Chapter 5. Greedy Algorithm

Always choosing the next piece that offers the most obvious and immediate benefit

5.1 Minimum spanning trees

- Optimal set of edges cannot contain a cycle

- Property 1. Removing a cycle edge cannot disconnect a graph.

Therefore, minimum spanning tree

Input: an undirected graph G = (V, E) -> output: a tree T = (V, E') that minimizes weight

5.1.1 A greedy approach

Repeatedly add the next largest edge that doesn't produce a cycle

5.1.2 The cut property

Cut: partition of vertices that divide into two non-empty sets

Cut property: for a cut, crossing edge of minimum weight is always MST

5.1.3 Kruskal's algorithm

When weight is arranged in ascending order, start with the edge of the smallest weight

If the edge's endpoints are not connected, choose the edge as the edge of MST

Repeat until number of edges is V-1, then you can get MST

5.1.4 A data structure for disjoint sets

Union by rank

Store a set as a directed tree

The root element is a convenient representative, or name for the set

Parent pointer is a self-loop

Each node has a rank

Makeset is a constant-time operation, find takes time proportional to the hight of the tree

Merging two sets: make the root of one point to the root of the other

∟ Make the root of the shorter tree point to the root of the taller tree

∟ Instead of explicitly computing heights of trees, we will use the rank numbers of their root modes

Property 1. For any x, rank(x) < rank(pi(x)).

Property 2. Any root node of rank k has at least 2^k nodes in its tree.

Property 3. If there are n elements overall, there can be at most n/2^k nodes of rank k.

Path compression

Kruskal's algorithm time: O(|E|log|V|) for sorting edges + O(|E|log|V|) for the union and find operations

How to shorten?

-During each find, change all these pointers so that they point directly to the root of a tree

-Need to look at sequences of find and union operations->barely more than O(1)

-Ranks of all nodes are unchanged by path compression b/c once a node ceases to be a root, it never resurfaces

5.1.5 Prim's algorithm

Alternative to Kruskal's algorithm.

The intermediate set of edges X always forms a subtree, and S is chosen to be the set of this tree's vertices

Difference with Dijkstra's algorithm: in key values by which the priority queue is ordered

5.2 Huffman encoding

Compress files using the minimum heap with the number of characters in the input file

For a character that is used frequently, the smallest bit is allocated, and for a larger bit, it is allocated to a character that is used once or twice

5.3 Horn formulas

-For expressing logical facts and deriving conclusions

-Most objective object is a Boolean variable

-Knowledge about variables represented by two clauses: Implications and negative clauses

-Goal: To determine whether there is a consistent explanation

5.4 Set cover

Given a whole set and subsets of that set, select as few sets as possible from the subsets and select a set so that the union of those sets becomes the original whole set(=the sets ``cover'' the original whole set)