CS 4102: Algorithms

Lecture 25: P vs. NP

David Wu Fall 2019

Today's Keywords

Reductions NP-Completeness P vs. NP

CLRS Readings: Chapter 34

Homework

HW9, HW10C due Thursday, December 5, 11pm

- Graphs, Reductions
- Written (LaTeX)

Final Exam

Monday, December 9, 7pm in Olsson 120

- Practice exam coming soon
- Review session likely the weekend before
- SDAC: Please sign-up for a time on December 9

P vs. NP

A language $\mathcal{L} \in \text{NP}$ if there exists a deterministic polynomial-time "verifier" \mathcal{R} such that $x \in \mathcal{L} \Leftrightarrow \exists w \in \{0,1\}^{\text{poly}(|x|)}$: $\mathcal{R}(x,w) = 1$

A language $\mathcal{L} \in P$ if there exists a deterministic polynomial-time "solver" \mathcal{R} such that $x \in \mathcal{L} \Leftrightarrow \mathcal{R}(x) = 1$



If we can decide a problem in polynomial time, we can verify a solution to the problem in polynomial time: $P \subseteq NP$

Biggest open problem in computer science: is this containment strict?

$$P = NP \text{ or } P \neq NP$$

Understanding the Landscape of NP

Question: What are the <u>hard</u> problems in NP?

- Can we systematically characterize these?
- Can we use insights from one problem to help solve another problem?

Strategy: Identify problems at least as "hard" as NP

 If any of these "hard" problems can be solved in polynomial time, then all NP problems can be solved in polynomial time

A problem (or language) B is NP-hard

- $\forall A \in NP, A \leq_p B$
- $A \leq_p B$ means A reduces to B in <u>polynomial</u> time



NP-Hardness



NP-hardness reduction

 $A \leq_p B$: there is a polynomial-time reduction from A to B

7

NP-Hardness



Very powerful: if we can solve even one NP-hard problem in polynomial time, we can solve <u>all</u> of them!

Understanding the Landscape of NP

Question: What are the <u>hardest</u> problems in NP?

- By definition, an efficient algorithm for an NP-hard problem implies an efficient algorithm for <u>every</u> NP problem
- Answer: the ones that are NP-hard (if there are any)

NP-complete = NP \cap **NP-hard**

"Complete" for NP in the sense that a solution to one implies a solution to <u>all</u>

- To show P = NP, just need a <u>single</u> polynomial-time algorithm for a single NP-complete (or NP-hard) problem
- To show P ≠ NP, just need a <u>single</u> lower-bound that some NP problem cannot be solved in polynomial time



Understanding the Landscape of NP

Question: What are the <u>hardest</u> problems in NP?

- By definition, an efficient algorithm for an NP-hard problem implies an efficient algorithm for <u>every</u> NP problem
- Answer: the ones that are NP-hard (if there are any)

NP-complete = NP \cap **NP-hard**

"Complete" for NP solution to <u>all</u>

• To show P = NI single NP-comple

Not only do our existing techniques for proving lower bounds not work here, we are able to <u>prove</u> that most of our techniques will <u>always</u> fail...

 To show P ≠ NP, just need a <u>single</u> lower-bound that some NP problem cannot be solved in polynomial time NP

Ρ

NP-complete

NP-hard

NP-Completeness

NP-complete = NP \cap **NP-hard**

To prove that a problem (or language) is NP-complete:

- Show it is in NP (i.e., construct a polynomial-time verifier)
- Show it is NP-hard (i.e., show <u>every</u> problem in NP reduces to it)



But there are a <u>lot</u> of problems in NP...



NP-Completeness

NP-complete = NP \cap **NP-hard**

To prove that a problem (or language) is NP-complete:

- Show it is in NP (i.e., construct a polynomial-time verifier)
- Show it is NP-hard (i.e., show <u>every</u> problem in NP reduces to it)
 - Sufficient to show that another NP-hard problem reduces to it
 - Suppose C is NP-hard and $C \leq_p B$; then for all $A \in NP$ $A \leq_p C \leq_p B \Rightarrow A \leq_p C$
 - **Challenge:** coming up with a first <u>NP-hard</u> problem



3-SAT (Satisfiability)

Shown to be NP-hard by Cook and Levin (independently)

Given a 3-CNF formula (logical AND of clauses, each an OR of 3 variables), is there an assignment of true/false to each variable to make the formula true (i.e., <u>satisfy</u> the formula)?



k-Independent Set is NP-Complete

- 1. Show that it belongs to NP
- 2. Show it is NP-Hard
 - Show 3-SAT $\leq_p k$ -Independent Set

k-Independent Set is in NP

Show: For any graph *G*:

- There is a short witness (i.e., proof) that G has a k-independent set
- The proof can be checked efficiently (in polynomial time)



Witness for $G: S = \{A, C, E, G, H, J\}$ (nodes in the *k*-independent set)

Checking the witness:

- Check that |S| = k O(k) = O(|V|)
- Check that every edge is incident on at most one node in S
 O(|V| + |E|)

Total time: O(|E| + |V|) = poly(|V| + |E|)

Graph G

k-Independent Set is NP-Complete

- 1. Show that it belongs to NP
- 2. Show it is NP-Hard
 - Show 3-SAT $\leq_p k$ -Independent Set





polynomial-time reduction

$(x \lor y \lor z) \land (x \lor \overline{y} \lor y) \land (u \lor y \lor \overline{z}) \land (z \lor \overline{x} \lor u) \land (\overline{x} \lor \overline{y} \lor \overline{z})$



For each clause, construct a <u>triangle graph</u> with its three variables as nodes Add an edge between each node and its negation

Let k = number of clauses

Claim. There is a *k*-independent set in this graph if and only if there is a satisfying assignment

$(x \lor y \lor z) \land (x \lor \overline{y} \lor y) \land (u \lor y \lor \overline{z}) \land (z \lor \overline{x} \lor u) \land (\overline{x} \lor \overline{y} \lor \overline{z})$



Suppose there is a *k*-independent set *S* in this graph *G*

- By construction of G, at most one node from each triangle is in S
- Since |S| = k and there are k triangles, each triangle contributes one node
- If a variable x is selected in one triangle, then \bar{x} is never selected in another triangle (since each variable is connected to its negation)
- There are no contradicting assignments, so can set variable chosen in each triangle to "true"; satisfying assignment by construction

$(x \lor y \lor z) \land (x \lor \overline{y} \lor y) \land (u \lor y \lor \overline{z}) \land (z \lor \overline{x} \lor u) \land (\overline{x} \lor \overline{y} \lor \overline{z})$



Suppose there is a satisfying assignment to the formula

- At least one variable in each clause must be true
- Add the node to that variable to the set *S*
- There are k clauses, so set S has exactly k nodes



polynomial-time reduction

k-Independent Set is NP-Complete

- 1. Show that it belongs to NP
- 2. Show it is NP-Hard
 - Show 3-SAT $\leq_p k$ -independent set



Max Independent Set \leq_p Min Vertex Cover

k-independent set



Reduction

k-vertex cover

k-Vertex Cover is NP-Complete

- 1. Show that it belongs to NP
 - Given a candidate cover, check that every edge is covered
- 2. Show it is NP-Hard
 - Show k-independent set $\leq_p k$ -vertex cover

k-Clique Problem

Clique: A complete subgraph *k***-Clique problem:** given a graph *G* and a number *k*, is there a clique of size *k*?



k-Clique is NP-Complete

- 1. Show that it belongs to NP
 - Give a polynomial time verifier
- 2. Show it is NP-Hard
 - Give a reduction from a known NP-Hard problem
 - We will show 3-SAT $\leq_p k$ -clique

k-Clique is in NP

Show: For any graph *G*:

- There is a short witness (i.e., proof) that G has a k-clique
- The proof can be checked efficiently (in polynomial time)



```
Suppose k = 4
```

Witness for $G: S = \{B, D, E, F\}$ (nodes in the *k*-clique)

Checking the witness:

- Check that |S| = k O(k) = O(|V|)
- Check that every pair of nodes in S share an edge $O(k^2) = O(|V|^2)$

Total time:
$$O(|V|^2) = poly(|V| + |E|)$$
 27

k-Clique is NP-Complete

1. Show that it belongs to NP



- Give a polynomial time verifier
- 2. Show it is NP-Hard
 - Give a reduction from a known NP-Hard problem
 - We will show 3-SAT $\leq_p k$ -clique



polynomial-time reduction

$(x \lor y \lor z) \land (x \lor \overline{y} \lor y) \land (u \lor y \lor \overline{z}) \land (z \lor \overline{x} \lor u) \land (\overline{x} \lor \overline{y} \lor \overline{z})$



(also do this for the other clauses, omitted due to clutter)

For each clause, introduce a node for each of its three variables

Add an edge from each node to all non-contradictory nodes in the other clauses (i.e., to all nodes that is not the negation of its own variable)

Let k = number of clauses

Claim. There is a *k*-clique in this graph if and only if there is a satisfying assignment

$(x \lor y \lor z) \land (x \lor \bar{y} \lor y) \land (u \lor y \lor \bar{z}) \land (z \lor \bar{x} \lor u) \land (\bar{x} \lor \bar{y} \lor \bar{z})$



Suppose there is a *k*-clique in this graph

- There are no edges between nodes for variables in the same clause, so k-clique must contain one node from each clause
- Nodes in clique cannot contain variable and its negation
- Nodes in clique must then correspond to a satisfying assignment

$(x \lor y \lor z) \land (x \lor \bar{y} \lor y) \land (u \lor y \lor \bar{z}) \land (z \lor \bar{x} \lor u) \land (\bar{x} \lor \bar{y} \lor \bar{z})$



Suppose there is a satisfying assignment to the formula

- For each clause, choose one node whose value is true
- There are k clauses, so this yields a collection of k nodes
- Since the assignment is consistent, there is an edge between every pair of nodes, so this constitutes a *k*-clique



polynomial-time reduction

k-Clique is NP-Complete

- 1. Show that it belongs to NP
 - Give a polynomial time verifier
- 2. Show it is NP-Hard
 - Give a reduction from a known NP-Hard problem
 - We will show 3-SAT $\leq_p k$ -clique



 \checkmark

Bonus Material: Coping with NP-Hardness

Material from subsequent slides will not be on the exam

Many optimization problems that come up in practice are NP-complete What do we do?

Approach 1: Find an algorithm that gives <u>nearly-optimal</u> solutions



Goal: Find a set of nodes such that every edge is incident on one of the nodes

Greedy approach?



Goal: Find a set of nodes such that every edge is incident on one of the nodes

Greedy approach?



Goal: Find a set of nodes such that every edge is incident on one of the nodes

Greedy approach?



Goal: Find a set of nodes such that every edge is incident on one of the nodes

Greedy approach?



Goal: Find a set of nodes such that every edge is incident on one of the nodes

Greedy approach?



Goal: Find a set of nodes such that every edge is incident on one of the nodes

Greedy approach?



Goal: Find a set of nodes such that every edge is incident on one of the nodes

Greedy approach?



Goal: Find a set of nodes such that every edge is incident on one of the nodes Greedy approach?



Goal: Find a set of nodes such that every edge is incident on one of the nodes Greedy approach?



Goal: Find a set of nodes such that every edge is incident on one of the nodes Greedy approach?



Goal: Find a set of nodes such that every edge is incident on one of the nodes Greedy approach?



Goal: Find a set of nodes such that every edge is incident on one of the nodes Greedy approach?

Greedy choice: Node with highest degree (e.g., node that covers the <u>most</u> edges)

Size of vertex cover: 5

In this case, actually optimal!

But not always optimal...



But is it "good enough?"

How do we measure good enough?

Let OPT(G) denote the size of the minimum vertex cover in G and |A(G)| be the size of the cover output by algorithm A

Define the approximation factor of *A* to be



But is it "good enough?"

How do we measure good enough?

Let OPT(G) denote the size of the minimum vertex cover in G and |A(G)| be the size of the cover output by algorithm A

Define the approximation factor of A to be

ApproxFactor(A) =
$$\frac{|A(G)|}{OPT}$$

Theorem. The greedy algorithm for vertex cover achieves an approximation factor of $\Omega(\log|V|)$ Not that great... quality of solution is worse for large instances



Goal: Obtain a <u>2-approximation</u> (i.e., vertex cover that is at most twice as large as the optimal)



Goal: Obtain a <u>2-approximation</u> (i.e., vertex cover that is at most twice as large as the optimal)

- Optimal vertex covering must contain either *u* or *v*
- **Our approach:** take <u>both</u> of them!
 - Add *u*, *v* to cover
 - Remove all edges incident on u and v
 - Repeat until no edges remain



Goal: Obtain a <u>2-approximation</u> (i.e., vertex cover that is at most twice as large as the optimal)

- Optimal vertex covering must contain either *u* or *v*
- **Our approach:** take <u>both</u> of them!
 - Add *u*, *v* to cover
 - Remove all edges incident on u and v
 - Repeat until no edges remain



Goal: Obtain a <u>2-approximation</u> (i.e., vertex cover that is at most twice as large as the optimal)

- Optimal vertex covering must contain either *u* or *v*
- **Our approach:** take <u>both</u> of them!
 - Add *u*, *v* to cover
 - Remove all edges incident on u and v
 - Repeat until no edges remain



Goal: Obtain a <u>2-approximation</u> (i.e., vertex cover that is at most twice as large as the optimal)

- Optimal vertex covering must contain either *u* or *v*
- **Our approach:** take <u>both</u> of them!
 - Add *u*, *v* to cover
 - Remove all edges incident on u and v
 - Repeat until no edges remain



Goal: Obtain a <u>2-approximation</u> (i.e., vertex cover that is at most twice as large as the optimal)

- Optimal vertex covering must contain either *u* or *v*
- **Our approach:** take <u>both</u> of them!
 - Add *u*, *v* to cover
 - Remove all edges incident on u and v
 - Repeat until no edges remain



Goal: Obtain a <u>2-approximation</u> (i.e., vertex cover that is at most twice as large as the optimal)

- Optimal vertex covering must contain either *u* or *v*
- **Our approach:** take <u>both</u> of them!
 - Add *u*, *v* to cover
 - Remove all edges incident on u and v
 - Repeat until no edges remain



Goal: Obtain a <u>2-approximation</u> (i.e., vertex cover that is at most twice as large as the optimal)

- Optimal vertex covering must contain either *u* or *v*
- **Our approach:** take <u>both</u> of them!
 - Add *u*, *v* to cover
 - Remove all edges incident on u and v
 - Repeat until no edges remain



Goal: Obtain a <u>2-approximation</u> (i.e., vertex cover that is at most twice as large as the optimal)

- Optimal vertex covering must contain either *u* or *v*
- **Our approach:** take <u>both</u> of them!
 - Add *u*, *v* to cover
 - Remove all edges incident on u and v
 - Repeat until no edges remain



Theorem. The approximate algorithm for vertex cover achieves an approximation factor of 2

- Optimal vertex covering must contain either *u* or *v*
- **Our approach:** take <u>both</u> of them!
 - Add *u*, *v* to cover
 - Remove all edges incident on u and v
 - Repeat until no edges remain

Many optimization problems that come up in practice are NP-complete What do we do?

Approach 1: Find an algorithm that gives <u>nearly-optimal</u> solutions

Question: Can we do better than a 2-approximation?

Slightly... there is an algorithm that achieves a $(2 - O(1/\sqrt{\log|V|}))$ approximation

Open Problem: Obtain a $(2 - \varepsilon)$ -approximation for constant $\varepsilon > 0$

Question: What's the best we could hope for? Can we have a 1.00001-approximation? Unlikely, computing a $\sqrt{2} \approx 1.41$ approximation is NP-hard (Khot-Minzer-Safra, 2018) **Earlier lower bounds:** 7/6 ≈ 1.17 (Håstad, 1997), $10\sqrt{5} - 21 \approx 1.36$ (Dinur-Safra, 2005)

Many optimization problems that come up in practice are NP-complete What do we do?

Approach 1: Find an algorithm that gives <u>nearly-optimal</u> solutions

Question: Can we do better than a 2-approximation?

Open Problem: Obta **Question:** What's the Hardness of approximation: many NP-hard problems are hard not only to solve exactly, but even hard to approximate (beautiful theory – see also PCP theorem)

Unlikely, computing a $\sqrt{2} \approx 1.41$ approximation is NP-hard (Khot-Minzer-Safra, 2018)

Earlier lower bounds: $7/6 \approx 1.17$ (Håstad, 1997), $10\sqrt{5} - 21 \approx 1.36$ (Dinur-Safra, 2005)

 $\sqrt{\log|V|}$ approximation

ation?

Many optimization problems that come up in practice are NP-complete What do we do?

Approach 1: Find an algorithm that gives <u>nearly-optimal</u> solutions

Approach 2: For small instances, solve using brute force or dynamic programming Can also improve (expected) run-time using heuristics

Many optimization problems that come up in practice are NP-complete What do we do?

Approach 1: Find an algorithm that gives <u>nearly-optimal</u> solutions

Approach 2: For small instances, solve using brute force or dynamic programming Can also improve (expected) run-time using heuristics

Approach 3: Special cases of the problems can be tractable

Vertex Cover on a Tree



Solve vertex cover on subtrees and take the minimum

Many optimization problems that come up in practice are NP-complete What do we do?

Approach 1: Find an algorithm that gives <u>nearly-optimal</u> solutions

Approach 2: For small instances, solve using brute force or dynamic programming Can also improve (expected) run-time using heuristics

Approach 3: Special cases of the problems can be tractable (see also parameterized complexity)