment to z; we don't care.) That's one clause lost. On the other hand, before this change, the clause $(\neg x_i \vee \neg x_j)$ was not satisfied: both x_i and x_j were true. After the change, with x_i false, this clause *is* satisfied. That's one clause gained. There might be other clauses involving x_i of the non-edge type; by the same argument, those clauses are satisfied after the change, so the number of such satisfied clauses is not reduced by the change.

The net effect of this change is, then, not to reduce the number of satisfied clauses. The change does, however, reduce the number of pairs (v_i, v_j) of vertices in V' with no edge, by at least one. Thus repeating this procedure results eventually in a truth assignment which satisfies at least as many clauses as the original assignment, that is, at least $|V| + j + |\bar{E}|$ clauses, but in which every pair of vertices in V' is joined by an edge, that is, a clique assignment. As previously computed, such an assignment satisfies $|V| + |V'| + |\bar{E}|$ clauses. Therefore $|V'| \geq j$, that is, the clique V' has size at least j; its existence establishes that the CLIQUE instance is yes.