P vs. NP

Data Structures and Algorithms

Goals for today

Define "P, NP, and NP-complete"

Explain the P vs. NP problem

- why it's the biggest open problem in CS.
- And what to do when a problem you want to solve is "NP-complete"

Problem For Today

You just started a job at Amazon.

They give you the following problem:

You have a weighted, directed graph, where each vertex is a place to deliver a package.

One vertex is marked as the amazon warehouse.

How do you direct the truck to deliver all the packages, and return to the warehouse, using the least amount of gas?

This is NOT just finding a minimum spanning tree.

Keep this problem in mind, today we'll discuss why you shouldn't hope to solve it, and what you should do instead.

Taking a Big Step Back

What has this quarter been about?

We've taken problems you probably knew how to solve slowly,

And we figured out how to solve them faster.

In some sense, that's the job of a computer scientist.

Figure out how to take our problems

And make the computer do the hard work for us.

Taking a Big Step Back

Let’s take a big step back, and try to break problems into two types:

Those for which a computer might be able to help.

And those which would take so long to solve **even on a computer** we wouldn’t expect to solve them.

There are problems we could solve in finite time…but we’ll all be long dead before our computer tells us the answer.

Efficient

We'll consider a problem "efficiently solvable" if it has a polynomial time algorithm.

I.e. an algorithm that runs in time $O(n^k)$ where k is a constant.

Are these algorithms always actually efficient?

Well………no

Your n^{10000} algorithm or even your $2^{2^{2^{2^2}}} \cdot n^3$ algorithm probably aren't going to finish anytime soon.

But these edge cases are rare, and polynomial time is good as a low bar

- If we can't even find an n^{10000} algorithm, we should probably rethink our strategy

Decision Problems

Let’s go back to dividing problems into solvable/not solvable.
For today, we’re going to talk about **decision problems**.

Problems that have a “yes” or “no” answer.

Why?

Theory reasons (ask me later).

But it’s not too bad

- most problems can be rephrased as very similar decision problems.

E.g. instead of “find the shortest path from s to t” ask
Is there a path from s to t of length at most k?

P

P (stands for "Polynomial")

The set of all decision problems that have an algorithm that runs in time $O(n^k)$ for some constant k.

The decision version of all problems we've solved in this class are in P.

P is an example of a "complexity class"

A set of problems that can be solved under some limitations (e.g. with some amount of memory or in some amount of time).

I’ll know it when I see it.

Another class of problems we want to talk about.

“I’ll know it when I see it” Problems.

Decision Problems such that:

If the answer is YES, you can prove the answer is yes by

- Being given a “proof” or a “certificate”
- Verifying that certificate in polynomial time.

What certificate would be convenient for short paths?

- The path itself. Easy to check the path is really in the graph and really short.

I’ll know it when I see it.

More formally,

NP (stands for “nondeterministic polynomial”)

The set of all decision problems such that if the answer is YES, there is a proof of that which can be verified in polynomial time.

It’s a common misconception that NP stands for “not polynomial”
Please never ever ever ever say that.

Please.

Every time you do a theoretical computer scientist sheds a single tear.

(That theoretical computer scientist is me)

NP

We can **verify** YES instances of NP problems efficiently, but can we **decide** whether the answer is YES or NO efficiently?

I.e. can you bootstrap the ability to check a certificate into the ability to find a certificate?

We don't know.

This is the P vs. NP problem.

P vs. NP

P (stands for “Polynomial”)

The set of all decision problems that have an algorithm that runs in time $\boldsymbol{O(n^k)}$ for some constant k.

NP (stands for “nondeterministic polynomial”)

The set of all decision problems such that if the answer is YES, there is a proof of that which can be verified in polynomial time.

Claim: P is a subset of NP, i.e. every problem in P is also in NP

(do you see why?)

P vs. NP

P vs. NP

Are P and NP the same complexity class?
That is, can every problem that can be verified in polynomial time also be solved in polynomial time.

No one knows the answer to this problem.

In fact, it’s the biggest open problem in Computer Science.

Hard Problems

Let’s say we want to prove that some problem in NP needs exponential time (i.e. that P is not equal to NP).

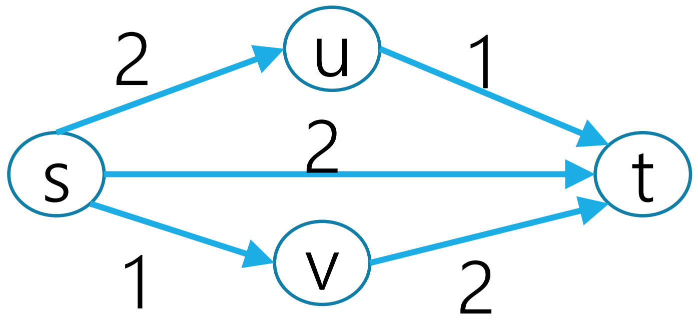
Ideally we’d start with a really hard problem in NP.

What is the hardest problem in NP?

What does it mean to be a hard problem?

Reduction (informally)

Using an algorithm for Problem B to solve Problem A.



Transform Input

Unweighted Shortest Paths

Transform Output

NP-Complete

Let's say we want to prove that some problem in NP needs exponential time (i.e. that P is not equal to NP).

Ideally we'd start with a really hard problem in NP.

What is the hardest problem in NP?

What does it mean to be a hard problem?

NP-complete

We say that a problem B is "NP-complete" if B is in NP and for all problems A in NP, A reduces to B.

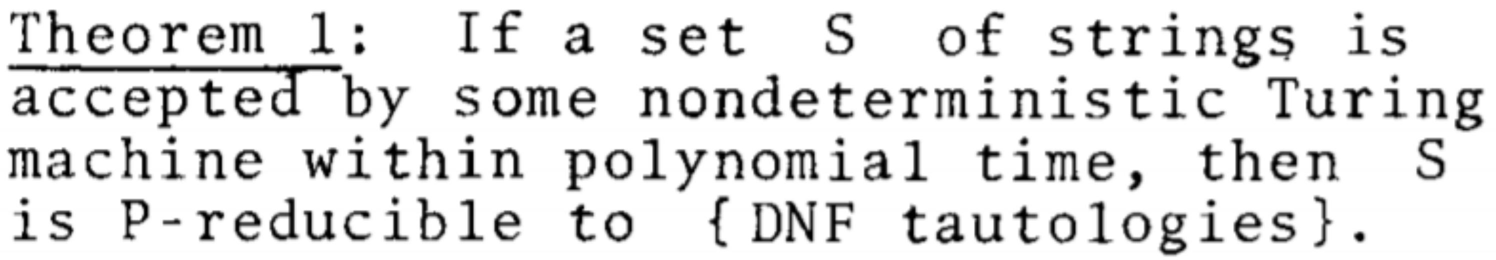
NP-Completeness

An NP-complete problem is a **universal language** for encoding "I'll know it when I see it" problems.

If you find an efficient algorithm for an NP-complete problem, you have an algorithm for **every** problem in NP

Cook-Levin Theorem (1971)

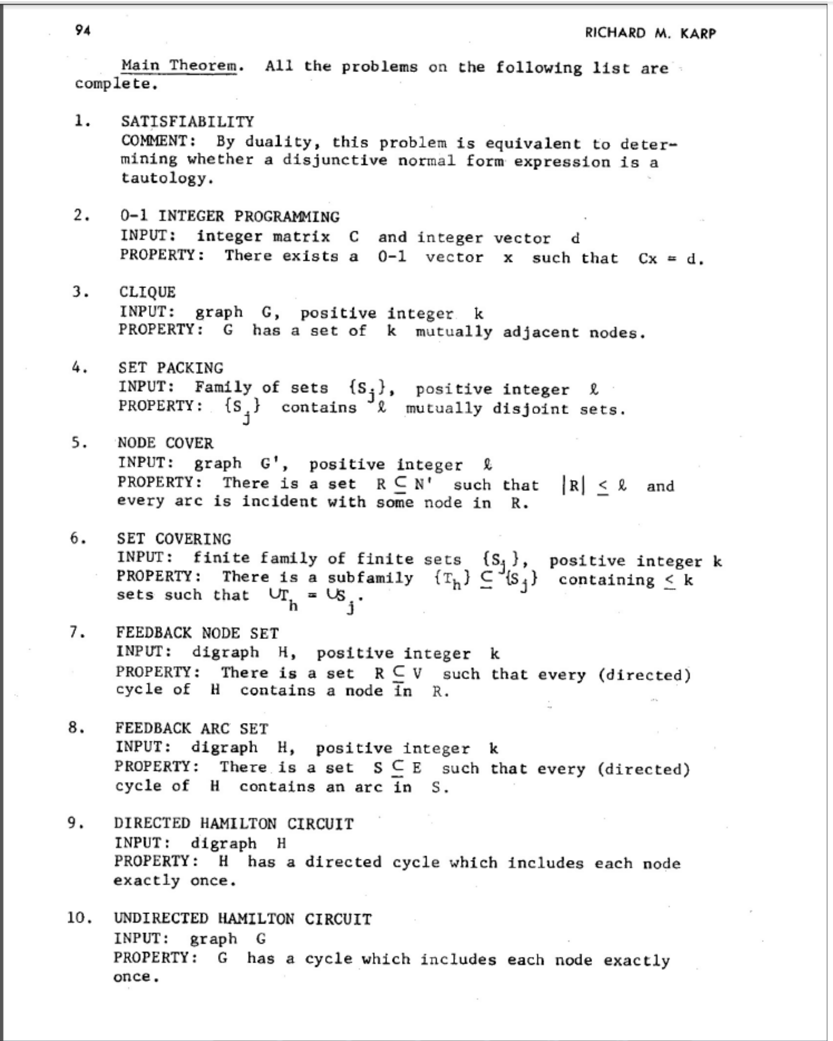
3-SAT is NP-complete



Theorem 1: If a set S of strings is accepted by some nondeterministic Turing machine within polynomial time, then S is P-reducible to {DNF tautologies}.

NP-Complete Problems

But Wait! There's more!



94 RICHARD M. KARP

<u>Main Theorem</u>. All the problems on the following list are complete.

1. SATISFIABILITY
COMMENT: By duality, this problem is equivalent to determining whether a disjunctive normal form expression is a tautology.

2. 0-1 INTEGER PROGRAMMING
INPUT: integer matrix C and integer vector d
PROPERTY: There exists a 0-1 vector x such that $Cx = d$.

3. CLIQUE
INPUT: graph G, positive integer k
PROPERTY: G has a set of k mutually adjacent nodes.

4. SET PACKING
INPUT: Family of sets $\{S_j\}$, positive integer ℓ
PROPERTY: $\{S_j\}$ contains ℓ mutually disjoint sets.

5. NODE COVER
INPUT: graph G', positive integer ℓ
PROPERTY: There is a set $R \subseteq N'$ such that $|R| \leq \ell$ and every arc is incident with some node in R.

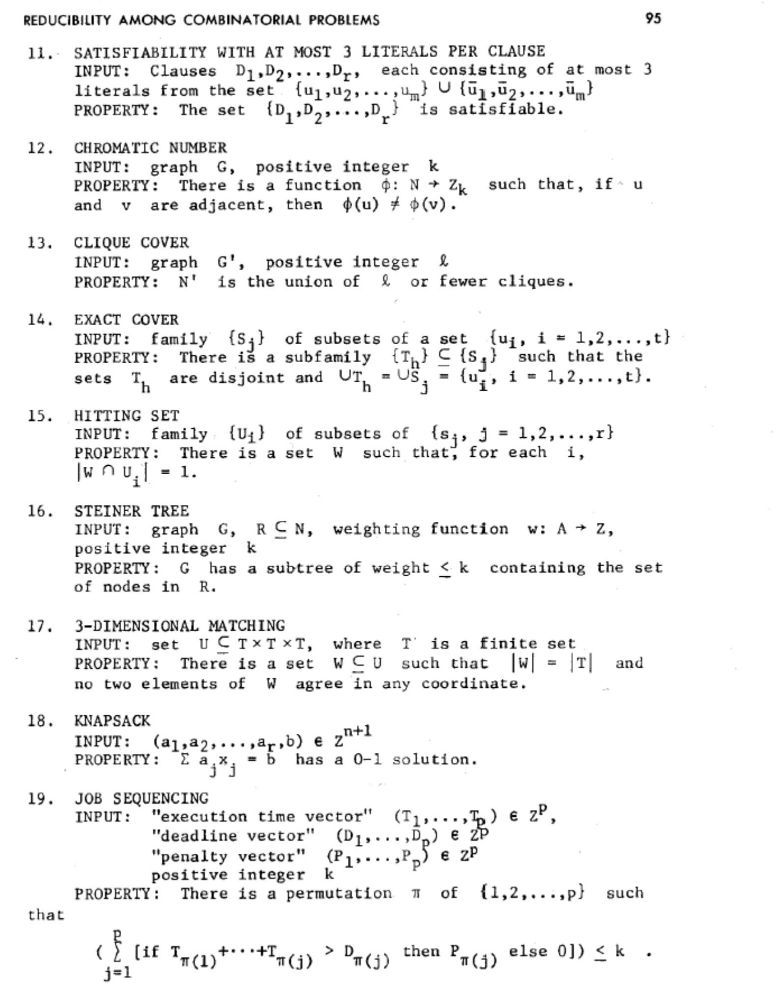
6. SET COVERING
INPUT: finite family of finite sets $\{S_j\}$, positive integer k
PROPERTY: There is a subfamily $\{T_h\} \subseteq \{S_j\}$ containing $\leq k$ sets such that $\cup T_h = \cup S_j$.

7. FEEDBACK NODE SET
INPUT: digraph H, positive integer k
PROPERTY: There is a set $R \subseteq V$ such that every (directed) cycle of H contains a node in R.

8. FEEDBACK ARC SET
INPUT: digraph H, positive integer k
PROPERTY: There is a set $S \subseteq E$ such that every (directed) cycle of H contains an arc in S.

9. DIRECTED HAMILTON CIRCUIT
INPUT: digraph H
PROPERTY: H has a directed cycle which includes each node exactly once.

10. UNDIRECTED HAMILTON CIRCUIT
INPUT: graph G
PROPERTY: G has a cycle which includes each node exactly once.



REDUCIBILITY AMONG COMBINATORIAL PROBLEMS 95

11. SATISFIABILITY WITH AT MOST 3 LITERALS PER CLAUSE
INPUT: Clauses $D_1, D_2, \ldots, D_r$, each consisting of at most 3 literals from the set $\{u_1, u_2, \ldots, u_m\} \cup \{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m\}$
PROPERTY: The set $\{D_1, D_2, \ldots, D_r\}$ is satisfiable.

12. CHROMATIC NUMBER
INPUT: graph G, positive integer k
PROPERTY: There is a function $\phi: N \to Z_k$ such that, if u and v are adjacent, then $\phi(u) \neq \phi(v)$.

13. CLIQUE COVER
INPUT: graph G', positive integer ℓ
PROPERTY: N' is the union of ℓ or fewer cliques.

14. EXACT COVER
INPUT: family $\{S_j\}$ of subsets of a set $\{u_i, i = 1,2,\ldots,t\}$
PROPERTY: There is a subfamily $\{T_h\} \subseteq \{S_j\}$ such that the sets T_h are disjoint and $\cup T_h = \cup S_j = \{u_i, i = 1,2,\ldots,t\}$.

15. HITTING SET
INPUT: family $\{U_i\}$ of subsets of $\{s_j, j = 1,2,\ldots,r\}$
PROPERTY: There is a set W such that, for each i, $|W \cap U_i| = 1$.

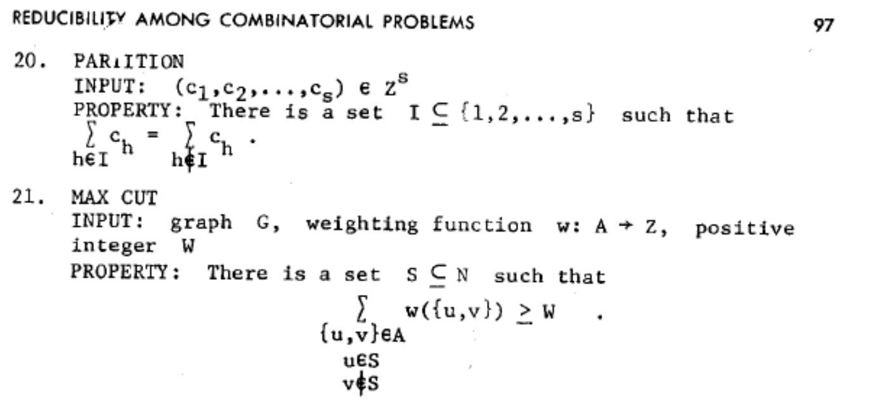
16. STEINER TREE
INPUT: graph G, $R \subseteq N$, weighting function $w: A \to Z$, positive integer k
PROPERTY: G has a subtree of weight $\leq k$ containing the set of nodes in R.

17. 3-DIMENSIONAL MATCHING
INPUT: set $U \subseteq T \times T \times T$, where T is a finite set
PROPERTY: There is a set $W \subseteq U$ such that $|W| = |T|$ and no two elements of W agree in any coordinate.

18. KNAPSACK
INPUT: $(a_1, a_2, \ldots, a_r, b) \in Z^{n+1}$
PROPERTY: $\Sigma a_j x_j = b$ has a 0-1 solution.

19. JOB SEQUENCING
INPUT: "execution time vector" $(T_1, \ldots, T_p) \in Z^p$,
"deadline vector" $(D_1, \ldots, D_p) \in Z^p$
"penalty vector" $(P_1, \ldots, P_p) \in Z^p$
positive integer k
PROPERTY: There is a permutation π of $\{1,2,\ldots,p\}$ such that

$$\left(\sum_{j=1}^{p} [\text{if } T_{\pi(1)}+\cdots+T_{\pi(j)} > D_{\pi(j)} \text{ then } P_{\pi(j)} \text{ else } 0]\right) \leq k \quad .$$



REDUCIBILITY AMONG COMBINATORIAL PROBLEMS 97

20. PARTITION
INPUT: $(c_1, c_2, \ldots, c_s) \in Z^s$
PROPERTY: There is a set $I \subseteq \{1,2,\ldots,s\}$ such that $\sum_{h \in I} c_h = \sum_{h \notin I} c_h$.

21. MAX CUT
INPUT: graph G, weighting function $w: A \to Z$, positive integer W
PROPERTY: There is a set $S \subseteq N$ such that

$$\sum_{\substack{\{u,v\} \in A \\ u \in S \\ v \notin S}} w(\{u,v\}) \geq W \quad .$$

Karp's Theorem (1972)

A lot of problems are NP-complete

NP-Complete Problems

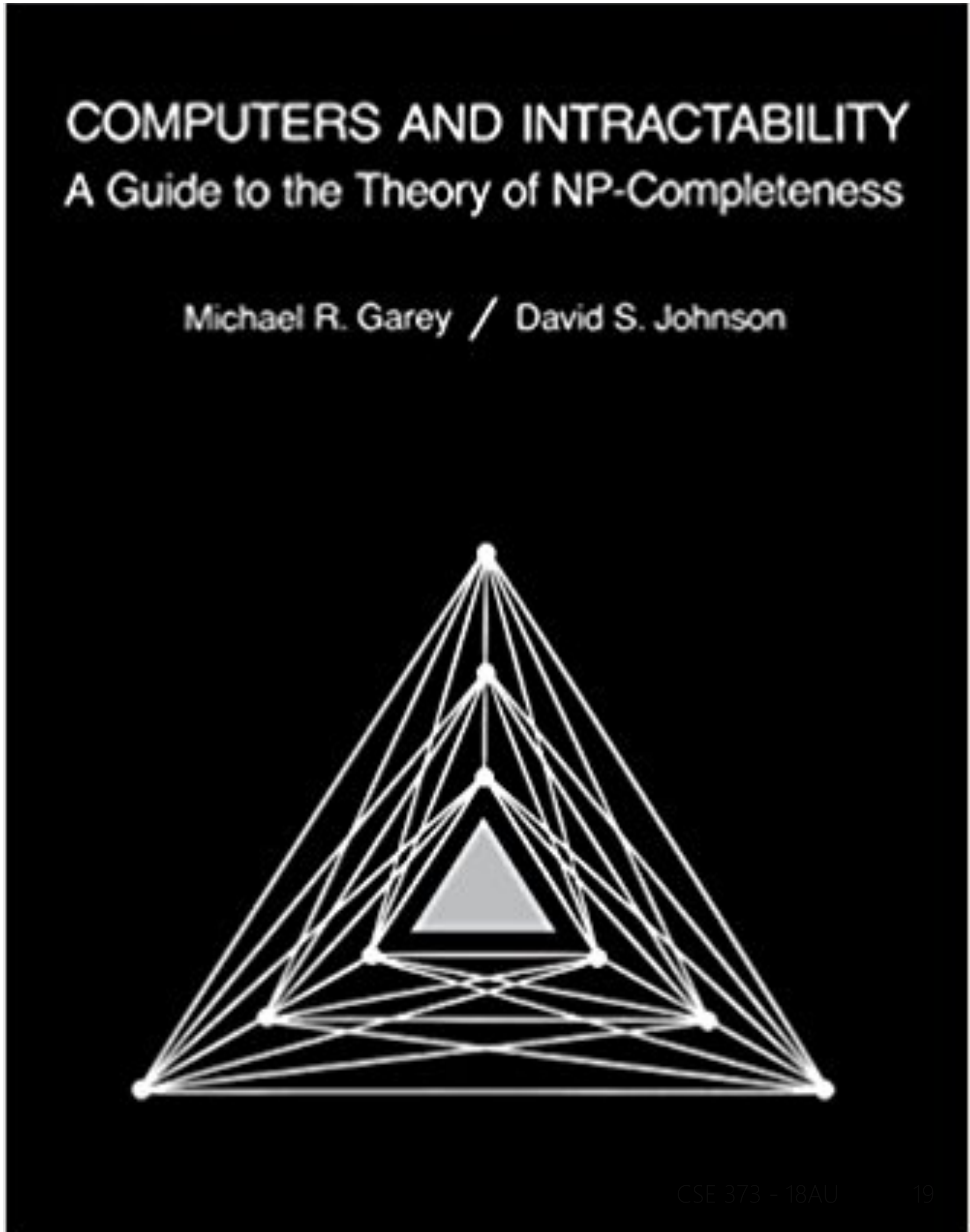
But Wait! There's more!

By 1979, at least 300 problems had been proven NP-complete.

Garey and Johnson put a list of all the NP-complete problems they could find in this textbook.

Took them almost 100 pages to just list them all.

No one has made a comprehensive list since.


COMPUTERS AND INTRACTABILITY
A Guide to the Theory of NP-Completeness

Michael R. Garey / David S. Johnson

NP-Complete Problems

But Wait! There's more!

In the last month, mathematicians and computer scientists have put papers on the arXiv claiming to show (at least) 11 more problems are NP-complete.

There are literally thousands of NP-complete problems known.

And some of them look weirdly similar to problems we've already studied.

Examples

There are literally thousands of NP-complete problems.
And some of them look weirdly similar to problems we do know efficient algorithms for.

In P

Short Path

Given a directed graph, report if there is a path from s to t of length at most k.

NP-Complete

Long Path

Given a directed graph, report if there is a path from s to t of length at least k.

Examples

In P

Light Spanning Tree

Given a weighted graph, find a spanning tree (a set of edges that connect all vertices) of weight at most k.

NP-Complete

Traveling Salesperson

Given a weighted graph, find a tour (a walk that visits every vertex and returns to its start) of weight at most k.

The electric company just needs a greedy algorithm to lay its wires.
Amazon doesn't know a way to optimally route its delivery trucks.

Dealing with NP-Completeness

Option 1: Maybe it's a special case we understand

Maybe you don't need to solve the general problem, just a special case

Option 2: Maybe it's a special case we *don't* understand (yet)

There are algorithms that are known to run quickly on "nice" instances. Maybe your problem has one of those.

One approach: Turn your problem into a SAT instance, find a solver and cross your fingers.

Dealing with NP-Completeness

Option 3: Approximation Algorithms

You might not be able to get an exact answer, but you might be able to get close.

Optimization version of Traveling Salesperson

Given a weighted graph, find a tour (a walk that visits every vertex and returns to its start) of minimum weight.

Algorithm:

Find a minimum spanning tree.

Have the tour follow the visitation order of a DFS of the spanning tree.

Theorem: This tour is at most twice as long as the best one.

Why should you care about P vs. NP

Most computer scientists are convinced that P≠NP.

Why should you care about this problem?

It's your chance for:

$1,000,000. The Clay Mathematics Institute will give $1,000,000 to whoever solves P vs. NP (or any of the 5 remaining problems they listed)

To get a Turing Award

Why should you care about P vs. NP

Most computer scientists are convinced that P≠NP.

Why should you care about this problem?

It’s your chance for:

$1,000,000. The Clay Mathematics Institute will give $1,000,000 to whoever solves P vs. NP (or any of the 5 remaining problems they listed)

To get ~~a Turing Award~~ the Turing Award renamed after you.

Why Should You Care if P=NP?

Suppose P=NP.

Specifically that we found a genuinely in-practice efficient algorithm for an NP-complete problem. What would you do?

- $1,000,000 from the Clay Math Institute obviously, but what’s next?

Why Should You Care if P=NP?

We found a genuinely in-practice efficient algorithm for an NP-complete problem. What would you do?

- Another $5,000,000 from the Clay Math Institute
- Put mathematicians out of work.
- Decrypt (essentially) all current internet communication.
- No more secure online shopping or online banking or online messaging…or online *anything.*

A world where P=NP is a very very different place from the world we live in now.

Why Should You Care if P≠NP?

We already expect P≠NP. Why should you care when we finally prove it?

P≠NP says something fundamental about the universe.

For some questions there is not a clever way to find the right answer

- Even though you'll know it when you see it.

There is actually a way to obscure information, so it cannot be found quickly no matter how clever you are.

Why Should You Care if P≠NP?

To prove P≠NP we need to better understand the differences between problems.

- Why do some problems allow easy solutions and others don't?
- What is the structure of these problems?

We don't care about P vs NP just because it has a huge effect about what the world looks like.

We will learn a lot about computation along the way.