COMP610 Information Structures Assignment 1

Due at 17:00 on Friday January 26th 2018. Via email to breed@cs.mcgill.ca. Questions 3-7 rely on the material presented in the lecture of Wednesday January 17th. Some of these questions require thought. Do not leave the assignment to the last minute.

1. You may assume the fact that the sum of the heights of the nodes of a tree with n nodes, whose interior nodes all have at least two children, is at least $n(log\ n) - 2n$.

   Show that in the best case, quicksort takes $n(log\ n) + O(n)$ comparisions.

2. Show how to implement our 2-pivot median finding algorithm so that its expected number of comparisions it makes is $(\frac{3}{2} + o(1))n$.

3. Consider the variant of SELECT where the elements are split into $\lceil \frac{n}{3} \rceil$ groups, all but at most one of size three. The median of each group is found by sorting it using at most three comparisions then the median of the medians is found by a recursive call to SELECT. Finally this median of the medians is used as a pivot.

   Give as good upper and lower bounds as you can on the worst case number of comparisions used by this algorithm on inputs of size n.

4. For the group of 3 variant of SELECT shown in the last question show that if we choose the groups of three by repeatedly selecting a random group of three elements from those not yet grouped with each choice equally likely then the expected number of comparisisons made is $O(n)$.

5. Consider the variant of SELECT where the elements are split into $\lceil \sqrt{n} \rceil$ groups each of size $\sqrt{n}$ and the median of each group is found by a recursive call to SELECT anf then the median of the medians is found

by another such recursive call. Finally this median of the medians is used as a pivot.

Give as good upper and lower bounds as you can on the worst case time complexity of this algorithm.

6. Show that the expected number of compairions made by any decision tree algorithm for finding the median is at least $\frac{101(n-1)}{100}$.

7. Challenge question: Unmarked \$100 for best written correct answer. Consider a further variant to the group of 3 version of SELECT. When we recursively call SELECT on the medians of the group, we may not find the median, rather we choose the order statistic of the medians that we find depending on the order statistic of the element of the entire set we are looking for. Show that a clever such choice allows us to solve the problem using $O(n)$ comparisions.