Lecture 15. NP-complete languages.

A literal is a Boolean variable or a negated Boolean variable, for example: $x_1, x_{15}, \bar{x}_{23}, z_{10}$.

A clause is several literals connected with logical OR, for example: $(x_1 \lor \bar{x}_2 \lor x_{15} \lor x_{23})$.

A Boolean formula is in conjunctive normal form, or a CNF-formula, if it consists of several clauses connected with logical AND, for example:

$$(x_1 \lor \bar{x}_2 \lor x_{15} \lor \bar{x}_{23}) \land (\bar{x}_1 \lor \bar{x}_2 \lor x_7) \land (x_3 \lor \bar{x}_{20}).$$

A Boolean formula is a 3-CNF-formula if it is a CNF-formula and all clauses have three literals, for example:

$$(x_1 \lor \bar{x}_2 \lor \bar{x}_3) \land (x_1 \lor x_2 \lor x_{15}) \land (x_2 \lor x_3 \lor \bar{x}_{11}).$$

Define languages:

$$\text{SAT} = \{ \langle \phi \rangle \mid \phi \text{ is a satisfiable CNF-formula} \},$$

$$3\text{-SAT} = \{ \langle \phi \rangle \mid \phi \text{ is a satisfiable 3-CNF-formula} \}.$$

**Definition.** A language $L$ called NP-complete, if it satisfies two conditions:

1. $L \in \text{NP}$;

2. $L$ is NP-hard, i.e. any problem $L' \in \text{NP}$ is polynomial-time reducible to $L$. 
From this definition it follows that if \( L \) is NP-complete and \( L \in P \) then \( P = NP \).

**Theorem.** If \( L \) is NP-complete and for some \( L' \in NP \) it holds that \( L \leq_P L' \), then \( L' \) is NP-complete too.

**Proof.** It is given that \( L' \in NP \). It remains to show that every \( L_0 \in NP \) is polynomial-time reducible to \( L' \). Since \( L \) is NP-complete, \( L_0 \leq_P L \). We have:

\[
L_0 \leq_P L \quad \text{ja} \quad L \leq_P L'.
\]

Therefore from definition of polynomial-time reduction we have that \( L_0 \leq_P L' \). We obtained that every \( L_0 \in NP \) satisfies \( L_0 \leq_P L' \).

**Cooki-Levini teoreem.** SAT is NP-complete.

**Proof idea.** First, showing that SAT \( \in NP \) is easy. To show that SAT is NP-hard, we construct polynomial-time reduction from any language \( A \in NP \) to SAT. The reduction takes a string \( w \) and produces a Boolean formula \( \phi \) that simulates the non-deterministic machine for \( A \) on \( \phi \). If the machine accepts – there is a satisfying assignment. If the machine rejects – there is no such assignment. Hence \( w \in A \) if and only if \( \phi \) is satisfiable.

More details can be found in the book.

**Theorem.** 3-SAT is NP-complete.

**Proof.** First, 3-SAT \( \in NP \), since we can non-deterministically “guess” assignment and verify that it satisfies the given 3-CNF-formula.

Second, let

\[
\phi = (l_{1,1} \lor l_{1,2} \lor \ldots \lor l_{1,i_1}) \land (l_{2,1} \lor l_{2,2} \lor \ldots \lor l_{2,i_2}) \land \ldots \land (l_{k,1} \lor l_{k,2} \lor \ldots \lor l_{k,i_k}),
\]

where \( l_{i,j} \) are literals (can contain “not” sign). We convert each clause into “and” of clauses with 3 literals, as follows:

\[
\varphi_j = (l_{j,1} \lor l_{j,2} \lor \ldots \lor l_{j,i_j})
\]

is converted into

\[
\hat{\varphi}_j = (l_{j,1} \lor l_{j,2} \lor z_1) \land (z_1 \lor l_{j,3} \lor z_2) \land (z_2 \lor l_{j,4} \lor z_3) \land \ldots \land (z_{i_j-3} \lor l_{i_j-1} \lor l_{i_j}),
\]

where \( z_1, z_2, \ldots, z_{i_j-3} \) are new Boolean variables. Then \( f(\phi) = \hat{\varphi}_1 \lor \hat{\varphi}_2 \lor \ldots \lor \hat{\varphi}_k \).

**Correctness.** Prove that \( \phi \in SAT \) if and only if \( f(\phi) \in 3-SAT \). In other words, \( \phi \) is satisfiable if and only if \( f(\phi) \) is satisfiable.
Direction 1) Let $\phi$ be satisfiable, and consider assignment to literals that satisfy $\phi$. Consider clause $\varphi_j = (l_{j,1} \lor l_{j,2} \lor \ldots \lor l_{j,i_j})$. There exists $l_{j,r}$ in that assignment such that $l_{j,r} = \text{TRUE}$. Then we choose assignments to $z_1, \ldots, z_{i_j-3}$ such that all clauses with three variables are satisfied:

\[
\begin{array}{cccc}
T & F & T & T \\
(l_{j,1} \lor l_{j,2} \lor z_1) \land (\bar{z}_1 \lor l_{j,3} \lor z_2) \land (\bar{z}_2 \lor l_{j,4} \lor z_3) \land \ldots \\
F & T & F & T \\
\ldots \land (\bar{z}_{r-2} \lor l_{j,r} \lor z_{r-1}) \land \ldots \land (\bar{z}_{i_j-3} \lor l_{j,i_j-1} \lor l_{i_j}).
\end{array}
\]

Here T and F stand for TRUE and FALSE, respectively.

Direction 2) Let $f(\phi)$ be satisfiable, and consider assignment to $l_{j,r}$ and $z_r$ that satisfy $f(\phi)$. Then, the same assignment to $l_{j,r}$ satisfies $\phi$ because there are $i_j - 2$ clauses with three variables, but only $i_j - 3$ variables $z_r$ can be TRUE. So at least one of the clauses is satisfied with $l_{j,r} = \text{TRUE}$.

Polynomiality. $f(\phi)$ can be constructed from $\phi$ in time polynomial in length of $\phi$. \qed

**Practise session**

1. Prove that the language

$$
\text{CLIQUE} = \{\langle G, k \rangle \mid G \text{ is an undirected graph with a clique of size } k\}
$$

is NP-complete.
Solutions to the practise session questions

1. Prove that the language

\[ \text{CLIQUE} = \{ \langle G, k \rangle \mid G \text{ is an undirected graph with a clique of size } k \} \]

is \text{NP}-complete.

\textit{Solution.} First, \text{CLIQUE} \in \text{NP} as it is possible to “guess” non-deterministically the clique, and then verify that it has size \( k \).

Next, we show that \text{CLIQUE} is \text{NP}-hard.

\textit{Idea.} We use polynomial-time reduction from 3-SAT, which has shown to be \text{NP}-complete in the lecture,

\[ 3\text{-SAT} \leq_p \text{CLIQUE.} \]

Let \( \phi = (l_{1,1} \lor l_{1,2} \lor l_{1,3}) \land (l_{2,1} \lor l_{2,2} \lor l_{2,3}) \land \ldots \land (l_{k,1} \lor l_{k,2} \lor l_{k,3}) \) be a 3-CNF formula, where \( l_{j,r} \) are literals. The reduction \( f \) generates a pair \( (G, k) \), where \( G \) is a graph and \( k \) is an integer.

The nodes of \( G \) are organised in \( k \) groups, each group has three nodes. Each such triple correspond to one of the clauses in \( \phi \), and each node corresponds to a literal in that triple.

More specifically, for clause \( (l_{i,1} \lor l_{i,2} \lor l_{i,3}) \) we have three nodes in graph, \( v_{i,1}, v_{i,2} \) and \( v_{i,3} \). There is an edge between any two nodes, except when

1. two nodes represent literals with contradicting labels, i.e. \( l_{j,r} \) and \( \overline{l_{j,r}} \),

or

2. nodes appear in the same clause.

For example:

\[ \phi = (x_1 \lor x_2 \lor \overline{x_2}) \land (\overline{x_1} \lor \overline{x_2} \lor x_3) \land (x_1 \lor x_2 \lor \overline{x_3}) \]

is transformed into a graph
We have a clique of size 3. The corresponding assignment is $\bar{x}_1 = \text{TRUE}, \bar{x}_2 = \text{TRUE}, \bar{x}_3 = \text{TRUE}$; or $x_1 = x_2 = x_3 = \text{FALSE}$.

**Correctness.** We need to show that $\phi$ has a satisfying assignment if and only if $G$ has a clique of size $k$.

1) Suppose that $\phi$ has a satisfying assignment. Then, at least one literal is TRUE in every clause. We choose nodes that correspond to those literals. (If more than one literal is TRUE in some clause, we choose one of these literals arbitrarily).

The chosen nodes form a clique because of the following:

- they do not have contradictory values (if $x_i$ is chosen then $\bar{x}_i$ is not chosen);
- they represent different clauses.

The size of this clique is $k$.

2) Suppose that $G$ has a clique of size $k$. All nodes of this clique appear in different clauses, because nodes in the same clause are not connected by edges. We assign values to the variables in $\phi$ such that each literal in the clique has value TRUE. This is always possible since any two literals having contradictory values are not connected by an edge. This assignment satisfies $\phi$ because each clause contains a clique node, which has value TRUE. Therefore, $\phi$ is satisfiable.

**Polynomiality.** Construction of $G$ from $\phi$ takes only polynomial number of steps.