Last week we formalized ways of showing that a problem is hard, specifically the definition of $NP$, $NP$-completeness, and $NP$-hardness.

We now discuss ways of saying something useful about these hard problems. Specifically, we show that the answer produced by our algorithm is within a certain fraction of the optimum.

For a maximization problem, suppose now that we have an algorithm $A$ for our problem which, given an instance $I$, returns a solution with value $A(I)$. The approximation ratio of algorithm $A$ is defined to be

$$\max_I \frac{OPT(I)}{A(I)}.$$

This is not restricted to NP-complete problems: we can apply this to matching to get a 2-approximation.

**Lemma 0.1.** The maximal matching has size at least $1/2$ of the optimum.

*Proof.* Let the maximum matching be $M^*$. Consider the process of taking a maximal matching: an edge in $M^*$ can’t be chosen only after one (or both) of its endpoints are picked.

Each edge we add to the maximal matching only has 2 endpoints, so it takes at least

$$\frac{M^*}{2}$$

edges to make all edges in $M^*$ ineligible. \hfill $\square$

This ratio is indeed close to tight: consider a graph that has many length 3 paths, and the maximal matching keeps on taking the middle edges.

We now apply this to the VERTEX-COVER problem. This problem’s decision version is whether there exists a vertex cover of size $\leq k$. It is in $NP$ because a verifier can just check the vertex cover. We now show that it’s NP-complete by reducing from CLIQUE. The following fact shows the close connection between CLIQUE and VERTEX-COVER.
Lemma 0.2. Let $\overline{G}$ be the complement of $G$: $uv$ is connected in $\overline{G}$ if and only if $u$ and $v$ are not connected in $G$.

$G$ has a clique of size $k$ if and only if $\overline{G}$ has a vertex cover of size $|V| - k$.

Proof. We will prove the two directions separately, although the two directions of the proof look very similar: they are based on the fact that $S$ is a clique in $G$ if and only if $V \setminus S$ is an independent set in $\overline{G}$.

For the if direction, Let $S$ be a vertex cover in $\overline{G}$, then we claim $V \setminus S$ is a clique in $G$. The fact that $S$ is a vertex cover means that there are no edges in $\overline{G}$ between two vertices in $V \setminus S$. Since $\overline{E}$ is the complement of $E$, this means all edges are present between edges in $V \setminus S$ in $G$.

The other direction follows similarly. Let $S$ be a clique in $G$, then we can show that $V \setminus S$ is a vertex cover in $\overline{G}$. The fact that $\overline{E}$ is the complement of $E$ means that there are no edges between vertices in $S$ in $\overline{G}$. Therefore $V \setminus S$ is a vertex cover in $\overline{G}$.

Corollary 0.3. $CLIQUE \rightarrow VERTEX\rightarrow COVER$

Note that this Lemma also implies $VERTEX\rightarrow COVER \rightarrow CLIQUE$, which gives an alternate proof that $CLIQUE \in NP$. However, because the proof of a problem is in NP can usually be done in directly (as in this case), we usually only focus on ONE side of the reduction when using such structural results.

We now give an 2-approximate algorithm for $VERTEX\rightarrow COVER$. This problem, unlike matching, is a minimization problem. So we will define the approximation ratio accordingly as:

$$\max_I \frac{A(I)}{OPT(I)}.$$

The algorithm is quite simple:

1. Compute a maximal matching $M$.
2. Return all $2|M|$ endpoints of edges in $M$.

This solution is a vertex cover because if there is an uncovered edge, it can be added to $M$ to form a larger matching.

Furthermore, we still have $|M| \leq OPT$ since each edge must have at least one end point chosen, and the edges in $M$ don’t share endpoints.

This means the size of our solution, $A(I) = 2|M|$, is at most $2OPT$. 

2