Advanced Algorithms assignment 1

Applied in c++

Ellis Rourke

S5057468

question 1: List storage

Problem

We are tasked with implementing a data structure able to store a list of n numbers and simultaneously keep track of the first k smallest numbers. This problem is easy solved using any sorted tree-based structure such as a min heap or balanced binary search tree.

The list needs to facilitate a very large number of random numbers being entered while still maintaining the k smallest. The target space complexity is O(k+n) although our tree-based solutions will maintain the list in just O(n) space.

Solution

As mentioned above, we could solve this problem with an array of different tree structure, but given our course work… I have chosen to implement a red-black tree.

The red-black tree allows us to insert and delete any number of numbers while maintaining a space complexity of O(n).

The red, black tree has complexities of O(Log n) for Insert, Delete and search allowing us to perform are most crucial operations in an efficient manner. Given this, we can find the first k elements in O(k log n) time.

RESULTS

AMORTISED ANALYSIS

question 2: maze generation

problem

The problem states we are to find a way to programmatically generate a simple maze by randomly knocking down walls in a 50 x 88 grid until a specified entry and exit points are connected: creating a valid solvable maze.

Solution

We can implement use a disjoint set structure to solve this problem. Disjoint sets work by keeping a list of items in unique sets where a set is some number of keys where there is no overlap with another set.

We can make use of this property by implementing the two main methods of a disjoint set, Union and Find. Union is used to join two subsets such that they are a single set and find is used to test if two items are in the same set.

We can solve the problem by first selecting two edge walls at random and breaking their respective walls to create an entry and exit point for the maze.

We then randomly knock down walls (union 2 random walls) until the entry and exit points are connected.

The complexity of the structure is O(Log n) for union and find in the worst case however the amortized cost is much closer to O(n).

RESULTS

Below is an example 5x5 maze generated by the algorithm. As you can see, while not incredibly difficult, the maze is complete in such a way that the start and endpoints are connected (in the same set) and other paths are present that makes the solution not obvious. At a larger scale, the maze would become increasingly difficult to navigate.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

question 3: red black vs veb

Problem

We are tasked with comparing the implementations of both Red Black, and Van Emde Boas trees.

A red, black tree structure provides a self-balancing binary search tree in which every node can be assigned a colour red or back. The tree uses rotations based on predefined rules to maintain a O(Log n) tree height at all times. The tree makes use of batch rotations to maintain efficiency, meaning the tree may not be rotated to an optimal configuration after each insertion or deletion if there is not a tree violation as these will all be corrected when a rotation is triggered.

The Red Black tree is very effective at managing insertions and deletions with relative frequency and given that it is self-balancing, it will maintain a time efficiency of O(Log n) for all operations and maintain this in a range of scenarios.

Alternatively, Veb or Van Emde Boas trees are implemented as an associative array and well outperform Red Black and most other tree structures. The Veb tree has time complexity of O(Log Log n) for most operations and O(1) for min and max. Veb’s are defined recursively and split the data into smaller veb clusters that store information regarding the summary and inclusivity of that given cluster.

When storing a large number of integers, space and time complexity becomes paramount the efficiency of the underlying system, A you can see below, with upto 500000 integers, The Veb tree outperforms a Red Black tree in every operation while with an n value less that 100000 the difference is quite small.

This is because the overhead of a Veb tree can become a burden of efficiency when the amount of data is too small. The overhead associated with the veb structure can be 2^(m/2)

question 4: kevin bacon

problem

The Kevin Bacon problem, or six degrees of bacon is a graph computation problem where we are tasked with linking one actor to another by connections that represent movies that the two actors appear in; It is said that any actor has a bacon number of no more than 6.

solution

We can solve the Kevin bacon problem by building a graph of movies and actors such that we can expand the graph and update the connections of each actor and movie as we load more data from the input file.  
  
Each line of input consists of the actor’s name, and the movie they appeared in. For each line we create and new node for the actor and movie. Unless they already exist in the graph, in this case we update the connections for the node. The resulting graph will resemble the figure below,

Diagram

Description automatically generated

Given the graph is acyclic and connected, we can simply use a modified depth first search algorithm to find a single, or the longest path between any two nodes.

The modification we can make to a standard BFS search lets us use a start and end point to set the two actors in which we are trying to connect. We also implement a findmax flag where we can search for the largest set of connections in the input file.

**Breadth\_first\_search**(graph g, node start, node end, bool findmax):

Define set s

Define queue q

**If** start **or** end == null:  
 **return** #Node does not exist

Q <- start

S <- start

**While** q not empty:

Current = q.front

Last = current

q.pop

**if** current.name == end.name **and** !findmax:

return current

**for** connection in current.connections:

**if** s.count(current.connections[connection] == 0:

s <- current.connections[connections]

q <- current.connections[connection]

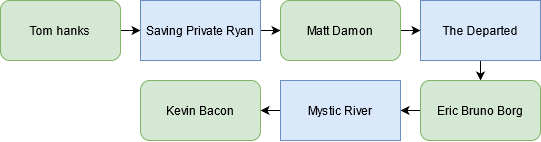
current.connections[connection].parent = current

**Return** last

RESULTS(bacon2.txt)

1. Find actors bacon number

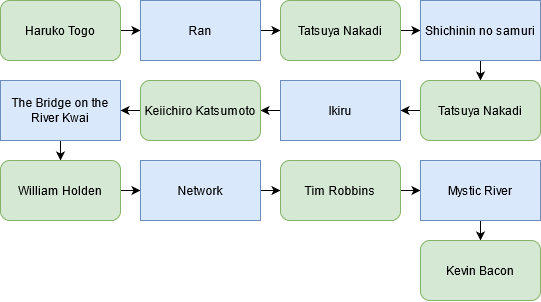
As you can see below, we can connect Kevin Bacon to Tom hanks in just 2 intermediate links, this is a low bacon number; Although, this could be expected as their careers coexisted in the same time frame.



1. Find the actor with the highest bacon number

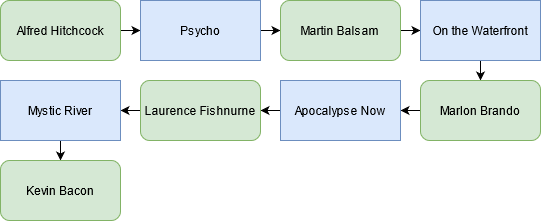
We can find the actor with the largest bacon number by slightly adapting our depth first search.

Here you can see Haruko Togo has a bacon number of 6, this is quite a large gap, although this could be expected as the actors do not have many similarities.



1. Finds the minimum number of links between two arbitrary actors

Here we are essentially doing a Bacon number of 2 given actors that do not necessarily need to be Kevin bacon. Although; in this example I have used Alfred Hitchcock and Kevin Bacon.  
As you can see, they have 4 links between them, this reflects what is expected as there is a significant time gap between their careers.



question 5: min vertex cover

Introduction

The min vertex problem is I widely studied optimization problem where given a graph G, we need to find the smallest set of vertices such that every edge in the graph has one of its endpoints in the cover. Vertex cover is an extensively researched graph theory problem.

A common derivative of this problem is to calculate the minimum number of security cameras that are required to cover every street in a city, where the graph G is the city, streets are edges and vertices are cameras.

The problem is NP-Hard and as such cannot be solved in polynomial time; rather approximation algorithms are required the calculate and search for possible solution, usually with so randomness or heuristic to discover global minimum.

Literature review

Many different algorithms have been implemented to solve the minimum vertex problem; Usually in the form of approximation algorithms as the problem is NP-hard. Detailed below are presented solutions to the problem that exist in current literature. We will cover a range of algorithmic methods ranging from incredibly basic and in optimal solutions to well regarded and efficient algorithmic solutions.

## Algorithm 1 – Brute Force

Kevin Wayne – Princeton University

The brute force approach of any algorithm is a method of finding a solution by checking all possible combinations for an optimal solution; Quite literally using brute force to get the desired solution from a given problem.

Brute force is generally considered the worst possible solution in most cases however they are usually complete as they will explore the entire search space. In the Princeton slides below, it is outlined that a brute force algorithm will work for the MVC however only when the size of k is exceedingly small or constant.

In small graphs of perhaps ~50 vertices or less, the brute force solution may operate with a feasible complexity however as n grows, the complexity will rise exponentially making the approach infeasible for most real applications of the problem as the complexity would follow .

Wayne, K. (2005). 10. Extending Tractability [Slides]. Https://Www.Cs.Princeton.Edu/. https://www.cs.princeton.edu/~wayne/kleinberg-tardos/pdf/10ExtendingTractability-2x2.pdf

## Algorithm 2 – Genetic Algorithm

GA based Approach to Find Minimal Vertex Cover - Sikkim Manipal Institute of Technology

Genetic Algorithms come from the area of study relating to optimization, they essential use natural selection to find the best option from a set of starting configurations, and then iterative choosing the best next solution based on some alteration of the current configuration.

Genetic algorithms could be used for a range of optimization problems like MVC and the TSP problem among many others by randomly generating an initial solution and the attempting to improve it by the mutation process.

The article above outlines an approach to use a GA to solve the mvc by following the proven process of generating an initial population, Defining a fitness function, Selection, Crossover, and mutation. The GA cannot be generalized for any given graph however it is widely accepted that it will always find a global maximum or minimum of the search space.

The time complexity of big o notation does not suit the GA well as there is many internal variables of the selection process that drastically effect the runtime of the algorithm. The big o could be represented as and as you can imagine, the individual runtime of each component would affect the overall complexity however if we set all these variables initially, the time complexity would be linear with relation to the search space.

Kr. Chakraborty, U., Konar, D., & Chakraborty, C. (2014). A GA based Approach to Find Minimal Vertex Cover. International Journal of Computer Applications, 0975(8887).

Algorithm description

The implementation of my algorithm used the depth first search algorithm to search in complement graphs. Firstly, we read the given CLQ file specified in the programs arguments then build our graph structure using an unordered map.

We then create a generic cover by adding all vertices of the graph and poll a number of random vertices to find one with the smallest degree.

When we have our start vertex, we use a generic depth first search algorithm where each time it is called, we remove the vertex with the smallest degree and call another instance of the dfs algorithm using a copy of the previous cover with the node removed. When this is returned, we check if it is smaller than our previous cover.

Our nodes also have a penalty factor attached to them that is used to score certain nodes against others when doing our comparison to find the smallest degree. These penalty factors are also randomly reset during execution.

Cover <- Init

Graph <- Input file

**Depth First Search**

**best cover size** <- Infinity

**for** node **in** graph:

**append** node to cover

**smallest** <- node in graph with smallest degree

remove node with smallest degree from cover

**duplicate cover** <- create a copy of the cover

search for next smallest degree node

add if to the duplicate cover

remove it from the list of options

**for** node **in** duplicate cover

**if** node is removed

recursive dfs

check if new cover is smaller than previous best

reset penalty factor sometimes

conclusioN

Overall, the algorithm developed is able to find a reasonable solution is short period of time.  
My algorithm does not find a vertex cover as optimal as that provide to us in the question brief, however in some cases is it much faster. For example, in the MANN\_a45.clq graph we find a solution of 690 in just 2.233 seconds while the provided results take 88.76 seconds to find a cover of 691; A less optimal solution.

Experimental results

|  |  |  |
| --- | --- | --- |
| Graph | Minimum Vertex Cover | CPU Time (Second) |
| Brock800\_1 | 783 | 0.583 |
| Brock800\_2 | 783 | 5.263 |
| Brock800\_3 | 783 | 5.556 |
| Brock800\_4c | 782 | 19.933 |
| C2000.9 | 1946 | 46.833 |
| C4000.5 | 3988 | 2.233 |
| MANN\_a45 | 690 | 0.282 |
| P\_hat1500-1 | 1492 | 0.259 |