CMPSC 465 Data Structures & Algorithms
Spring 2021 Paul Medvedev and Chunhao Wang

HW 9

Due April 1st, 10:00 pm

Instructions: You may work in groups of up to three people to solve the homework. You must write your own solutions and explicitly acknowledge up everyone whom you have worked with or who has given you any significant ideas about the HW solutions. You may also use books or online resources to help solve homework problems. All consulted references must be acknowledged.

You are encouraged to solve the problem sets on your own using only the textbook and lecture notes as a reference. This will give you the best chance of doing well on the exams. Relying too much on the help of group members or on online resources will hinder your performance on the exams.

Late HWs will be accepted until 11:59pm with a 20% penalty. HWs not submitted by 11:59pm will receive 0. There will be no exceptions to this policy, as we post the solutions soon after the deadline. However, you will be able to drop the three lowest HW grades.

For the full policy on HW assignments, please consult the syllabus.

1. (0 pts.) Acknowledgements. The assignment will receive a 0 if this question is not answered.

(a) If you worked in a group, list the members of the group. Otherwise, write "I did not work in a group."

(b) If you received significant ideas about the HW solutions from anyone not in your group, list their names here. Otherwise, write "I did not consult without anyone my group members".

(c) List any resources besides the course material that you consulted in order to solve the material. If you did not consult anything, write "I did not consult any non-class materials."

2. (12 pts.) Heaviest edge in a cycle

Let e^* be the heaviest edge in a cycle of an undirected graph G. Prove that e^* cannot appear in any MST of G.

3. (12 pts.) Huffman encoding

We use Huffman's algorithm to obtain an encoding of alphabet $\{a,b,c\}$ with frequencies $\{f_a,f_b,f_c\}$. In each of the following cases, either give an example of frequencies $\{f_a,f_b,f_c\}$ that would yield the specified code, or explain why the code cannot possibly be obtained (no matter what frequencies are).

(a) Code: $\{0,10,11\}$

(b) Code: $\{0,1,00\}$

(c) Code: $\{10,01,00\}$

4. (10 pts.) Cost of a prefix-free encoding

The basic intuition behind Huffman's algorithm is that frequent symbols should have short encodings and infrequent symbols should have long encodings. This intuition also has effect in English: typical words like I, you, is, and, to, from, and so on are short, and rarely used words like velociraptor are longer.

However, words like fire!, help!, and run! are short not because they are frequent, but perhaps because time is precious in situations where they are used.

To make things theoretical, suppose we have a file composed of m different words, with frequencies $f_1, \ldots, f_m$. Suppose also that for the i-th word, the cost per bit of encoding is c_i. Thus if we find a prefix-free code where the i-th word has a codeword of length l_i, then the total cost of the encoding will be $\sum_i f_i \cdot c_i \cdot l_i$.

Give an algorithm that finds the prefix-free encoding of minimal total cost (you don't need to prove the correctness). Analyze the running time of this algorithm (in terms of m).