

Biostatistics 615/815  
Problem Set 1  
Due Wednesday September 22

### **Union – Find Algorithms**

1. Write a program that generates random connections between pairs of objects. Estimate how many random connections must be sampled before all objects are connected to each other in a set of 1000 objects.
2. The Weighted Quick Union algorithm generates a tree that summarizes connectability between different objects. When all objects are connected, they all lead to the same “root” object. Develop and implement an algorithm that computes the average distance between each node and the root of the tree.

Modify the program you generated for exercise 1, so that it outputs the average distance between each node and the root of the tree after enough random connections have been generated for all objects to be connected.

What is the average distance between each node and the root of the tree? What is the maximum distance between each node and the root of the tree?

3. Propose a modified version of the Weighted Quick Union algorithm that uses the height of trees (longest path to any node from root) rather than the weight (total number of nodes in tree) to decide where to add the new pointer. Provide your example code.

Which version do you expect to be more efficient? [You can make a theoretical argument, conduct empirical studies or even use the code you developed in problem 2 to justify your answer.]

Please indicate how much time it took you to complete this assignment. This will help me calibrate future assignments.