

CS180 Homework 3

Zhehao Wang 404380075 (Dis 1B)

Apr 16, 2016

1 Strongly connected components in a directed graph

(a) Prove SCC graph is a DAG.

Proof by contradiction: assume that a path s_i, \dots, s_j, s_i exists in the SCC graph, where s_k ($i \leq k \leq j$) are the newly created distinct SCC nodes.

There exists a path from any node p_i in SCC s_i to any node p_{i+1} in SCC s_{i+1} , since there exists a node q_i in SCC s_i that has a directed edge to a node q_{i+1} in SCC s_{i+1} , and there exists a path from p_i to q_i , q_{i+1} to p_{i+1} .

Applying the above conclusion repeatedly till we reach s_j , we have the conclusion there exists a path from any node p_i in SCC s_i to any node p_j in SCC s_j . Similarly, we have there exists a path from any node p_j in SCC s_j to any node p_i in s_i . Thus the nodes p_i and p_j should belong to the same SCC node which is the combination of SCC s_i and s_j , and contradicts with the nodes being distinct in the assumption. And we have the directed SCC graph is acyclic, which makes it a DAG by definition.

(b) The algorithm is given in alg 1.

Time complexity: $O(|E|)$

Correctness:

2 Longest path in DAG

(a) algorithm is given in alg ??.

Time complexity: this algorithm is $O(|E|)$.

Correctness: this algorithm does a similar thing as topological sort, except that it counts the steps needed to remove all the nodes, instead of labeling each removed nodes. The recursive idea is that length of longest path in a DAG G is $1 + \text{length of longest path in } G'$, where G' is the remaining graph after all sources in G are removed.

(b) algorithm is given in alg ??. Similar with the idea of problem (a), we base the algorithm on topological sort. We associate a weight with each node, and each time the algorithm removes a source node i , the nodes j that it connects to will have new weight $w(j) = \max(w(j), \text{weight}(i, j) + w(i))$. The largest weight in the DAG will be returned as the length of the longest path.

Time complexity: this algorithm is $O(|E|)$.

Correctness:

(c) algorithm is given in alg 4. Similar idea as in (b).

Time complexity: this algorithm is $O(\max(|E|, |V| \log |V|))$.

Correctness:

3 Optimal order of files

Algorithm is given in alg 5, which is a greedy algorithm that orders files as $f_{t_1} \dots f_{t_n}$, such that for each $t_i > t_j$, $\frac{p_{t_i}}{l_{t_i}} \geq \frac{p_{t_j}}{l_{t_j}}$.

Time complexity: this algorithm is $O(n \log n)$, where n is the number of files.

Correctness: proof by the algorithm's greedy choice property and optimal substructure property.

Greedy choice property: there exists an optimal solution $S = f_{t_1} \dots f_{t_n}$, such that $\frac{p_{t_1}}{l_{t_1}}$ is the largest among all jobs.

Proof: suppose S' is an optimal solution with the sequence $f_{t'_1} f_{t'_2} \dots f_{t'_{k-1}} f_{t_1} f_{t_{k+1}} \dots f_{t_n}$ where $t'_1 \neq t_1$. The average access time of S' is

$$t(S') = \sum_{i=2}^n ((\sum_{j=1}^{i-1} l_{t'_j}) \cdot p_{t'_i})$$

Consider the solution S'' with f_{t_1} and $f_{t_{k-1}}$ swapped, the difference between average access time of S'' and that of S' is

$$t(S') - t(S'') = p_{t_1} * l_{t_{k-1}} - p_{t_{k-1}} * l_{t_1} = (\frac{p_{t_1}}{l_{t_1}} - \frac{p_{t_{k-1}}}{l_{t_{k-1}}}) \cdot l_{t_1} \cdot l_{t_{k-1}} \geq 0$$

Similarly, starting from $t(S'')$, each time we swap f_{t_1} with its previous file, we'll have a smaller or equal access time than before. Thus the optimal solution should contain f_{t_1} as its first element, whose $\frac{p_{t_1}}{l_{t_1}}$ is the largest.

Optimal substructure property: let $S = f_{t_1} \dots f_{t_n}$ be an optimal solution, then $S_1 = f_{t_2} \dots f_{t_n}$ is the optimal solution for the sub problem without f_{t_1} .

Proof by contradiction: assume there's a better solution $S'_1 = f_{t'_2} \dots f_{t'_n}$, $t(S'_1) < t(S_1)$ for the sub problem without f_{t_1} . Then $S' = f_{t_1} S'_1$ has the average access time of $t(S') = t(S'_1) + l_{t_1} * (1 - p_{t_1}) < t(S_1) + l_{t_1} * (1 - p_{t_1}) = t(S)$, which contradicts with S being the optimal solution for the original problem.

With both properties, the greedy algorithm in question is correct.

4 Sorting from SC

Suppose that $SC(p_i, d_i)$ takes in n jobs, and each has a p_i and d_i . SC returns the optimal list of jobs, whose array of deadlines can be extracted by `jobs.deadlines`. The sorting algorithm is described in alg 6.

Time complexity: SC is $o(n \log n)$, and other calls are $O(1)$, so the overall *sort* is $o(n \log n)$.

Correctness: suppose that the algorithm gives back an array $S = d_1 \dots d_n$, we want to prove for each $i < j$, $d_i \leq d_j$.

We start by proving that the first element d_1 is a smallest element.

Suppose we have a smallest element d_k in S , $d_k \leq d_i, 1 \leq i \leq n$. Let the total lateness of S be $t(S)$. Consider the sequence S' with d_k and d_{k-1} swapped, its difference in lateness with $t(S)$ is $t(S) - t(S') = \sigma_{i=1}^{k-1} d_i - \sigma_{i=1}^{k-2} d_i - d_k = d_{k-1} - d_k \geq 0$, and we have $t(S) \geq t(S')$. Since $t(S)$ is given back by SC , and satisfies $t(S) \leq t(S_i)$, where S_i is any schedule, we have $t(S) = t(S')$ and $d_{k-1} = d_k$.

Similarly, starting from S' , we repeatedly switch d_k with its previous element, and can prove $d_1 = \dots = d_{k-1} = d_k$. Thus d_1 is a smallest element.

Consider the array S_1 with d_1 removed, using the above described process we can prove that d_2 is a smallest element in S_1 . We continue until the array S is exhausted, and we have for each $i < j$, $d_i \leq d_j$.

Algorithm 1 SCC building algorithm

```
1: function DFS(v, do_label, smallest_node, node_remaining, node_removed, SCC_graph)
2:   v.visited  $\leftarrow$  true
3:   if do_label = false then
4:     node_remaining.remove(v)
5:     node_removed.add(v)
6:     if v = smallest_node then
7:       smallest_node  $\leftarrow$  node_remaining.nextSmallest()
8:   for  $\{i | (v, i) \in E, i.visited = false\}$  do
9:     if do_label = true then
10:      DFS(i, do_label, smallest_node, node_remaining, node_removed, SCC_graph)
11:    else
12:      if i  $\in$  SCC_graph then
13:        SCC_graph.addEdge(v, i)
14:      else
15:        DFS(i, do_label, smallest_node, node_remaining, node_removed, SCC_graph)
16:   if do_label then
17:     id  $\leftarrow$  label(v)
18:     if id < smallest_node then
19:       smallest_node  $\leftarrow$  v
20: function GETSCC(G)
21:   smallest_node  $\leftarrow$  nil
22:   DFS(G.firstNode(), true, smallest_node, G.nodes, [], nil)
23:   smallest_node_copy  $\leftarrow$  smallest_node.copy()
24:   G.resetVisited()
25:   SCC_graph  $\leftarrow$  nil
26:   while G.node_count > 0 do
27:     G.resetVisited()
28:     node_removed  $\leftarrow$  []
29:     DFS(smallest_node, false, smallest_node_copy, G.nodes, node_removed, SCC_graph)
30:     smallest_node  $\leftarrow$  smallest_node_copy
31:     SCC_graph.addNode(node_removed)
32:   return SCC_graph
```

Algorithm 2 Longest path in an unweighted DAG

```
1: function LONGESTPATH( $G$ )
2:    $nodeCount \leftarrow len(V)$ 
3:    $sourceNodes \leftarrow []$ 
4:   for  $\{i | i \in V\}$  do
5:     if  $i.inDegree = 0$  then
6:        $sourceNodes.add(i)$ 
7:    $length \leftarrow 0$ 
8:   while  $nodeCount > 0$  do
9:      $newSourceNodes \leftarrow []$ 
10:    for  $\{i | i \in sourceNodes\}$  do
11:      for  $\{v | (i, v) \in E\}$  do
12:         $v.inDegree \leftarrow v.inDegree - 1$ 
13:        if  $v.inDegree = 0$  then
14:           $newSourceNodes.add(v)$ 
15:       $G.remove(i)$ 
16:       $nodeCount \leftarrow nodeCount - 1$ 
17:       $sourceNodes \leftarrow newSourceNodes$ 
18:       $length \leftarrow length + 1$ 
19:   return  $length$ 
```

Algorithm 3 Longest path in a weighted DAG

```
1: function LONGESTPATH( $G$ )
2:    $nodeCount \leftarrow len(V)$ 
3:    $sourceNodes \leftarrow []$ 
4:   for  $\{i | i \in V\}$  do
5:     if  $i.inDegree = 0$  then
6:        $sourceNodes.add(i)$ 
7:        $i.length \leftarrow 0$ 
8:    $maxLength \leftarrow min\_real$ 
9:   while  $nodeCount > 0$  do
10:     $newSourceNodes \leftarrow []$ 
11:    for  $\{i | i \in sourceNodes\}$  do
12:      for  $\{v | (i, v) \in E\}$  do
13:         $v.inDegree \leftarrow v.inDegree - 1$ 
14:         $v.weight \leftarrow \max(v.weight, i.weight + w(v, i))$ 
15:        if  $maxLength < v.weight$  then
16:           $maxLength \leftarrow v.weight$ 
17:        if  $v.inDegree = 0$  then
18:           $newSourceNodes.add(v)$ 
19:       $G.remove(i)$ 
20:       $nodeCount \leftarrow nodeCount - 1$ 
21:       $sourceNodes \leftarrow newSourceNodes$ 
22:   return  $maxLength$ 
```

Algorithm 4 Weighted DAG job scheduling

```
1: function LONGESTPATH( $G$ )
2:    $nodeCount \leftarrow len(V)$ 
3:    $sourceNodes \leftarrow []$ 
4:   for  $\{i | i \in V\}$  do
5:     if  $i.inDegree = 0$  then
6:        $sourceNodes.add(i)$ 
7:        $i.weight \leftarrow 0$ 
8:   while  $nodeCount > 0$  do
9:      $newSourceNodes \leftarrow []$ 
10:    for  $\{i | i \in sourceNodes\}$  do
11:      for  $\{v | (i, v) \in E\}$  do
12:         $v.inDegree \leftarrow v.inDegree - 1$ 
13:         $v.weight \leftarrow \max(v.weight, i.weight + w(i, v))$ 
14:        if  $v.inDegree = 0$  then
15:           $newSourceNodes.add(v)$ 
16:         $G.remove(i)$ 
17:         $nodeCount \leftarrow nodeCount - 1$ 
18:       $sourceNodes \leftarrow newSourceNodes$ 
19:   return  $sort(v.weight)$ 
```

Algorithm 5 Optimal order of files

```
1: function OPTIMALORDER( $files$ )
2:    $weightedFiles \leftarrow []$ 
3:   for  $\{i | i \in files\}$  do
4:      $weightedFiles.add(\frac{p_i}{l_i})$ 
5:   return  $sort(weightedFiles)$ 
```

Algorithm 6 Sorting using SC

```
1: function SORT( $array$ )
2:    $jobs \leftarrow SC(array, array)$ 
3:   return  $jobs.deadlines$ 
```
