This is a "core dump" of potential questions for Midterm 3. This should give you a good idea of the *types* of questions that we will ask on the exam, 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. In particular, there are several multipart questions where we would ask only one or two parts on an exam.

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: 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!

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

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

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

1.A. Describe a polynomial-time algorithm to solve DNF-SAT.

1.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,*
>
> $$(x \vee y \vee \overline{z}) \wedge (\overline{x} \vee \overline{y}) \iff (x \wedge \overline{y}) \vee (y \wedge \overline{x}) \vee (\overline{z} \wedge \overline{x}) \vee (\overline{z} \wedge \overline{y})$$
>
> *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.

2.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).

2.B. ***Prove*** that it is NP-hard to determine whether a given graph has a relaxed 3-coloring.

3. An ***ultra-Hamiltonian cycle*** in G is a closed walk C that visits every vertex of G exactly once, except for *at most one* vertex that C visits more than once.

3.A. Give an example of a graph that contains a ultra-Hamiltonian cycle, but does not contain a Hamiltonian cycle (which visits every vertex exactly once).

3.B. ***Prove*** that it is NP-hard to determine whether a given graph contains a ultra-Hamiltonian cycle.

4. An ***infra-Hamiltonian cycle*** in G is a closed walk C that visits every vertex of G exactly once, except for *at most one* vertex that C does not visit at all.

4.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).

4.B. ***Prove*** that it is NP-hard to determine whether a given graph contains a infra-Hamiltonian cycle.

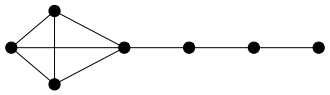
5. A ***quasi-satisfying assignment*** for a 3CNF boolean formula Φ is an assignment of truth values to the variables such that *at most one* clause in Φ does not contain a true literal. ***Prove*** that it is NP-hard to determine whether a given 3CNF boolean formula has a quasi-satisfying assignment.

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 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.

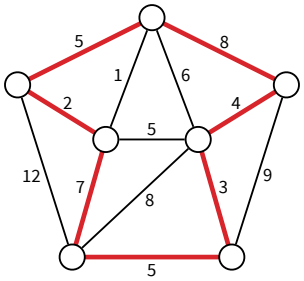
8. A subset S of vertices in an undirected graph G is ***almost independent*** if at most 374 edges in G have both endpoints in S. Prove that finding the size of the largest almost-independent set of vertices in a given undirected graph is NP-hard.

9. A ***kite*** of size k is a complete graph (clique) on k vertices plus a tail of k vertices. ***Prove*** that it is NP-hard to determine, if given an undirected graph G and integer k, whether G contains a kite of size k as a subgraph.



A kite of size 4.

10. 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.

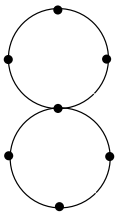


A heavy Hamiltonian cycle. The cycle has total weight 34; the graph has total weight 67.

11. An undirected graph is an ***eight-graph*** if it has an odd number of nodes, say $2n-1$, and consists of two cycles C_1 and C_2 on n vertices each and C_1 and C_2 share exactly one vertex. ***Prove*** that is NP-hard to determine, if given an undirected graph G and an integer k, whether G has a subgraph which is an eight-graph on $2k-1$ nodes.

12.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.

12.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.)



An eight-graph on 7 vertices.

13. Prove that the following variants of SAT is NP-hard. (**Hint:** Describe reductions from 3SAT.)

13.A. Given a boolean formula Φ in conjunctive normal form, where *each variable appears in at most three clauses*, determine whether F has a satisfying assignment. (**Hint:** First consider the variant where each variable appears in at most **five** clauses.)

13.B. Given a boolean formula Φ in conjunctive normal form *and given one satisfying assignment for* Φ, determine whether Φ has at least one other satisfying assignment.

14. 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) groups: 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 to be sure to get at least one person who fits the category "had extra-terrestrial affair". 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 guarantee that his show includes at least one guest who confesses to "really enjoying eating toothpaste". Each potential guest may fall into any number of different categories, such as the person who enjoys eating toothpaste more than the extra-terrestrial affair they had.

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.

14.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.)

14.B. Describe a polynomial-time reduction from NoMixedClauses3Sat to 3Sat.

15. 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))

16. To celebrate the end of the semester, Professor Jarling want 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), assume that he 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 this problem is NP-hard.

17. Prove that the following problems are NP-hard.

17.A. Given an undirected graph G, does G contain a simple path that visits all but 17 vertices?

17.B. Given an undirected graph G, does G have a spanning tree in which every node has degree at most 23?

17.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.

18.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?

18.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 the semester, the instructors need to solve the following EXAMDESIGN problem. They have a list of problems, and they know for each problem which students will *really enjoy* that problem. They need 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, they do 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.

20.A. The construction of a polynomial time algorithm for some problem in NP.

20.B. A polynomial-time reduction from 3SAT to the language $\{0^n1^n \mid n \geq 0\}$.

20.C. A polynomial-time reduction from $\{0^n1^n \mid n \geq 0\}$ to 3SAT.

20.D. A polynomial-time reduction from 3COLOR to MINVERTEXCOVER.

20.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.

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 $S_1, S_2, \ldots, S_m$ of a set S, what is the size of the smallest subcollection whose union is S?

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?

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, closure properties, or some combination of the above. Recall that w^R denotes the reversal of string w.

1. $\varnothing$

2. $\{0^n 1^n 2^n \mid n \geq 0\}$

3. $\{A \in \{0,1\}^{n \times n} \mid n \geq 0$ and A is the adjacency matrix of a dag with n vertices$\}$

4. $\{A \in \{0,1\}^{n \times n} \mid n \geq 0$ and A is the adjacency matrix of a 3-colorable graph with n 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. $\{\langle M, w \rangle \mid M \text{ accepts } ww^R\}$

8. $\{\langle M \rangle \mid M$ accepts at least one palindrome$\}$

9. $\Sigma^* \setminus \{\langle M \rangle \mid M$ accepts at least one palindrome$\}$

10. $\{\langle M \rangle \mid M$ rejects at least one palindrome$\}$

11. $\{\langle M \rangle \mid M$ accepts exactly one string of length ℓ, for each integer $\ell \geq 0\}$

12. $\{\langle M \rangle \mid L(M)$ has an infinite fooling set$\}$

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.

ACCEPT: $L_{Accept} = \{\langle M, w\rangle \mid M \text{ accepts on } w\}$

HALT: $L_{HALT} = \{\langle M\rangle \mid M \text{ halts on blank input}\}$

HALTONINPUT: $L_{HaltOnInput} = \{\langle M, w\rangle \mid M \text{ halts on } w\}$