Recap from Last Time
The Complexity Class $\mathbf{P}$

- The *complexity class* $\mathbf{P}$ (for *polynomial* time) contains all problems that can be solved in polynomial time.

- Formally:

\[
\mathbf{P} = \{ L \mid \text{There is a polynomial-time decider for } L \} 
\]
The Complexity Class $\textbf{NP}$

- The complexity class $\textbf{NP}$ (\textit{nondeterministic polynomial time}) contains all problems that can be verified in polynomial time.

- Formally:

$$\textbf{NP} = \{ L \mid \text{There is a polynomial-time verifier for } L \}$$

- This means a verifier $V$'s runtime is a polynomial in $|w|$ (that is, $V$'s runtime is $O(|w|^k)$ for some integer $k$).
So how *are* we going to reason about P and NP?
New Stuff!
A Challenge
Problems in **NP** vary widely in their difficulty, even if **P = NP**.

How can we rank the relative difficulties of problems?
Reducibility
Maximum Matching

- Given an undirected graph $G$, a matching in $G$ is a set of edges such that no two edges share an endpoint.
- A maximum matching is a matching with the largest number of edges.
Maximum Matching

- Given an undirected graph \( G \), a matching in \( G \) is a set of edges such that no two edges share an endpoint.
- A maximum matching is a matching with the largest number of edges.
Maximum Matching

- Given an undirected graph $G$, a matching in $G$ is a set of edges such that no two edges share an endpoint.
- A maximum matching is a matching with the largest number of edges.

A matching, but not a maximum matching.
Maximum Matching

- Given an undirected graph $G$, a \textit{matching} in $G$ is a set of edges such that no two edges share an endpoint.
- A \textit{maximum matching} is a matching with the largest number of edges.
Maximum Matching

- Given an undirected graph $G$, a matching in $G$ is a set of edges such that no two edges share an endpoint.
- A maximum matching is a matching with the largest number of edges.
Maximum Matching

- Given an undirected graph $G$, a matching in $G$ is a set of edges such that no two edges share an endpoint.
- A maximum matching is a matching with the largest number of edges.
Maximum Matching

• Jack Edmonds' paper “Paths, Trees, and Flowers” gives a polynomial-time algorithm for finding maximum matchings.
  • (This is the same Edmonds as in “Cobham-Edmonds Thesis.”)
• Using this fact, what other problems can we solve?
Domino Tiling
Domino Tiling
Domino Tiling
Domino Tiling
Domino Tiling
Domino Tiling
Solving Domino Tiling
Solving Domino Tiling
Solving Domino Tiling
Solving Domino Tiling
Solving Domino Tiling
Solving Domino Tiling
Solving Domino Tiling
In Pseudocode

```java
boolean canPlaceDominoes(Grid G, int k) {
    return hasMatching(gridToGraph(G), k);
}
```
Based on this connection between maximum matching and domino tiling, which of the following statements would be more proper to conclude?

A. Finding a maximum matching isn’t any more difficult (in BigO/P-NP terms) than tiling a grid with dominoes.

B. Tiling a grid with dominoes isn’t any more difficult (in BigO/P-NP terms) than finding a maximum matching.
**Intuition:**

Tiling a grid with dominoes can't be "harder" than solving maximum matching, because if we can solve maximum matching efficiently, we can solve domino tiling efficiently.
Another Example
Reachability

• Consider the following problem:

  Given an directed graph \( G \) and nodes \( s \) and \( t \) in \( G \), is there a path from \( s \) to \( t \)?

• It's known that this problem can be solved in polynomial time (use DFS or BFS).

• Given that we can solve the reachability problem in polynomial time, what other problems can we solve in polynomial time?
Converter Conundrums

- Suppose that you want to plug your laptop into a projector.
- Your laptop only has a VGA output, but the projector needs HDMI input.
- You have a box of connectors that convert various types of input into various types of output (for example, VGA to DVI, DVI to DisplayPort, etc.)
- **Question:** Can you plug your laptop into the projector?
Converter Conundrums

Connectors
RGB to USB
VGA to DisplayPort
DB13W3 to CATV
DisplayPort to RGB
DB13W3 to HDMI
DVI to DB13W3
S-Video to DVI
FireWire to SDI
VGA to RGB
DVI to DisplayPort
USB to S-Video
SDI to HDMI
Converter Conundrums

Connectors
RGB to USB
VGA to DisplayPort
DB13W3 to CATV
DisplayPort to RGB
DB13W3 to HDMI
DVI to DB13W3
S-Video to DVI
FireWire to SDI
VGA to RGB
DVI to DisplayPort
USB to S-Video
SDI to HDMI
Converter Conundrums

VGA ➔ RGB ➔ USB

DisplayPort ➔ DB13W3 ➔ CATV

HDMI ➔ DVI ➔ S-Video

FireWire ➔ SDI
Converter Conundrums
Converter Conundrums

- VGA
- DisplayPort
- RGB
- HDMI
- FireWire
- USB
- DB13W3
- DVI
- SDI
- CATV
- S-Video
Converter Conundrums

**Connectors**
- RGB to USB
- VGA to DisplayPort
- DB13W3 to CATV
- DisplayPort to RGB
- DB13W3 to HDMI
- DVI to DB13W3
- S-Video to DVI
- FireWire to SDI
- VGA to RGB
- DVI to DisplayPort
- USB to S-Video
- SDI to HDMI
In Pseudocode

```java
boolean canPlugIn(List<Plug> plugs) {
    return isReachable(plugsToGraph(plugs), VGA, HDMI);
}
```
Based on this connection between plugging a laptop into a projector and determining reachability, which of the following statements would be more proper to conclude?

A. Plugging a laptop into a projector isn’t any more difficult than computing reachability in a directed graph.

B. Computing reachability in a directed graph isn’t any more difficult than plugging a laptop into a projector.
Intuition:

Finding a way to plug a computer into a projector can't be "harder" than determining reachability in a graph, since if we can determine reachability in a graph, we can find a way to plug a computer into a projector.
```cpp
bool solveProblemA(string input) {
    return solveProblemB(transform(input));
}
```

**Intuition:**

Problem A can't be “harder” than problem B, because solving problem B lets us solve problem A.
```c
bool solveProblemA(string input) {
    return solveProblemB(transform(input));
}
```

- If $A$ and $B$ are problems where it's possible to solve problem $A$ using the strategy shown above\(^*\), we write
  $$A \leq_p B.$$  
- We say that $A$ is polynomial-time reducible to $B$.

\(^*\) Assuming that $\text{transform}$ runs in polynomial time.
Polynomial-Time Reductions

- If $A \leq_p B$ and $B \in \text{P}$, then $A \in \text{P}$.

\[ P \]

\[ * \]
Polynomial-Time Reductions

- If $A \leq_p B$ and $B \in \mathbf{P}$, then $A \in \mathbf{P}$. 
Polynomial-Time Reductions

- If \( A \leq_p B \) and \( B \in \text{P} \), then \( A \in \text{P} \).
Polynomial-Time Reductions

- If $A \leq_p B$ and $B \in \mathsf{P}$, then $A \in \mathsf{P}$.
- If $A \leq_p B$ and $B \in \mathsf{NP}$, then $A \in \mathsf{NP}$.
Polynomial-Time Reductions

- If $A \leq_p B$ and $B \in \mathbf{P}$, then $A \in \mathbf{P}$.
- If $A \leq_p B$ and $B \in \mathbf{NP}$, then $A \in \mathbf{NP}$.
Polynomial-Time Reductions

- If \( A \leq_p B \) and \( B \in \mathbf{P} \), then \( A \in \mathbf{P} \).
- If \( A \leq_p B \) and \( B \in \mathbf{NP} \), then \( A \in \mathbf{NP} \).
 Polynomial-Time Reductions

- If $A \leq_p B$ and $B \in \mathbf{P}$, then $A \in \mathbf{P}$.
- If $A \leq_p B$ and $B \in \mathbf{NP}$, then $A \in \mathbf{NP}$. 
This $\leq_p$ relation lets us rank the relative difficulties of problems in $\mathbf{P}$ and $\mathbf{NP}$.

What else can we do with it?
NP-Hardness and NP-Completeness
Question: What makes a problem hard to solve?
**Intuition:** If $A \leq_p B$, then problem $B$ is at least as hard* as problem $A$.

* for some definition of “at least as hard as.”
**Intuition:** To show that some problem is hard, show that lots of other problems reduce to it.
**NP-Hardness**

- A language $L$ is called **NP-hard** if for every $A \in \text{NP}$, we have $A \leq_p L$.

The class $\text{NPC}$ is the set of NP-complete problems.
A language $L$ is called **NP-hard** if for every $A \in \text{NP}$, we have $A \leq_p L$.

The class $\text{NPC}$ is the set of NP-complete problems.

**NP-Hardness**

- A language $L$ is called **NP-hard** if for every $A \in \text{NP}$, we have $A \leq_p L$. 

A Venn diagram is shown with circles representing $P$, $\text{NP}$, and $\text{NP-Hard}$, illustrating the relationships between these classes.
A language \( L \) is called **NP-hard** if for every \( A \in \text{NP} \), we have \( A \leq_p L \).

A language in \( L \) is called **NP-complete** iff \( L \) is \( \text{NP} \)-hard and \( L \in \text{NP} \).

The class \( \text{NPC} \) is the set of \( \text{NP} \)-complete problems.
A language $L$ is called **NP-hard** if for every $A \in \text{NP}$, we have $A \leq_p L$.

Intuitively: $L$ has to be at least as hard as every problem in $\text{NP}$, since an algorithm for $L$ can be used to decide all problems in $\text{NP}$.
A language $L$ is called **NP-hard** if for every $A \in \text{NP}$, we have $A \leq_p L$.

The class $\text{NP-Hard}$ is the set of **NP-complete problems**.
A language $L$ is called $\textbf{NP-hard}$ if for every $A \in \textbf{NP}$, we have $A \leq_p L$.

A language in $L$ is called $\textbf{NP-complete}$ iff $L$ is $\textbf{NP}$-hard and $L \in \textbf{NP}$.

The class $\textbf{NPC}$ is the set of $\textbf{NP}$-complete problems.
A language $L$ is called **NP-hard** if for every $A \in \text{NP}$, we have $A \leq_p L$.

A language in $L$ is called **NP-complete** if $L$ is NP-hard and $L \in \text{NP}$.

The class **NPC** is the set of NP-complete problems.
A language $L$ is called **NP-hard** if for every $A \in \text{NP}$, we have $A \leq_p L$.

A language in $L$ is called **NP-complete** if $L$ is NP-hard and $L \in \text{NP}$.

The class $\text{NPC}$ is the set of NP-complete problems.
The Tantalizing Truth

**Theorem:** If *any* NP-complete language is in $\mathbf{P}$, then $\mathbf{P} = \mathbf{NP}$. 
The Tantalizing Truth

**Theorem**: If any NP-complete language is in P, then $P = NP$. 
The Tantalizing Truth

Theorem: If any NP-complete language is in P, then P = NP.
The Tantalizing Truth

**Theorem:** If any NP-complete language is in P, then P = NP.
The Tantalizing Truth

*Theorem:* If any \( \text{NP} \)-complete language is in \( \text{P} \), then \( \text{P} = \text{NP} \).
The Tantalizing Truth

**Theorem:** If any \textbf{NP}-complete language is in \textbf{P}, then \textbf{P} = \textbf{NP}.

**Proof:** Suppose that \( L \) is \textbf{NP}-complete and \( L \in \textbf{P} \). Now consider any arbitrary \textbf{NP} problem \( A \). Since \( L \) is \textbf{NP}-complete, we know that \( A \leq_p L \). Since \( L \in \textbf{P} \) and \( A \leq_p L \), we see that \( A \in \textbf{P} \). Since our choice of \( A \) was arbitrary, this means that \textbf{NP} \subseteq \textbf{P}, so \( \textbf{P} = \textbf{NP} \). ■
**The Tantalizing Truth**

**Theorem:** If any **NP**-complete language is not in **P**, then **P** \(\not=\) **NP**.

**Proof:** Suppose that \(L\) is an **NP**-complete language not in **P**. Since \(L\) is **NP**-complete, we know that \(L \in \text{NP}\). Therefore, we know that \(L \in \text{NP}\) and \(L \notin \text{P}\), so **P** \(\not=\) **NP**. ■
How do we even know NP-complete problems exist in the first place?
Satisfiability

- A propositional logic formula $\varphi$ is called **satisfiable** if there is some assignment to its variables that makes it evaluate to true.
  - $p \land q$ is satisfiable.
  - $p \land \lnot p$ is unsatisfiable.
  - $p \rightarrow (q \land \lnot q)$ is satisfiable.
- An assignment of true and false to the variables of $\varphi$ that makes it evaluate to true is called a **satisfying assignment**.
The **boolean satisfiability problem** (**SAT**) is the following:

**Given a propositional logic formula** $\varphi$, **is** $\varphi$ **satisfiable?**

Formally:

$$SAT = \{ \langle \varphi \rangle \mid \varphi \text{ is a satisfiable PL formula} \}$$
The language SAT happens to be in \textbf{NP}. Think about how a polynomial-time verifier for SAT might work. Which of the following would work as certificates for such a verifier, given that the input is a propositional formula $\varphi$?

A. The truth table of $\varphi$.
B. One possible variable assignment to $\varphi$.
C. A list of all possible variable assignments for $\varphi$.
D. None of the above, or two or more of the above.

Answer at \textbf{PollEv.com/cs103} or text \textbf{CS103} to \textbf{22333} once to join, then \textbf{A, B, C,} or \textbf{D}.

\[
SAT = \{ \langle \varphi \rangle \mid \varphi \text{ is a satisfiable PL formula} \}
\]
Theorem (Cook-Levin): SAT is NP-complete.

Proof Idea: To see that SAT ∈ NP, show how to make a polynomial-time verifier for it. Key idea: have the certificate be a satisfying assignment.

To show that SAT is NP-hard, given a polynomial-time verifier V for an arbitrary NP language L, for any string w you can construct a polynomials-sized formula φ(w) that says “there is a certificate c where V accepts ⟨w, c⟩.” This formula is satisfiable if and only if w ∈ L, so deciding whether the formula is satisfiable decides whether w is in L.

Proof: Take CS154!
Why All This Matters

• Resolving $\mathbf{P} \neq \mathbf{NP}$ is equivalent to just figuring out how hard SAT is.
  • If $\text{SAT} \in \mathbf{P}$, then $\mathbf{P} = \mathbf{NP}$.
    If $\text{SAT} \notin \mathbf{P}$, then $\mathbf{P} \neq \mathbf{NP}$.
Sample NP-Hard Problems

- **Computational biology:** Given a set of genomes, what is the most probable evolutionary tree that would give rise to those genomes? *(Maximum parsimony problem)*
- **Game theory:** Given an arbitrary perfect-information, finite, twoplayer game, who wins? *(Generalized geography problem)*
- **Operations research:** Given a set of jobs and workers who can perform those tasks in parallel, can you complete all the jobs within some time bound? *(Job scheduling problem)*
- **Machine learning:** Given a set of data, find the simplest way of modeling the statistical patterns in that data *(Bayesian network inference problem)*
- **Medicine:** Given a group of people who need kidneys and a group of kidney donors, find the maximum number of people who can end up with kidneys *(Cycle cover problem)*
- **Systems:** Given a set of processes and a number of processors, find the optimal way to assign those tasks so that they complete as soon as possible *(Processor scheduling problem)*
Coda: What if $P \neq NP$ is resolved?
Intermediate Problems

- With few exceptions, every problem we've discovered in \( \text{NP} \) has either
  - definitely been proven to be in \( \text{P} \), or
  - definitely been proven to be \( \text{NP} \)-complete.
- A problem that's \( \text{NP} \), not in \( \text{P} \), but not \( \text{NP} \)-complete is called \textit{NP-intermediate}.
- \textit{Theorem (Ladner):} There are \( \text{NP} \)-intermediate problems if and only if \( \text{P} \neq \text{NP} \).
What if $P \neq NP$?
A Good Read:

“A Personal View of Average-Case Complexity” by Russell Impagliazzo
What if \( \mathbf{P} = \mathbf{NP} \)?
And a Dismal Third Option