

## 🌀 Final Exam Study Questions 🌀

This is a “core dump” of potential questions for the final exam. This should give you a good idea of the *types* of questions that we will ask on the exam. In particular, there will be a series of True/False or short-answer questions—but the actual exam questions may or may not appear in this handout. This list intentionally includes a few questions that are too long or difficult for exam conditions; these are indicated with a \*star.

**Don't forget to review the study problems for Midterms 1 and 2; the final exam is cumulative!**

---

### 🌀 How to Use These Problems 🌀

Solving every problem in this handout is *not* the best way to study for the exam. Memorizing the solutions to every problem in this handout is the *absolute worst* way to study for the exam.

What we recommend instead is to work on a *sample* of the problems. Choose one or two problems at random from each section and try to solve them from scratch under exam conditions—by yourself, in a quiet room, with a 30-minute timer, *without* your notes, *without* the internet, and if possible, even without your cheat sheet. If you're comfortable solving a few problems in a particular section, you're probably ready for that type of problem on the exam. Move on to the next section.

Discussing problems with other people (in your study groups, in the review sessions, in office hours, or on Piazza) and/or looking up old solutions can be *extremely* helpful, but *only after* you have (1) made a good-faith effort to solve the problem on your own, and (2) you have either a candidate solution or some idea about where you're getting stuck.

If you find yourself getting stuck on a particular type of problem, try to figure out *why* you're stuck. Do you understand the problem statement? Are you stuck on choosing the right high-level approach? Are you stuck on the technical details? Or are you struggling to express your ideas clearly? (We *strongly* recommend writing solutions that follow the homework grading rubrics bullet-by-bullet.)

Similarly, if feedback from other people suggests that your solutions to a particular type of problem are incorrect or incomplete, try to figure out what you missed. For NP-hardness proofs: Are you choosing a good problem to reduce from? Are you reducing in the correct direction? Are you designing your reduction with both good instances and bad instances in mind? You're not trying *solve* the problem, are you? For undecidability proofs: Does the problem have the right structure to apply Rice's theorem? If you are arguing by reduction, are you reducing in the correct direction? You're not using *pronouns*, are you?

Remember that your goal is *not* merely to “understand” the solution to any particular problem, but to become more comfortable with solving a certain *type* of problem on your own. **“Understanding” is a trap; aim for mastery.** If you can identify specific steps that you find problematic, read more *about those steps*, focus your practice *on those steps*, and try to find helpful information *about those steps* to write on your cheat sheet. Then work on the next problem!

## Minimum Spanning Trees

1. What are the definitions of safe and unsafe edges?
2. Suppose  $G = (V, E)$  is an undirected graph with distinct edge weights. Prove that every edge is either safe or unsafe.
3. Prove that the Reverse Delete algorithm is a correct algorithm to compute an MST
4. Suppose we are given both an undirected graph  $G$  with weighted edges and a minimum spanning tree  $T$  of  $G$ .
  - (a) Describe an efficient algorithm to update the minimum spanning tree when the weight of one edge  $e \in T$  is decreased.
  - (b) Describe an efficient algorithm to update the minimum spanning tree when the weight of one edge  $e \notin T$  is increased.
  - (c) Describe an efficient algorithm to update the minimum spanning tree when the weight of one edge  $e \in T$  is increased.
  - (d) Describe an efficient algorithm to update the minimum spanning tree when the weight of one edge  $e \notin T$  is decreased.
5. Let  $G = (V, E)$  an undirected edge-weighted graph. The bottleneck weight of a spanning tree  $T$  is the weight of the maximum weight edge in  $T$ . Prove that an MST is a spanning tree that minimizes the bottleneck weight.
6. Suppose we have a graph  $G = (V, E)$  with only  $k$  distinct edge weights for some small value of  $k$  (say 5). Describe an algorithm that computes an MST in  $O(mk)$  time.
7. We saw in lecture that Borouvka's algorithm for MST can be implemented in  $O(m \log n)$  time where  $m$  is the number of edges and  $n$  is the number of nodes. We also saw that Prim's algorithm can be implemented in  $O(m + n \log n)$  time. Obtain an algorithm for MST with running time  $O(m \log \log n)$  by running Borouvka's algorithm for some number of steps and then switching to Prim's algorithm.
8. Let  $G_1 = (V, E_1)$  be a connected edge-weighted graph on vertex  $V$  and let  $T_1$  be an MST of  $G_1$ . Let  $G_2 = (V, E_2)$  be a connected edge-weighted graph on vertex  $V$  and let  $T_2$  be an MST of  $G_2$ . We wish to compute the MST of the graph  $G = (V, E_1 \cup E_2)$ . Is it true that we can compute the MST in the graph  $H = (V, E(T_1) \cup E(T_2))$ ? Prove or disprove.

### Greedy Algorithms

1. Red street in the city Shampoo-Banana can be modeled as a straight line starting at 0. The street has  $n$  houses at locations  $x_1, x_2, \dots, x_n$  on the line. The local cable company wants to install some new fiber optic equipment at several locations such that every house is within distance  $r$  from one of the equipment locations. The city has granted permits to install the equipment, but only at some  $m$  locations on the street given  $y$  locations  $y_1, y_2, \dots, y_m$ . For simplicity assume that all the  $x$  and  $y$  values are distinct. You can also assume that  $x_1 < x_2 < \dots < x_n$  and that  $y_1 < y_2 < \dots < y_m$ . Describe a greedy algorithm that finds the minimum number of equipment locations that the cable company can build to satisfy the desired constraint that every house is within distance  $r$  from one of them. Your algorithm has to detect if a feasible solution does not exist. Prove the correctness of the algorithm.
  
2. Let  $N$  be a set of  $n$  items. Each item has a profit  $p_i \geq 0$ .  $N$  is partitioned into  $h$  groups  $N_1, N_2, \dots, N_h$ . We wish to pick items from  $N$  to maximize the total profit under two constraints: (i) we get to pick a total of at most  $k$  items and (ii) from each group  $N_i$  we get to pick at most  $k_i$  items. Here  $k, k_1, \dots, k_h$  are given integer parameters in addition to  $N$  and the profit values. Describe a simple greedy algorithm and prove its correctness.
  
3. Let  $S = \{x_1, x_2, \dots, x_n\}$  be a set of  $n$  points on the real line  $\mathbb{R}$ . For simplicity assume that the points are in sorted order, that is  $x_1 < x_2 < \dots < x_n$ . We say that an interval  $[a, b] \subseteq \mathbb{R}$  covers a point  $x_i$  if  $x_i \in [a, b]$ . A unit interval is an interval of the form  $[a, a + 1]$ . Describe an efficient algorithm to find the *smallest* number of unit intervals that cover every point in  $S$ . In other words, every point in  $S$  must belong to some unit interval chosen by your algorithm, and the set of unit intervals chosen must be smallest possible. Note that you get to decide where to place the unit intervals. See figure below for an example of set of points and a feasible solution; although this particular solution does not have any overlapping intervals, they are allowed.



### True or False? (All from previous final exams)

For each statement below, write “YES” or “True” if the statement is *always* true and “NO” or “False” otherwise, and give a brief (at most one short sentence) explanation of your answer. **Assume  $P \neq NP$ .** If there is any other ambiguity or uncertainty about an answer, write “NO” or “False”. For example:

- $x + y = 5$   
**NO** — Suppose  $x = 3$  and  $y = 4$ .
- 3SAT can be solved in polynomial time.  
**NO** — 3SAT is NP-hard.
- If  $P = NP$  then Jeff is the Queen of England.  
**YES** — The hypothesis is false, so the implication is true.

1. Which of the following are clear English specifications of a recursive function that could possibly be used to compute the edit distance between two strings  $A[1..n]$  and  $B[1..n]$ ?

- (a)  $Edit(i, j)$  is the answer for  $i$  and  $j$ .
- (b)  $Edit(i, j)$  is the edit distance between  $A[i]$  and  $B[j]$ .

$$(c) Edit[i, j] = \begin{cases} i & \text{if } j = 0 \\ j & \text{if } i = 0 \\ Edit[i - 1, j - 1] & \text{if } A[i] = B[j] \\ \max \begin{cases} 1 + Edit[i, j - 1] \\ 1 + Edit[i - 1, j] \\ 1 + Edit[i - 1, j - 1] \end{cases} & \text{otherwise} \end{cases}$$

- (d)  $Edit[1..n, 1..n]$  stores the edit distances for all prefixes.
- (e)  $Edit(i, j)$  is the edit distance between  $A[i..n]$  and  $B[j..n]$ .
- (f)  $Edit[i, j]$  is the value stored at row  $i$  and column  $j$  of the table.
- (g)  $Edit(i, j)$  is the edit distance between the last  $i$  characters of  $A$  and the last  $j$  characters of  $B$ .
- (h)  $Edit(i, j)$  is the edit distance when  $i$  and  $j$  are the current characters in  $A$  and  $B$ .
- (i) Iterate through both strings and update  $Edit[·, ·]$  at each character.
- (j)  $Edit(i, j, k, l)$  is the edit distance between substrings  $A[i..j]$  and  $B[k..l]$ .
- (k) [I don't need an English description; my pseudocode is clear enough!]

2. Which of the following statements is true for *every* directed graph  $G = (V, E)$ ?

- (a)  $E \neq \emptyset$ .
- (b) Given the graph  $G$  as input, Floyd-Warshall runs in  $O(E^3)$  time.

- (c) If  $G$  has at least one source and at least one sink, then  $G$  is a dag.
  - (d) We can compute a spanning tree of  $G$  using whatever-first search.
  - (e) If the edges of  $G$  are weighted, we can compute the shortest path from any node  $s$  to any node  $t$  in  $O(E \log V)$  time using Dijkstra's algorithm.
- 

3. Which of the following statements are true for *every* language  $L \subseteq \{0, 1\}^*$ ?

- (a)  $L$  is non-empty.
  - (b)  $L$  is infinite.
  - (c)  $L$  contains the empty string  $\varepsilon$ .
  - (d)  $L^*$  is infinite.
  - (e)  $L^*$  is regular.
  - (f)  $L$  is accepted by some DFA if and only if  $L$  is accepted by some NFA.
  - (g)  $L$  is described by some regular expression if and only if  $L$  is rejected by some NFA.
  - (h)  $L$  is accepted by some DFA with 42 states if and only if  $L$  is accepted by some NFA with 42 states.
  - (i) If  $L$  is decidable, then  $L$  is infinite.
  - (j) If  $L$  is not decidable, then  $L$  is infinite.
  - (k) If  $L$  is not regular, then  $L$  is undecidable.
  - (l) If  $L$  has an infinite fooling set, then  $L$  is undecidable.
  - (m) If  $L$  has a finite fooling set, then  $L$  is decidable.
  - (n) If  $L$  is the union of two regular languages, then its complement  $\bar{L}$  is regular.
  - (o) If  $L$  is the union of two regular languages, then its complement  $\bar{L}$  is context-free.
  - (p) If  $L$  is the union of two decidable languages, then  $L$  is decidable.
  - (q) If  $L$  is the union of two undecidable languages, then  $L$  is undecidable.
  - (r) If  $L \notin \mathcal{P}$ , then  $L$  is not regular.
  - (s)  $L$  is decidable if and only if its complement  $\bar{L}$  is undecidable.
  - (t) Both  $L$  and its complement  $\bar{L}$  are decidable.
- 

4. Which of the following statements are true for *at least one* language  $L \subseteq \{0, 1\}^*$ ?

- (a)  $L$  is non-empty.
- (b)  $L$  is infinite.
- (c)  $L$  contains the empty string  $\varepsilon$ .
- (d)  $L^*$  is finite.
- (e)  $L^*$  is not regular.
- (f)  $L$  is not regular but  $L^*$  is regular.
- (g)  $L$  is finite and  $L$  is undecidable.

- (h)  $L$  is decidable but  $L^*$  is not decidable.
  - (i)  $L$  is not decidable but  $L^*$  is decidable.
  - (j)  $L$  is the union of two decidable languages, but  $L$  is not decidable.
  - (k)  $L$  is the union of two undecidable languages, but  $L$  is decidable.
  - (l)  $L$  is accepted by an NFA with 374 states, but  $L$  is not accepted by a DFA with 374 states.
  - (m)  $L$  is accepted by a DFA with 374 states, but  $L$  is not accepted by a NFA with 374 states.
  - (n)  $L$  is regular and  $L \notin P$ .
  - (o) There is a Turing machine that accepts  $L$ .
  - (p) There is an algorithm to decide whether an arbitrary given Turing machine accepts  $L$ .
- 

5. Which of the following languages over the alphabet  $\{0, 1\}$  are **regular**?

- (a)  $\{0^m 1^n \mid m \geq 0 \text{ and } n \geq 0\}$
  - (b) All strings with the same number of 0s and 1s
  - (c) Binary representations of all positive integers divisible by 17
  - (d) Binary representations of all prime numbers less than  $10^{100}$
  - (e)  $\{ww \mid w \text{ is a palindrome}\}$
  - (f)  $\{wxw \mid w \text{ is a palindrome and } x \in \{0, 1\}^*\}$
  - (g)  $\{\langle M \rangle \mid M \text{ accepts a regular language}\}$
  - (h)  $\{\langle M \rangle \mid M \text{ accepts a finite number of non-palindromes}\}$
- 

6. Which of the following languages/decision problems are **decidable**?

- (a)  $\emptyset$
  - (b)  $\{0^n 1^{2n} 0^n 1^{2n} \mid n \geq 0\}$
  - (c)  $\{ww \mid w \text{ is a palindrome}\}$
  - (d)  $\{\langle M \rangle \mid M \text{ accepts } \langle M \rangle \cdot \langle M \rangle\}$
  - (e)  $\{\langle M \rangle \mid M \text{ accepts a finite number of non-palindromes}\}$
  - (f)  $\{\langle M \rangle \cdot w \mid M \text{ accepts } ww\}$
  - (g)  $\{\langle M \rangle \cdot w \mid M \text{ accepts } ww \text{ after at most } |w|^2 \text{ steps}\}$
  - (h) Given an NFA  $N$ , is the language  $L(N)$  infinite?
  - (i) CIRCUITSAT
  - (j) Given an undirected graph  $G$ , does  $G$  contain a Hamiltonian cycle?
  - (k) Given encodings of two Turing machines  $M$  and  $M'$ , is there a string  $w$  that is accepted by both  $M$  and  $M'$ ?
-

7. Which of the following languages can be proved undecidable *using Rice's Theorem*?
- (a)  $\emptyset$
  - (b)  $\{0^n 1^{2n} 0^n 1^{2n} \mid n \geq 0\}$
  - (c)  $\{ww \mid w \text{ is a palindrome}\}$
  - (d)  $\{\langle M \rangle \mid M \text{ accepts an infinite number of strings}\}$
  - (e)  $\{\langle M \rangle \mid M \text{ accepts a finite number of strings}\}$
  - (f)  $\{\langle M \rangle \mid M \text{ accepts either } \langle M \rangle \text{ or } \langle M \rangle^R\}$
  - (g)  $\{\langle M \rangle \mid M \text{ accepts both } \langle M \rangle \text{ and } \langle M \rangle^R\}$
  - (h)  $\{\langle M \rangle \mid M \text{ does not accept exactly 374 palindromes}\}$
  - (i)  $\{\langle M \rangle \mid M \text{ accepts some string } w \text{ after at most } |w|^2 \text{ steps}\}$
  - (j)  $\{\langle M \rangle \cdot w \mid M \text{ rejects } w \text{ after at most } |w|^2 \text{ steps}\}$
  - (k) Given the encodings of two Turing machines  $M$  and  $M'$ , is there a string  $w$  that is accepted by both  $M$  and  $M'$ ?
- 

8. Recall the halting language  $\text{HALT} = \{\langle M \rangle \cdot w \mid M \text{ halts on input } w\}$ . Which of the following statements about its complement  $\overline{\text{HALT}} = \Sigma^* \setminus \text{HALT}$  are true?

- (a)  $\overline{\text{HALT}}$  is empty.
  - (b)  $\overline{\text{HALT}}$  is regular.
  - (c)  $\overline{\text{HALT}}$  is infinite.
  - (d)  $\overline{\text{HALT}}$  is in NP.
  - (e)  $\overline{\text{HALT}}$  is decidable.
- 

9. Suppose some language  $A \in \{0, 1\}^*$  reduces to another language  $B \in \{0, 1\}^*$ . Which of the following statements *must* be true?

- (a) A Turing machine that recognizes  $A$  can be used to construct a Turing machine that recognizes  $B$ .
  - (b)  $A$  is decidable.
  - (c) If  $B$  is decidable then  $A$  is decidable.
  - (d) If  $A$  is decidable then  $B$  is decidable.
  - (e) If  $B$  is NP-hard then  $A$  is NP-hard.
  - (f) If  $A$  has no polynomial-time algorithm then neither does  $B$ .
-

10. Suppose there is a *polynomial-time* reduction from problem  $A$  to problem  $B$ . Which of the following statements *must* be true?

- (a) Problem  $B$  is NP-hard.
  - (b) A polynomial-time algorithm for  $B$  can be used to solve  $A$  in polynomial time.
  - (c) If  $B$  has no polynomial-time algorithm then neither does  $A$ .
  - (d) If  $A$  is NP-hard and  $B$  has a polynomial-time algorithm then  $P = NP$ .
  - (e) If  $B$  is NP-hard then  $A$  is NP-hard.
  - (f) If  $B$  is undecidable then  $A$  is undecidable.
- 

11. Consider the following pair of languages:

- $\text{HAMPATH} := \{G \mid G \text{ is an undirected graph with a Hamiltonian path}\}$
- $\text{CONNECTED} := \{G \mid G \text{ is a connected undirected graph}\}$

(For concreteness, assume that in both of these languages, graphs are represented by their adjacency matrices.) Which of the following *must* be true, assuming  $P \neq NP$ ?

- (a)  $\text{CONNECTED} \in NP$
  - (b)  $\text{HAMPATH} \in NP$
  - (c)  $\text{HAMPATH}$  is decidable.
  - (d) There is no polynomial-time reduction from  $\text{HAMPATH}$  to  $\text{CONNECTED}$ .
  - (e) There is no polynomial-time reduction from  $\text{CONNECTED}$  to  $\text{HAMPATH}$ .
- 

12. Consider the following pair of languages:

- $\text{DIRHAMPATH} := \{G \mid G \text{ is a directed graph with a Hamiltonian path}\}$
- $\text{ACYCLIC} := \{G \mid G \text{ is a directed acyclic graph}\}$

(For concreteness, assume that in both of these languages, graphs are represented by their adjacency matrices.) Which of the following *must* be true, assuming  $P \neq NP$ ?

- (a)  $\text{ACYCLIC} \in NP$
  - (b)  $\text{ACYCLIC} \cap \text{DIRHAMPATH} \in P$
  - (c)  $\text{DIRHAMPATH}$  is decidable.
  - (d) There is a polynomial-time reduction from  $\text{DIRHAMPATH}$  to  $\text{ACYCLIC}$ .
  - (e) There is a polynomial-time reduction from  $\text{ACYCLIC}$  to  $\text{DIRHAMPATH}$ .
- 

13. Consider the following pair of languages:

- $\text{3COLOR} := \{G \mid G \text{ is a 3-colorable undirected graph}\}$
- $\text{TREE} := \{G \mid G \text{ is a connected acyclic undirected graph}\}$



(For concreteness, assume that in both of these languages, graphs are represented by adjacency matrices.) Which of the following *must* be true, assuming  $P \neq NP$ ?

- (a) TREE is NP-hard.
  - (b)  $TREE \cap 3COLOR \in P$
  - (c) 3COLOR is undecidable.
  - (d) There is a polynomial-time reduction from 3COLOR to TREE.
  - (e) There is a polynomial-time reduction from TREE to 3COLOR.
-

## NP-hardness

1. A boolean formula is in *disjunctive normal form* (or *DNF*) if it consists of a *disjunction* (OR) or several *terms*, each of which is the conjunction (AND) of one or more literals. For example, the formula

$$(\bar{x} \wedge y \wedge \bar{z}) \vee (y \wedge z) \vee (x \wedge \bar{y} \wedge \bar{z})$$

is in disjunctive normal form. DNF-SAT asks, given a boolean formula in disjunctive normal form, whether that formula is satisfiable.

- (a) Describe a polynomial-time algorithm to solve DNF-SAT.
- (b) What is the error in the following argument that P=NP?

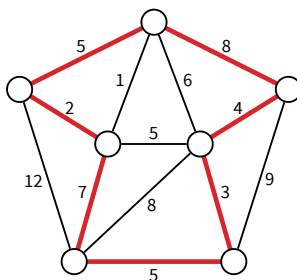
*Suppose we are given a boolean formula in conjunctive normal form with at most three literals per clause, and we want to know if it is satisfiable. We can use the distributive law to construct an equivalent formula in disjunctive normal form. For example,*

$$(x \vee y \vee \bar{z}) \wedge (\bar{x} \vee \bar{y}) \iff (x \wedge \bar{y}) \vee (y \wedge \bar{x}) \vee (\bar{z} \wedge \bar{x}) \vee (\bar{z} \wedge \bar{y})$$

*Now we can use the algorithm from part (a) to determine, in polynomial time, whether the resulting DNF formula is satisfiable. We have just solved 3Sat in polynomial time. Since 3Sat is NP-hard, we must conclude that P=NP!*

2. A **relaxed 3-coloring** of a graph  $G$  assigns each vertex of  $G$  one of three colors (for example, red, green, and blue), such that **at most one** edge in  $G$  has both endpoints the same color.
  - (a) Give an example of a graph that has a relaxed 3-coloring, but does not have a proper 3-coloring (where every edge has endpoints of different colors).
  - (b) **Prove** that it is NP-hard to determine whether a given graph has a relaxed 3-coloring.
3. An **ultra-Hamiltonian tour** in  $G$  is a closed walk  $W$  that visits every vertex of  $G$  exactly once, except for *at most one* vertex that  $W$  visits more than once.
  - (a) Give an example of a graph that contains a ultra-Hamiltonian tour, but does not contain a Hamiltonian cycle (which visits every vertex exactly once).
  - (b) **Prove** that it is NP-hard to determine whether a given graph contains a ultra-Hamiltonian tour.
4. An **infra-Hamiltonian cycle** in  $G$  is a closed walk  $W$  that visits every vertex of  $G$  exactly once, except for *at most one* vertex that  $W$  does not visit at all.
  - (a) Give an example of a graph that contains a infra-Hamiltonian cycle, but does not contain a Hamiltonian cycle (which visits every vertex exactly once).
  - (b) **Prove** that it is NP-hard to determine whether a given graph contains a infra-Hamiltonian cycle.
5. A **quasi-satisfying assignment** for a 3CNF boolean formula  $\Phi$  is an assignment of truth values to the variables such that *at most one* clause in  $\Phi$  does not contain a true literal. **Prove** that it is NP-hard to determine whether a given 3CNF boolean formula has a quasi-satisfying assignment.

6. A subset  $S$  of vertices in an undirected graph  $G$  is **half-independent** if each vertex in  $S$  is adjacent to *at most one* other vertex in  $S$ . Prove that finding the size of the largest half-independent set of vertices in a given undirected graph is NP-hard.
7. A subset  $S$  of vertices in an undirected graph  $G$  is **sort-of-independent** if each vertex in  $S$  is adjacent to *at most 374* other vertices in  $S$ . Prove that finding the size of the largest sort-of-independent set of vertices in a given undirected graph is NP-hard.
8. A subset  $S$  of vertices in an undirected graph  $G$  is **almost independent** if at most 374 edges in  $G$  have both endpoints in  $S$ . Prove that finding the size of the largest almost-independent set of vertices in a given undirected graph is NP-hard.
9. Let  $G$  be an undirected graph with weighted edges. A **heavy Hamiltonian cycle** is a cycle  $C$  that passes through each vertex of  $G$  exactly once, such that the total weight of the edges in  $C$  is more than half of the total weight of all edges in  $G$ . Prove that deciding whether a graph has a heavy Hamiltonian cycle is NP-hard.



A heavy Hamiltonian cycle. The cycle has total weight 34; the graph has total weight 67.

10. (a) A **tonian path** in a graph  $G$  is a path that goes through at least half of the vertices of  $G$ . Show that determining whether a graph has a tonian path is NP-hard.  
 (b) A **tonian cycle** in a graph  $G$  is a cycle that goes through at least half of the vertices of  $G$ . Show that determining whether a graph has a tonian cycle is NP-hard. [Hint: Use part (a). Or not.]
11. Prove that the following variants of SAT is NP-hard. [Hint: Describe reductions from 3SAT.]  
 (a) Given a boolean formula  $\Phi$  in conjunctive normal form, where *each variable appears in at most three clauses*, determine whether  $\Phi$  has a satisfying assignment. [Hint: First consider the variant where each variable appears in at most **five** clauses.]  
 (b) Given a boolean formula  $\Phi$  in conjunctive normal form *and given one satisfying assignment for  $\Phi$* , determine whether  $\Phi$  has at least one other satisfying assignment.
12. Jerry Springer and Maury Povich have decided not to compete with each other over scheduling guests during the next talk-show season. There is only one set of Weird People who either host would consider having on their show. The hosts want to divide the Weird People into two disjoint subsets: those to appear on Jerry’s show, and those to appear on Maury’s show. (Neither wants to “recycle” a guest that appeared on the other’s show.)

Both Jerry and Maury have preferences about which Weird People they are particularly interested in. For example, Jerry wants at least one guest who fits the description “was abducted by a flying saucer”. Thus, on his list of preferences, he writes “ $w_1$  or  $w_3$  or  $w_{45}$ ”, since weird people numbered 1, 3, and 45 are the only ones who fit that description. Jerry has other preferences as well, so he lists those also. Similarly, Maury might like to include at least one guest who “really enjoys Rice’s theorem”. Each potential guest may fall into any number of different categories, such as the person who enjoys Rice’s theorem more than their involuntary flying-saucer voyage.

Jerry and Maury each prepare a list reflecting all of their preferences. Each list contains a collection of statements of the form “( $w_i$  or  $w_j$  or  $w_k$ )”. Your task is to prove that it is NP-hard to find an assignment of weird guests to the two shows that satisfies all of Jerry’s preferences and all of Maury’s preferences.

- (a) The problem `NOMIXEDCLAUSES3SAT` is the special case of `3SAT` where the input formula cannot contain a clause with both a negated variable and a non-negated variable. Prove that `NOMIXEDCLAUSES3SAT` is NP-hard. [Hint: Reduce from the standard `3SAT` problem.]
- (b) Describe a polynomial-time reduction from `NOMIXEDCLAUSES3SAT` to `3SAT`.

13. The president of Sham-Poobanana University is planning An Unofficial St. Brigid’s Day party for the university staff.<sup>1</sup> His staff has a hierarchical structure; that is, the supervisor relation forms a directed, acyclic graph, with the president as the only source, and with an edge from person  $i$  to person  $j$  in the graph if and only if person  $i$  is an immediate supervisor of person  $j$ . (Many staff members have multiple positions, and thus have several immediate supervisors.) In order to make the party fun for all guests, the president wants to ensure that if a person  $i$  attends, then none of  $i$ ’s immediate supervisors can attend.

By mining each staff member’s email and social media accounts, Sham-Poobanana University Human Resources has determined a “party-hound” rating for each staff member, which is a non-negative real number reflecting how likely it is that the person will leave the party wearing a monkey suit and a lampshade.

Show that it is NP-hard to determine a guest-list that *maximizes* the sum of the party-hound ratings of all invited guests, subject to the supervisor constraint.

[Hint: This problem can be solved in polynomial time when the input graph is a tree!]

14. Prove that the following problem (which we call `MATCH`) is NP-hard. The input is a finite set  $S$  of strings, all of the same length  $n$ , over the alphabet  $\{0, 1, 2\}$ . The problem is to determine whether there is a string  $w \in \{0, 1\}^n$  such that for every string  $s \in S$ , the strings  $s$  and  $w$  have the same symbol in at least one position.

For example, given the set  $S = \{01220, 21110, 21120, 00211, 11101\}$ , the correct output is `TRUE`, because the string  $w = 01001$  matches the first three strings of  $S$  in the second position, and matches the last two strings of  $S$  in the last position. On the other hand, given the set  $S = \{00, 11, 01, 10\}$ , the correct output is `FALSE`.

[Hint: Describe a reduction from `SAT` (or `3SAT`)]

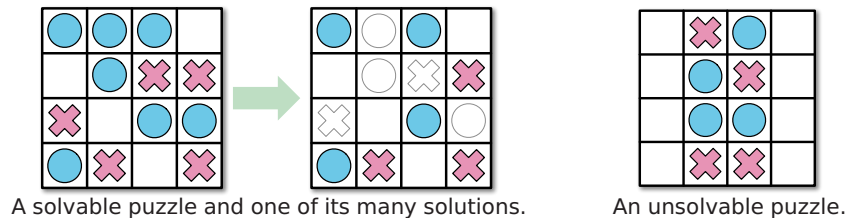
<sup>1</sup>As I’m sure you already know, St. Brigid of Kildare is one of the patron saints of Ireland, chicken and dairy farmers, and academics.

15. Consider the following solitaire game. The puzzle consists of an  $n \times m$  grid of squares, where each square may be empty, occupied by a red stone, or occupied by a blue stone. The goal of the puzzle is to remove some of the given stones so that the remaining stones satisfy two conditions:

- (1) Every row contains at least one stone.
- (2) No column contains stones of both colors.

For some initial configurations of stones, reaching this goal is impossible; see the example below.

Prove that it is NP-hard to determine, given an initial configuration of red and blue stones, whether this puzzle can be solved.



A solvable puzzle and one of its many solutions.

An unsolvable puzzle.

16. To celebrate the end of the semester, Professor Jarling wants to treat himself to an ice-cream cone at the *Polynomial House of Flavors*. For a fixed price, he can build a cone with as many scoops as he'd like. Because he has good balance (and because we want this problem to work out), Prof. Jarling can balance any number of scoops on top of the cone without it tipping over. He plans to eat the ice cream one scoop at a time, from top to bottom, and doesn't want more than one scoop of any flavor.

However, he realizes that eating a scoop of bubblegum ice cream immediately after the scoop of potatoes-and-gravy ice cream would be unpalatable; these two flavors clearly should not be placed next to each other in the stack. He has other similar constraints; certain pairs of flavors cannot be adjacent in the stack.

He'd like to get as much ice cream as he can for the one fee by building the tallest cone possible that meets his flavor-incompatibility constraints. Prove that Prof. Jarling's problem is NP-hard.

17. Prove that the following problems are NP-hard.
- (a) Given an undirected graph  $G$ , does  $G$  contain a simple path that visits all but 17 vertices?
  - (b) Given an undirected graph  $G$ , does  $G$  have a spanning tree in which every node has degree at most 23?
  - (c) Given an undirected graph  $G$ , does  $G$  have a spanning tree with at most 42 leaves?
18. Prove that the following problems are NP-hard.
- (a) Given an undirected graph  $G$ , is it possible to color the vertices of  $G$  with three different colors, so that at most 31337 edges have both endpoints the same color?
  - (b) Given an undirected graph  $G$ , is it possible to color the vertices of  $G$  with three different colors, so that each vertex has at most 8675309 neighbors with the same color?

19. At the end of every semester, Jeff needs to solve the following EXAMDESIGN problem. He has a list of problems, and he knows for each problem which students will *really enjoy* that problem. He needs to choose a subset of problems for the exam such that for each student in the class, the exam includes at least one question that student will really enjoy. On the other hand, he does not want to spend the entire summer grading an exam with dozens of questions, so the exam must also contain as few questions as possible. Prove that the EXAMDESIGN problem is NP-hard.
20. Which of the following results would resolve the P vs. NP question? Justify each answer with a short sentence or two.
- (a) The construction of a polynomial time algorithm for some problem in NP.
  - (b) A polynomial-time reduction from 3SAT to the language  $\{0^n 1^n \mid n \geq 0\}$ .
  - (c) A polynomial-time reduction from  $\{0^n 1^n \mid n \geq 0\}$  to 3SAT.
  - (d) A polynomial-time reduction from 3COLOR to MINVERTEXCOVER.
  - (e) The construction of a nondeterministic Turing machine that cannot be simulated by any deterministic Turing machine with the same running time.

**Some useful NP-hard problems.** You are welcome to use any of these in your own NP-hardness proofs, except of course for the specific problem you are trying to prove NP-hard. **The final exam will include a copy of this list.**

**CIRCUITSAT:** Given a boolean circuit, are there any input values that make the circuit output TRUE?

**3SAT:** Given a boolean formula in conjunctive normal form, with exactly three distinct literals per clause, does the formula have a satisfying assignment?

**MAXINDEPENDENTSET:** Given an undirected graph  $G$ , what is the size of the largest subset of vertices in  $G$  that have no edges among them?

**MAXCLIQUE:** Given an undirected graph  $G$ , what is the size of the largest complete subgraph of  $G$ ?

**MINVERTEXCOVER:** Given an undirected graph  $G$ , what is the size of the smallest subset of vertices that touch every edge in  $G$ ?

**MINSETCOVER:** Given a collection of subsets  $S_1, S_2, \dots, S_m$  of a set  $S$ , what is the size of the smallest subcollection whose union is  $S$ ?

**MINHITTINGSET:** Given a collection of subsets  $S_1, S_2, \dots, S_m$  of a set  $S$ , what is the size of the smallest subset of  $S$  that intersects every subset  $S_i$ ?

**3COLOR:** Given an undirected graph  $G$ , can its vertices be colored with three colors, so that every edge touches vertices with two different colors?

**HAMILTONIANPATH:** Given graph  $G$  (either directed or undirected), is there a path in  $G$  that visits every vertex exactly once?

**HAMILTONIANCYCLE:** Given a graph  $G$  (either directed or undirected), is there a cycle in  $G$  that visits every vertex exactly once?

**TRAVELINGSALESMAN:** Given a graph  $G$  (either directed or undirected) with weighted edges, what is the minimum total weight of any Hamiltonian path/cycle in  $G$ ?

**LONGESTPATH:** Given a graph  $G$  (either directed or undirected, possibly with weighted edges), what is the length of the longest simple path in  $G$ ?

**STEINERTREE:** Given an undirected graph  $G$  with some of the vertices marked, what is the minimum number of edges in a subtree of  $G$  that contains every marked vertex?

**SUBSETSUM:** Given a set  $X$  of positive integers and an integer  $k$ , does  $X$  have a subset whose elements sum to  $k$ ?

**PARTITION:** Given a set  $X$  of positive integers, can  $X$  be partitioned into two subsets with the same sum?

**3PARTITION:** Given a set  $X$  of  $3n$  positive integers, can  $X$  be partitioned into  $n$  three-element subsets, all with the same sum?

**INTEGERLINEARPROGRAMMING:** Given a matrix  $A \in \mathbb{Z}^{n \times d}$  and two vectors  $b \in \mathbb{Z}^n$  and  $c \in \mathbb{Z}^d$ , compute  $\max\{c \cdot x \mid Ax \leq b, x \geq 0, x \in \mathbb{Z}^d\}$ .

**FEASIBLEILP:** Given a matrix  $A \in \mathbb{Z}^{n \times d}$  and a vector  $b \in \mathbb{Z}^n$ , determine whether the set of feasible integer points  $\max\{x \in \mathbb{Z}^d \mid Ax \leq b, x \geq 0\}$  is empty.

**DRAUGHTS:** Given an  $n \times n$  international draughts configuration, what is the largest number of pieces that can (and therefore must) be captured in a single move?

**STEAMEDHAMS:** Aurora borealis? At this time of year, at this time of day, in this part of the country, localized entirely within your kitchen? May I see it?



## Turing Machines and Undecidability

For each of the following languages, either *sketch* an algorithm to decide that language or *prove* that the language is undecidable, using a diagonalization argument, a reduction argument, Rice's theorem, closure properties, or some combination of the above. Recall that  $w^R$  denotes the reversal of string  $w$ . Since we did not teach Rice's theorem this semester we will not expect you to be able to use it. You should be able to prove undecidability *without* using Rice's theorem.

1.  $\emptyset$
2.  $\{0^n 1^n 2^n \mid n \geq 0\}$
3.  $\{A \in \{0, 1\}^{n \times n} \mid n \geq 0 \text{ and } A \text{ is the adjacency matrix of a dag with } n \text{ vertices}\}$
4.  $\{A \in \{0, 1\}^{n \times n} \mid n \geq 0 \text{ and } A \text{ is the adjacency matrix of a 3-colorable graph with } n \text{ vertices}\}$
5.  $\{\langle M \rangle \mid M \text{ accepts } \langle M \rangle^R\}$
6.  $\{\langle M \rangle \mid M \text{ accepts } \langle M \rangle^R\} \cap \{\langle M \rangle \mid M \text{ rejects } \langle M \rangle^R\}$
7.  $\{\langle M \rangle \# w \mid M \text{ accepts } ww^R\}$
8.  $\{\langle M \rangle \mid M \text{ accepts RICESTHEOREM}\}$
9.  $\{\langle M \rangle \mid M \text{ rejects RICESTHEOREM}\}$
10.  $\{\langle M \rangle \mid M \text{ accepts at least one palindrome}\}$
11.  $\Sigma^* \setminus \{\langle M \rangle \mid M \text{ accepts at least one palindrome}\}$
12.  $\{\langle M \rangle \mid M \text{ rejects at least one palindrome}\}$
13.  $\{\langle M \rangle \mid M \text{ accepts exactly one string of length } \ell, \text{ for each integer } \ell \geq 0\}$
14.  $\{\langle M \rangle \mid \text{ACCEPT}(M) \text{ has an infinite fooling set}\}$
15.  $\{\langle M \rangle \# \langle M' \rangle \mid \text{ACCEPT}(M) \cap \text{ACCEPT}(M') \neq \emptyset\}$
16.  $\{\langle M \rangle \# \langle M' \rangle \mid \text{ACCEPT}(M) \oplus \text{REJECT}(M') \neq \emptyset\}$  — Here  $\oplus$  means exclusive-or.

**Some useful undecidable problems.** You are welcome to use any of these in your own undecidability proofs, except of course for the specific problem you are trying to prove undecidable.

$$\text{SELFREJECT} := \{ \langle M \rangle \mid M \text{ rejects } \langle M \rangle \}$$

$$\text{SELFACCEPT} := \{ \langle M \rangle \mid M \text{ accepts } \langle M \rangle \}$$

$$\text{SELFHALT} := \{ \langle M \rangle \mid M \text{ halts on } \langle M \rangle \}$$

$$\text{SELFDIVERGE} := \{ \langle M \rangle \mid M \text{ does not halt on } \langle M \rangle \}$$

$$\text{REJECT} := \{ \langle M \rangle \# w \mid M \text{ rejects } w \}$$

$$\text{ACCEPT} := \{ \langle M \rangle \# w \mid M \text{ accepts } w \}$$

$$\text{HALT} := \{ \langle M \rangle \# w \mid M \text{ halts on } w \}$$

$$\text{DIVERGE} := \{ \langle M \rangle \# w \mid M \text{ does not halt on } w \}$$

$$\text{NEVERREJECT} := \{ \langle M \rangle \mid \text{REJECT}(M) = \emptyset \}$$

$$\text{NEVERACCEPT} := \{ \langle M \rangle \mid \text{ACCEPT}(M) = \emptyset \}$$

$$\text{NEVERHALT} := \{ \langle M \rangle \mid \text{HALT}(M) = \emptyset \}$$

$$\text{NEVERDIVERGE} := \{ \langle M \rangle \mid \text{DIVERGE}(M) = \emptyset \}$$