**NP-Completeness of Independent Set**

Let $G = (V, E)$ be a simple undirected graph. We say a subset $I \subseteq V$ in $G$ is an independent set if no two vertices in $I$ are connected by an edge of $G$. That is, for any $u, v \in V$ we have $(u, v) \notin E$. An independent set is maximal if $I$ cannot be expanded further; that is, there exists no vertex $w \in V - I$ such that $I \cup \{w\}$ is also an independent set. The problem of finding a maximal independent set of maximum cardinality is a hard problem.

The **INDEPENDENT SET (IS) decision problem** is the following.

**INSTANCE**: Given a graph $G = (V, E)$ and a lower bound $k$, where $1 \leq k \leq |V|$.

**QUESTION**: Does $G$ contain an independent set of cardinality $k$?

**Theorem**: IS is NP-complete.

**Proof**: For each vertex in $I$ we check every edge incident to it in $G$ to see if that edge connects the vertex to at least one other vertex in $I$. If we ever find such an edge, we reject. Otherwise we accept $I$ as the independent set of the graph $G$. The overall algorithm runs in polynomial time, so IS is in NP.

We use reduction from 3SAT to IS. Let $E = e_1 \land e_2 \land e_3 \land ... \land e_m$ be a CNF expression where each clause $e_i$ has three literals. Based on $E$ we construct a graph $G$ with $3m$ vertices, labeled $[i, j]$, $1 \leq i \leq m$, $j = 1, 2, 3$. The vertex $[i, j]$ represents the $j$th literal in the clause $e_i$.

For example, consider $E = (x_1 + x_2 + x_3)(\overline{x_1} + x_2 + x_4)(\overline{x_2} + x_3 + x_5)(\overline{x_3} + \overline{x_4} + \overline{x_5})$. The graph $G$ is given below. The columns correspond to the clauses.

![Graph](image)

There following are the two ideas in forming the edges.

1. We want to make sure that only one vertex corresponding to a given clause is chosen in building an independent set. We do this by putting edges between all pairs of vertices in a column.
2. If two vertices represent complementary vertices, then in the selection of independent set we want only one of the vertices. Thus if \([i_1, j_1]\) represents \(x\) and \([i_2, j_2]\) represents \(\overline{x}\), we place an edge between them.

The bound \(k\) for the graph \(G\) constructed by these two rules is taken as \(m\).

It is not hard to see how \(G\) can be constructed from \(E\) in time that is proportional to the square of the length of \(E\). The claim is that this correctly reduces 3SAT to IS. That is:

\[E\text{ is satisfiable if and only if } G\text{ has an independent set of size } m.\]

We first assume that \(G\) has an independent set of size \(m\). In each column there can be only one vertex in the independent set. \(I\) must include exactly one vertex from each clause if its size is \(m\). Also the set cannot have vertices representing a variable \(x\) and its negation \(\overline{x}\) since all \(x\) and \(\overline{x}\) vertices are edge-connected. Thus the independent set \(I\) of size \(m\) yields a satisfying truth assignment \(\tau\) for \(E\) as follows. If a vertex equivalent to a variable \(x\) is in \(I\), make \(\tau(x) = 1\); if a vertex equivalent to a negated variable \(\overline{x}\) is in \(I\), choose \(\tau(x) = 0\). For vertices equivalent to variables not in \(I\), pick \(\tau(x)\) arbitrarily.

We now assume \(E\) is satisfied by a truth assignment \(\tau\). Since \(\tau\) makes each clause of \(E\) true, we can identify one literal from each clause that \(\tau\) makes true. For some clauses we may have a choice - we then arbitrarily pick one. Construct a set of \(m\) vertices \(I\) by picking the vertex corresponding to the selected literal from each clause.

We claim that \(I\) is an independent set. Only one vertex from each clause implies one vertex from each column which again implies a potential independent set. An edge connecting a variable and its negation cannot be both in \(I\) as we have selected only vertices that correspond to literals made true by the truth assignment \(\tau\). Of course, \(\tau\) will make one of \(x\) and \(\overline{x}\) true but never both. Thus if \(E\) is satisfiable, then \(G\) has an independent set of size \(m\).

Thus, there is a polynomial time reduction from 3SAT to IS. Since 3SAT is known to be NP-complete, so is IS.

The following two corollaries are immediate from the above theorem.

**Corollary 1** : The VERTEX COVER problem is NP-complete.

A vertex cover of a simple undirected graph \(G = (V, E)\) is a set of vertices such that each edge has at least one of its ends at a vertex of the set. That is, a vertex cover of \(G\) is a subset \(V' \subseteq V\) such that if \((u, v)\) is an edge of \(G\), then either \(u \in V'\) or \(v \in V'\) (or both). A vertex cover is minimal if \(V'\) cannot be contracted further; that is, there exists no vertex \(w \in V'\) such that \(V' - \{w\}\) is also a vertex cover. Finding a minimal vertex cover of minimum cardinality is a hard problem.

Vertex covers and independent sets are closely related. In fact, the complement of an independent set is a vertex cover, and vice-versa. That is:

\(I\) is an independent set of graph \(G\) if and only if \(V - I\) is a vertex cover of the same graph.

In the above figure, the set of vertices that do not have boldface outlines form a vertex cover.
Corollary 2: The CLIQUE problem is NP-complete.

A clique in a simple undirected graph $G = (V, E)$ is a subset $C \subseteq V$ of vertices, each pair of which is connected by an edge in $E$. In other words, a clique is a complete subgraph of $G$. The size of a clique is the number of vertices it contains; that is, $|C|$. The CLIQUE problem is to find a clique of maximum size in a graph $G$.

It is easy to see that the independent set of a graph $G$ is precisely the clique in the complementary graph $G'$ of $G$, the graph that contains all edges that are not in $G$; and vice-versa. That is:

$I$ is an independent set of graph $G$ if and only if $I$ is a clique of the complementary graph $G'$.

Reference