This assignment is designed to be an introduction to the concepts used in our study of NP-completeness later in the semester. Some of the questions are trivial; some are subtle. You should have enough background to do these questions.

The senario: You are working as a computer programmer for the Alpha Beta Gamma Software company. It is your dream job.

1 Sorting

Your manager calls you into the office with the following comment:

> Alpha Beta Gamma is becoming a big-data company. Starting next month, every morning we will be receiving a large number of large lists of large values. We will need to sort the lists quickly. Each value in a list is an unsigned integer stored in four words of memory; so each value uses 256 bits ($4 \cdot 64$).
>
> We hired you because you have a Computer Science degree from one of those fancy colleges. Nobody else here at Alpha Beta Gamma Software has your education or ability. You have two weeks to write a lightning fast program to sort these lists.

1. As you leave, you realize that you have no idea how to write such a program because ... [Pick one. If you would like, you can answer for an aquaintance or for a fictional character]:
   (a) You never actually graduated from college. You just pretended to have a degree on your job application.
   (b) You did graduate from college, but you did not major in computer science.
   (c) You majored in computer science, but your algorithms course barely discussed sorting.
   (d) Your algorithms course discussed sorting, but you did not go to class or do the homework.
   (e) You did do the work, but you still did not understand the material.
   (f) You understood the course material, but have long since forgotten it.
   (g) You still remember the course material, but the theoretical material you learned in class has little to do with practical, real life sorting algorithms.
   (h) OTHER: FILL IN THE BLANK.

In any case you have to deal with the situation. You scour the Internet and find a well-regarded sorting package. It is just what you need: It allows you to input a list to be sorted with a parameter for how many words each value uses. You are all set, and can relax for the rest of the month, while pretending you are writing the sorting program.

In order to get exact results, for the remaining problems in this section, ignore the cost of preparing the lists for input into the sorting package or processing the results (assuming those costs are within reason).

2. Assume you are sorting k lists (of four-word values) with sizes (number of values) $n_1, n_2, \ldots, n_k$ where $n = \sum_{i=1}^{k} n_i$.

(a) Assume that the sorting program takes αwm time to sort a list with m values, where each value consists of w words, for some constant α. How long does it take to sort all k lists. Simplify.

(b) Assume that the sorting program takes αwm^2 time. How long does it take to sort all k lists, where each list has size n/k? Simplify.

(c) In general, assume that the sorting program takes αwm^r time, for some constant r. How long does it take to sort all k lists, where each list has size n/k? Simplify.

3. It turns out that your software license only allows you to run the sorting program once per day.

(a) It is too late to write your own program or find a new software package. What should you do, using the software that you have and being limited by one (large) sort per day? If you do any pre or post processing on the data it must be at most linear time in the worst-case. For example, you cannot use a hash table, because that only has good average-case performance.

(b) If it takes αwm time to sort a list of size m, how long does it take to sort all k lists? Simplify. How does this compare with your previous time from Problem 2a?

(c) If takes αwm^2 time to sort a list of size m, how long does it take to sort the k lists, each of size n/k? Simplify. How does this compare with your previous time from Problem 2b?

(d) In general, if takes αwm^r time to sort a list of size m, how long does it take to sort k lists, each of size n/k? Simplify. How does this compare with your previous time from Problem 2c?

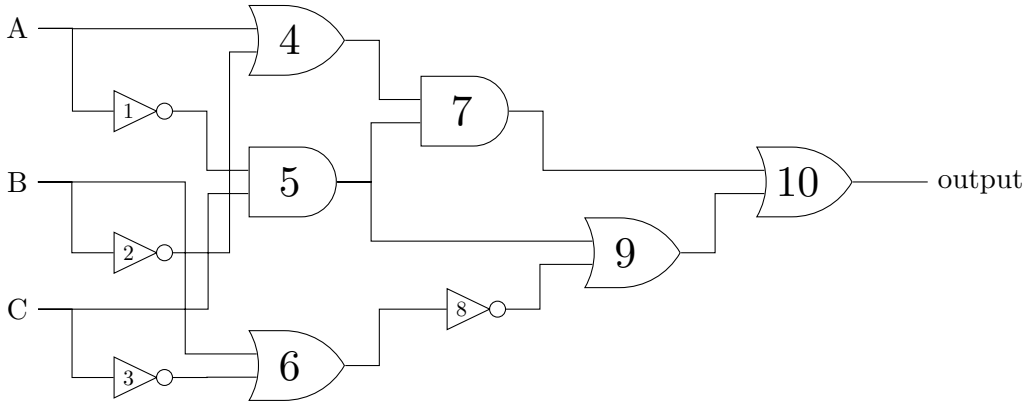
2 Boolean Formula Evaluation

4. (a) Consider the formula

$$((A \wedge \overline{B}) \vee (\overline{A} \wedge C)) \wedge (B \wedge \overline{C})$$

with assignment to the variables $A, B \equiv$ TRUE, $C \equiv$ FALSE. Evaluate the formula. You do not need to show your work.

(b) Consider the Boolean circuit, with assignment to the inputs $A, B \equiv$ TRUE, $C \equiv$ FALSE. Evaluate the Boolean circuit. Show your work by indicating the truth value produced by



each gate. Use the table in the back of this assignment.

Your manager calls you into the office with the following comment:

> We are now moving into the business of Boolean formula evaluation. Starting next month, every morning we will be receiving a large number of large Boolean formulas. For each formula, the assignment of TRUE's and FALSE's to the variables will be given. We will evaluate each formula.
>
> After you did such an impressive job on the sorting program, we at Alpha Beta Gamma believe that we can rely on you to write a lightning fast program to evaluate these Boolean formulas.

Given the last minute difficulties you had with the sorting program, and your not completely satisfactory solution, perhaps your manager's comment about your "impressive job" was sarcastic. Not being sure, you accept the compliment graciously as if it were real.

Of course, you have no idea how to write such a program. You scour the internet but cannot find a satisfactory program to evaluate Boolean formulas. However, you do find a great program to evaluate Boolean circuits.

5. (a) Explain very briefly in English how you would use the Boolean circuit evaluation program to evaluate Boolean formulas.
   (b) Show what you would do on the following example (from above):

$$((A \wedge \overline{B}) \vee (\overline{A} \wedge C)) \wedge (B \wedge \overline{C})$$

6. (a) Assume that the Boolean circuit evaluation program works in linear time $\Theta(n)$, where n is the number of gates and/or wires in the Boolean circuit. How fast can you evaluate a formula of size n, where n is the number of connectives (AND's, OR's, and NOT's)? Justify.
   (b) Assume that the Boolean circuit evaluation program works in time $\Theta(n^r)$, for some constant $r \geq 1$. How fast can you evaluate a formula of size n, where n is the number of connectives (AND's, OR's, and NOT's)? Justify.

7. Assume the scenario is reversed: You need to evaluate Boolean circuits; you have available a great program that evaluates Boolean formulas.
   (a) Explain briefly how to use the Boolean formula program to evaluate Boolean circuits.
   (b) Assume that the Boolean formula evaluation program works in time $\Theta(n^r)$ for some constant $r \geq 1$. How fast can you evaluate a Boolean circuit of size n? Justify.

3 Boolean Formula Satisfiability

A Boolean formula is *satisfiable* if there is an assignment of TRUE's and FALSE's to the variables that makes the formula TRUE. A Boolean circuit with one output wire is *satisfiable* if there is an assignment of TRUE's and FALSE's (or 1's and 0's) to the inputs that makes the output wire TRUE (or 1).

8. This question works with Boolean formulas in Conjuntive Normal Form (CNF). If you are not sure what that is, look it up. A CNF formula is in *k-CNF* if every clause has exactly k literals (for some natural number k).
   (a) Give a 2-CNF formula that is satisfiable, where no variable occurs twice in the same clause. No justification needed.
   (b) Give a 2-CNF formula that is not satisfiable, where no variable occurs twice in the same clause. No justification needed.
   (c) Consider the following 3-CNF formula (with four variables and sixteen clauses):

$$(A \vee B \vee C)(A \vee B \vee \overline{C})(A \vee \overline{B} \vee C)(\overline{A} \vee B \vee \overline{C})(A \vee B \vee \overline{D})(\overline{A} \vee B \vee \overline{D})(A \vee \overline{B} \vee \overline{D})(\overline{A} \vee B \vee D)$$
$$(A \vee C \vee D)(A \vee C \vee \overline{D})(\overline{A} \vee C \vee D)(\overline{A} \vee \overline{C} \vee D)(B \vee C \vee D)(B \vee \overline{C} \vee D)(\overline{B} \vee C \vee D)(\overline{B} \vee \overline{C} \vee \overline{D})$$

   After much struggling, you discover that the formula is satisfiable with the assignment $A, D \equiv$ FALSE and $B, C \equiv$ TRUE. Confirm this by circling exactly one literal in each clause such that this assignment makes the clause TRUE. Use the table in the back of this assignment.

9. Imagine that there is a large Boolean formula (not necessarily in CNF) written on the white-board, where n is the number of connectives (AND's, OR's, and NOT's). Perhaps $n = 200$. You claim that the formula is satisfiable; I claim that it is not.
   (a) What do you have to do to convince me that you are right? How much time (in order notation) do you need (as a function of n)? Note: Only count the time needed to actually show me that you are right, not any time needed to figure out how to do so.
   (b) What do I have to do to convince you that I am right? How much time (in order notation) do I need (as a function of n)? Note: Only count the time needed to actually show you that I am right, not any time needed to figure out how to do so.

   Your manager calls you into the office with the following comment:

   > We are now moving into the business of Boolean formula satisfiability. Starting next month, every morning we will be receiving a large number of large Boolean formulas. For each formula, we will need to determine if it is satisfiable. Note that we do not have to actually find the satisfying assignment; we just need a YES/NO answer for each formula.
   >
   > Once again we need your unique skills. You have two weeks to write a lightning fast program to solve satisfiability for these formulas.

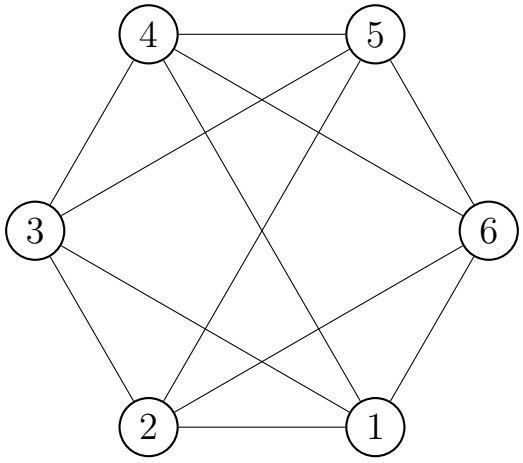
10. Of course, you have no idea how to write such a program. You scour the internet but cannot find a satisfactory program to solve satisfiability for Boolean formulas. However, you do find a great program to solve satisfiability for Boolean circuits.
    (a) Explain very briefly in English how you would use this Boolean circuit program to solve satisfiability for Boolean formulas.
    (b) Assume that the Boolean circuit satisfiability program works in linear time $\Theta(m)$, where m is the number of gates and/or wires in the Boolean circuit. How fast can you determine if a formula with n connectives is satisfiable? Justify.
    (c) Assume that the Boolean circuit satisfiability program works in time $\Theta(m^r)$, where m is the number of gates and/or wires in the Boolean circuit. How fast can you determine if a formula with n connectives is satisfiable? Justify.

4 Graph Coloring

An undirected graph $G = (V, E)$ is *c-colorable* if each vertex can be assigned a color such that at most c colors are used and no two vertices that share an edge have the same color.

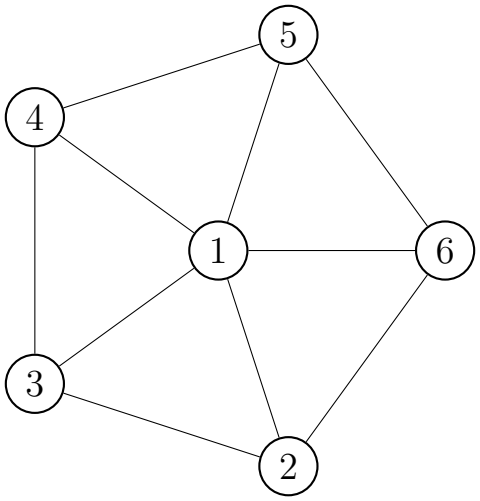
11. You must solve these problems as described. There may be much more clever methods, but we do not care.

    (a) Show that the following graph is 3-colorable, by assigning a color to each vertex and showing that no edge has the same color on its two endpoints.



    The three vertices in triangle (1,2,3) must be different colors. By symmetry, we can set those three colors to Red (R), Blue (B), and Green (G), respectively. This way we can guarantee that there is a unique solution to this problem. Use the two tables provided in the back of this assignment.

    (b) Show that the following graph is not 3-colorable, by showing that each possible assignment of colors to the vertices is not a 3-coloring. This is done by finding an edge in which both



    endpoints have the same color.

There are too many possible colorings to make this practical to do by hand. Once again, the three vertices in the triangle (1,2,3) must be different colors, so by symmetry, we can set those three colors once and for all to, say, Red, Blue, and and Green, respectively. Then we only need to color the remaining three vertices. Use the table in the back of this assignment.

12. Imagine that there is a large undirected graph with n vertices drawn on the whiteboard. Perhaps $n = 200$. You claim that the graph is c-colorable; I claim that it is not.

    (a) What do you have to do to convince me that you are right? How much time (in order notation) do you need (as a function of n and c)? Note: Only count the time needed to actually show me that you are right, not any time needed to figure out how to do so.

    (b) What do I have to do to convince you that I am right? How much time (in order notation) do I need (as a function of n and c)? Note: Only count the time needed to actually show you that I am right, not any time needed to figure out how to do so.

5 Finding Satisfying Assignments

Your manager calls you into the office with the following comment:

> It turns out that not only do we need to determine whether a Boolean formula is satisfiable, but, if so, we need to find a satisfying assignment. Furthermore, the inputs are more complicated than we expected: Boolean formulas can not only have variables, but they can also have TRUE's and FALSE's. So an input might, for example, be
>
> $$((A \wedge (\overline{B} \wedge \text{TRUE})) \vee (\overline{A} \wedge C)) \wedge ((B \wedge \overline{C}) \vee \text{FALSE})$$
>
> Once again we need your unique skills. You have two weeks to write a fast program to determine if a formula is satisfiable, and, if so, to find the satisfying assignment.

You find a program

`satisfiable(`H`)`

that allows TRUE's and FALSE's in the input, H, and returns YES or NO depending on whether the Boolean formula is satisfiable. It runs in time $\Theta(n^r)$, for some constant $r \geq 1$, where n is the number of variables.

13. (a) Show how to use (the Boolean formula satisfiability program) `satisfiable` to efficiently find a satisfying assignment. Write the pseudo-code.
    (b) How fast is your algorithm? Justify.

Your manager calls you into the office, compliments you on a great job, and continues:

> It turns out that not only do we need to find a satisfying assignment, but the assignment must minimize the number of variables set to TRUE.

You find a progam

`satisfiable_num(`H`,`k`)`

that allows TRUE's and FALSE's in the input, H, and returns YES or NO depending on whether the Boolean formula is satisfiable with k or fewer variables set to TRUE. It runs in time $\Theta(n^r)$, for some constant $r \geq 1$, where n is the number of variables.

14. (a) Show how to use (the Boolean formula satisfiability program) `satisfiable_num(`H`,`k`)` to find a satisfying assignment that minimizes the number of variables set to TRUE. Write the pseudo-code.

        HINT: eurtottesebtsumselbairavynamwohenimretedtsrif.

        HINT for hint: sdrawkcabtidaer.

    (b) How fast is your algorithm? Justify.

6 Exercising and Socializing

Alpha Beta Gamma came up with the following clever scheme to make its employees exercise and socialize more: Each day the company creates a group consisting of about 20% of the employees, and assigns each employee in the group a unique group member to visit that day. In this way, each employee in the group would have to take a walk at least once that day, and socialize at least twice (once with the employee visited, and once with the visiting employee). Of course, no pair of employees can be assigned to visit each other. The goal is that the employees visit other employees that they do not already know: there is a constant λ between 0 and 1 so that at least a fraction λ of the visits must be between employees who do not already know each other. Furthermore, in order to guarantee that participants get enough exercise, on average, each employee must walk at least target distance D.

After selecting 20% of the employees, the company checks if it is feasible to arrange the visits to meet the conditions. If not, they swap some employees in the group with employees not in the group, and try again. They continue this process until some group meets the conditions. At that point the company finds an assignment that maximizes the number of visits between employees who do not already know each other (still satisfying the distance requirement).

There are two computational problems that must be solved. Note that n is the number of employees in the group (not in the company).

(1) Given a set of n employees with their distances between offices and their friendships, and a constant λ between 0 and 1, determine if there is an assignment of visits such that each employee walks at least distance D, on average, and at least a fraction λ of the visits are between employees who do not already know each other. This is a YES/NO question.

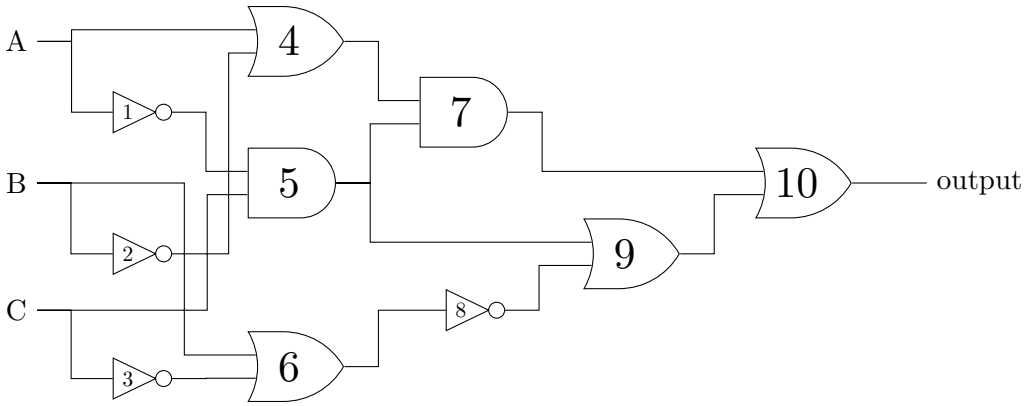
(2) Given a set of n employees with their distances between offices and their friendships, find an assignment of visits such that each employee walks at least distance D, on average, that maximizes the number of visits between employees who do not already know each other. This is an OPTIMIZATION problem.

These two problems seem to be hard to solve efficiently. As usual your manager calls you in and asks you to write programs to solve the two problems. As usual you have no idea how to write such programs. For each problem, you find an efficient program on the Internet that solves that problem. Unfortunately your budget will only allow you to buy one program.

15. Assume that you have a program that solves the *second* problem in time $\Theta(n^r)$, for $r \geq 1$. Can you use it to solve the *first* problem in polynomial time? If so, how, and how fast is your algorithm?

16. Assume that you have a program that solves the *first* problem in time $\Theta(n^s)$, for $s \geq 1$. Can you use it to solve the *second* problem in polynomial time? If so, how, and how fast is your algorithm?

Assignment to the inputs $A, B \equiv$ TRUE, $C \equiv$ FALSE.



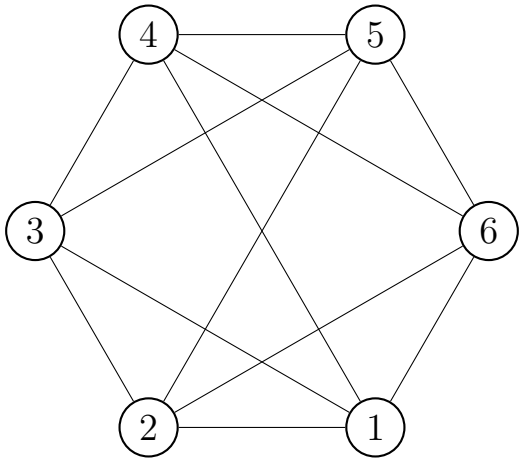
Gate	Output value
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Satisfying assignment: $A, D \equiv$ FALSE and $B, C \equiv$ TRUE.

Clause
$(A \vee B \vee C)$
$(A \vee B \vee \overline{C})$
$(A \vee \overline{B} \vee C)$
$(\overline{A} \vee B \vee \overline{C})$
$(A \vee B \vee \overline{D})$
$(\overline{A} \vee B \vee \overline{D})$
$(A \vee \overline{B} \vee \overline{D})$
$(\overline{A} \vee B \vee D)$
$(A \vee C \vee D)$
$(A \vee C \vee \overline{D})$
$(\overline{A} \vee C \vee D)$
$(\overline{A} \vee \overline{C} \vee D)$
$(B \vee C \vee D)$
$(B \vee \overline{C} \vee D)$
$(\overline{B} \vee C \vee D)$
$(\overline{B} \vee \overline{C} \vee \overline{D})$

Vertex	1	2	3	4	5	6
Color (R,B,G)	R	B	G			

Edge	Color left endpoint	Color right endpoint
(1,2)		
(1,3)		
(1,4)		
(1,6)		
(2,3)		
(2,5)		
(2,6)		
(3,4)		
(3,5)		
(4,5)		
(4,6)		
(5,6)		



123456	Bad edge	Color on endpoints
RBGRRR		
RBGRRB		
RBGRRG		
RBGRBR		
RBGRBB		
RBGRBG		
RBGRGR		
RBGRGB		
RBGRGG		
RBGBRR		
RBGBRB		
RBGBRG		
RBGBBR		
RBGBBB		
RBGBBG		
RBGBGR		
RBGBGB		
RBGBGG		
RBGGRR		
RBGGRB		
RBGGRG		
RBGGBR		
RBGGBB		
RBGGBG		
RBGGGR		
RBGGGB		
RBGGGG		