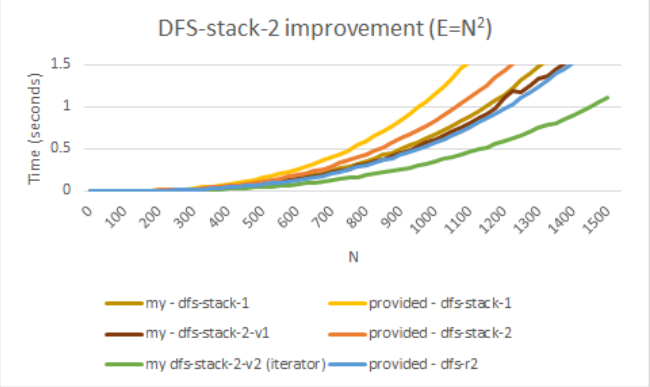
TheTips performance followup / improvements

I was intrigued as to why the DFS-R2 performed better than DFS-Stack-2 so I set out to improve DFS-Stack-2.



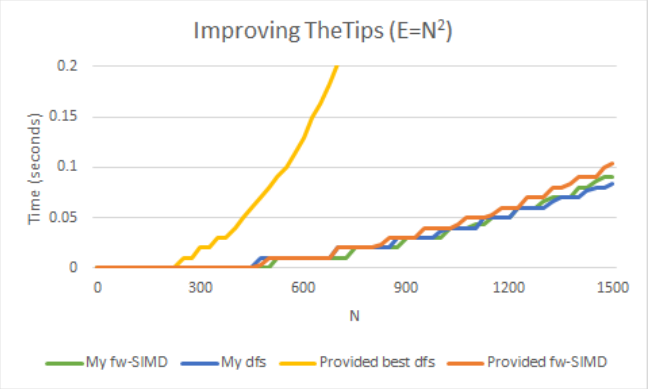
The first improvement was to reduce the additions to the stack. This can be done my setting a node as visited when adding it to the stack and removing the check when popping from the stack (v1 on the above chart).

```
visited[Adj[j][i]] = 1;
```

This can be further improved by using an iterator to traverse the adjacency list (v2 on the above chart).

```
for (Adjji = Adj[j].begin(); Adjji != Adj[j].end(); ++Adjji)
```

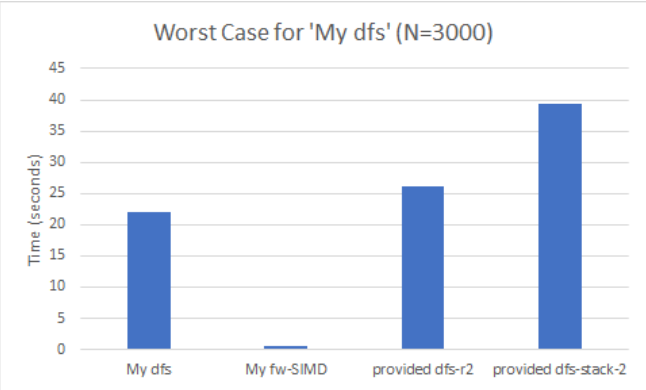
Not yet satisfied with this performance I decided to continue by improving the algorithm.



For large values of E (relative to N^2) the DFS algorithm will spend most of the cpu-time checking if it can reach nodes that have already been visited. By counting the number of nodes visited, the loop to add more nodes can be skipped when all nodes have been visited. This greatly improves the average performance. The above chart shows this implementation outperforming not only the provided DFS implementations but also the provided fw-SIMD implementation. This change can be implemented as follows:

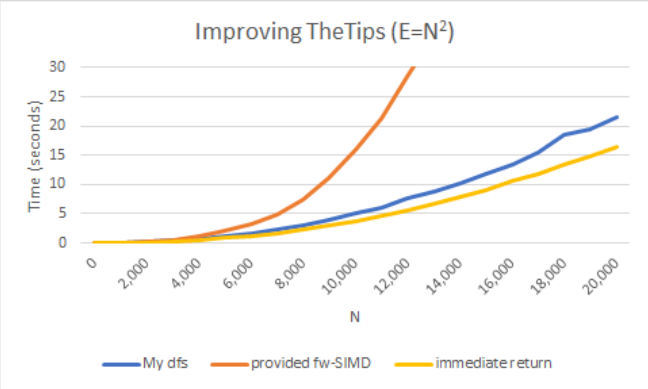
```
visCount = 1;
while (sp != 0) {
  j = stack[--sp];
  p[j] += ((1 - p[j])*pr[v]);
  if(visCount != Adj.size()){
    tmpSP = sp;
    for (Adjji = Adj[j].begin(); Adjji != Adj[j].end(); ++Adjji) {
      if (!visited[*Adjji]){
        visited[*Adjji] = 1;
        stack[sp++] = *Adjji;
      }
    }
    visCount += (tmpSP - sp);
  }
}
```

This improvement reduces the average complexity to $O(N^2)$. However, there are still limitations due to the nature of DFS. If a node is not connected or edges linking to it are not processed early, performance reverts to previous implementations (worst case is still $O(N^2 + NE)$). This is unlikely from randomly generated input but could happen. The following chart shows a hand crafted example at N=3000:



In this example all but one of the 3000 nodes are completely connected and the last node is only connected by one edge that has been placed such that it is processed last.

For reference, a comparison of immediately returning when entering the solve function instead of calculating the correct answer:



This shows the limiting factor when $E = N^2$ is no longer the calculation when using the provided main function.

lab5 lectures

~ An instructor (James S Plank) thinks this is a good note ~

Updated 4 days ago by Gregory Rouleau

followup discussions *for lingering questions and comments*