Worth: 11% **Due:** Monday March 29

For each question, please write up detailed answers carefully. Make sure that you use notation and terminology correctly, and that you explain and justify what you are doing.

1. Consider a list of cities $c_1, c_2, \ldots, c_n$. Assume we have a relation R such that, for any i, j, $R(c_i, c_j)$ is 1 if cities c_i and c_j are in the same province, and 0 otherwise.

(a) If R is stored as a table, how much space does it require?

Solution:
R must have an entry for every pair of cities. There are $\Theta(n^2)$ of these.

(b) Using a Disjoint Sets ADT , write pseudo-code for an algorithm that puts each city in a set such that c_i and c_j are in the same set if and only if they are in the same province. Justify coorectness and running time of your algorithm.

Solution:

```
For i <- 1 to n do
  MAKE-SET(c_i)
  For j <- 1 to i-1 do
    If R(c_j, c_i) then
      UNION(c_j, c_i)
      BreakLoop
    EndIf
  EndFor
EndFor
```

(c) When the cities are stored in the Disjoint Sets ADT, if you are given two cities c_i and c_j, how do you check if they are in the same province?

Solution:

c_i and c_j are in the same province if and only if $FIND-SET(c_i) = FIND-SET(c_j)$.

(d) If we use linked-lists with union-by-weight to implement the union-find ADT, how much space do we use to store the cities?

Solution:

There is one node per city, so the space is $\Theta(n)$.

(e) If we use trees with union-by-rank, what is the worst-case running time of the algorithm from (b) (Hint: the unions from your algorithm probably have a special form). Explain.

Solution:

Whenever we do a union in the algorithm from part (b), the second argument is a tree of size 1. Therefore, all trees have height 1, so each union takes time $O(1)$. The worst-case running time is then $\Theta(n^2)$.

(f) If we use trees without union-by-rank, what is the worst-case running time of the algorithm from (b). Explain. Are there more worst-case scenarios than in (e)?

Solution:

Because of the special case of the unions, union-by-rank does not make a difference for our algorithm. Hence, everything is the same as in part (e).

2. Give a sequence of m MAKE-SET, UNION, and FIND-SET operations, n of which are MAKE-SET, that takes $\Omega(m \log n)$ time when we use union by rank only.

Solution:

Perform the n MAKESET operations. Then, using $2^{\lfloor \log_2 n \rfloor} - 1$ UNION operations, create a binomial tree of degree $\lfloor \log_2 n \rfloor$. Recall that a binomial tree of degree $k+1$ is formed by taking the UNION of any two elements in different binomial trees of degree k. For example, UNION(1,2), UNION(3,4), UNION(1,3) creates a binomial tree of degree 2.

The resulting binomial tree has one node at depth $\lfloor \log_2 n \rfloor$. Say it has value k. Then the last $m - n - 2^{\lfloor \log_2 n \rfloor} + 1$ operations of the sequence are FIND-SET(k). Each of these takes time proportional to $\lfloor \log_2 n \rfloor$. The total time for the entire sequence is bounded below by the time to perform all the FIND-SET operations, which is in $\Omega((m - 2n)(\log n))$. Provided m is sufficiently large, i.e. $m - 2n \in \Omega(n)$, it follows that $\Omega((m - 2n) \log n) = \Omega(m \log n)$, as desired.

3. Consider an implementation of the disjoint-sets ADT using rooted trees, where each node contains one member and each tree represents one set. As discussed in class, each node points only to its parent. Moreover, the root of a tree is the set representative and is its own parent. MAKE-SET(x) simply creates a tree with one node for x. FIND-SET(x) follows parent pointers until the root of the tree containing x is found. UNION(x, y) first finds the roots r_x and r_y of xs and ys trees, respectively, and then sets parent(r_y) = r_x.

a Assume we have performed n MAKE-SET operations and no other operations. Give a sequence of $n1$ UNION operations that cumulatively take time $\Omega(n^2)$. Justify your answer.

b Assume we have performed n MAKE-SET operations and no other operations. Give a sequence consisting rst of $n1$ UNION operations and then n unique FIND-SET operations (i.e., each element is searched for exactly once) such that (i) the UNION operations cumulatively take time $O(n)$ and (ii) the FIND-SET operations cumulatively take time $\Omega(n)^2$). Justify your answer.

4. A bipartite graph $G = (V, E)$ is a graph where V can be partitioned into two sets V_1 and V_2 (one of which may be empty) such that there are no edges between any two nodes in V_1 or between any two nodes in V_2.

1. Show that any graph that contains an odd cycle is not bipartite.

   Solution:

   Let the cycle be $\{v_1, v_2\}, \{v_2, v_3\}, \cdots, \{v_{m-1}, v_m\}, \{v_m, v_1\}$, where m is odd. Assume there is some partition V_1, V_2 like the one described above. Assume without loss of generality that $v_1 \in V_1$. Then it must be the case that $v_2 \in V_2, v_3 \in V_1, \cdots, v_m \in V_1$. But this is a contradiction since there is an edge between v_1 and v_m.

2. Show how to modify DFS so that, given any graph, it creates a partition like the one described above (that is, it assigns to each node either a 1 or a 2) when the graph is bipartite, and it outputs "not bipartite" when the graph is not bipartite.

   Solution:

   For simplicity, we'll assume the graph is connected. This solution refers to the pseudocode for DFS in lecture 8. We'll let $part[u]$ store the partition of each node u (i.e. 1 or 2). When we initialize $p[s]$, we'll initialize $part[s] = 1$. Whenever we set $p[v] = u$, we'll set $part[v] = 3 - part[u]$ (if there is an edge between u and v, then they should be in different partitions). Finally after the (*) line, we can check if $colour[v] == gray$ AND $part[v] == part[u]$. If that check is true, then we output "not bipartite"

3. Is it true that any graph that does not contain an odd cycle is bipartite?

   Solution:

   Yes. If there are no odd cycles, then the algorithm from (b) will successfully partition the graph.