Fault Tolerant and Fully Dynamic DFS in Undirected Graphs: Simple Yet Efficient

Surender Baswana
Department of Computer Science & Engineering, IIT Kanpur, Kanpur, India
sbaswana@cs.iitk.ac.in

Shiv Gupta
Department of Computer Science & Engineering, IIT Kanpur, Kanpur, India
shivguptamails@gmail.com

Ayush Tulsyan
Department of Computer Science & Engineering, IIT Kanpur, Kanpur, India
ayushtulsyan01@gmail.com

Abstract
We present an algorithm for a fault tolerant Depth First Search (DFS) Tree in an undirected graph. This algorithm is drastically simpler than the current state-of-the-art algorithms for this problem, uses optimal space and optimal preprocessing time, and still achieves better time complexity. This algorithm also leads to a better time complexity for maintaining a DFS tree in a fully dynamic environment.

2012 ACM Subject Classification Theory of computation → Dynamic graph algorithms

Keywords and phrases Depth first search, DFS, Dynamic graph algorithms, Fault tolerant

Digital Object Identifier 10.4230/LIPIcs.MFCS.2019.65


Funding Surender Baswana: This research work was partly done during a visit to University of Paderborn supported by a fellowship from the Alexander von Humboldt Foundation.

1 Introduction

Depth First Search (DFS) is a widely popular graph traversal method. The traversal routine, formalized by Tarjan [44] in 1972, has played a crucial role in various graph problems including reachability, bi-connectivity, topological sorting, and strongly connected components.

Given an undirected graph $G = (V, E)$ with $n = |V|$ vertices and $m = |E|$ edges, DFS traversal on the graph takes $O(m + n)$ time and results in a DFS tree of $G$.

Most of the graphs in real-world applications keep changing with time. Vertices and edges keep entering and leaving the graph at various time steps. This dynamic aspect has motivated researchers to design algorithms that can update the solution of the corresponding problem efficiently after each such change in the graph. There are two models used for solving these graph problems, namely, fault tolerant algorithms and dynamic graph algorithms.

The fault tolerant version of any problem $\mathcal{P}$ on a graph $G$ is to construct a compact data structure, using which, for any given set of failed edges or vertices $F$, one can efficiently report the solution of $\mathcal{P}$ on $G \setminus F$. Many elegant fault tolerant algorithms have been designed for problems including connectivity [13, 21, 25], shortest paths [10, 11, 16, 20, 28], and spanners [12, 15].

The dynamic version of any problem $\mathcal{P}$ on $G$ is modeled as follows. For any online sequence of updates (insertion or deletion of an edge/vertex), one has to report the solution of $\mathcal{P}$ efficiently after every update. Note that, unlike the fault tolerant version, the updates are persistent in dynamic version, i.e., after each update, the solution has to be reported.
taking into account all the updates made so far. Algorithms which handle both insertion and deletion of vertices/edges are called fully dynamic graph algorithms, whereas the algorithms that handle either insertions or deletions are called partially dynamic graph algorithms, more specifically incremental or decremental graph algorithms, respectively. The prominent results for dynamic graph problems include connectivity [22, 30, 32], reachability [39, 41], shortest path [19, 29, 40], matching [6, 9, 43], spanner [8, 26, 38], min cut [45], Even-Shiloach tree [23], minimum spanning tree [31, 35], and graph sparsifiers [1].

1.1 Previous Results on Fault Tolerant and Dynamic DFS

Franciosa et al. [24] presented an incremental algorithm for maintaining a DFS tree in a directed acyclic graph (DAG) which takes overall $O(mn)$ time for any sequence of $m$ edge insertions. For undirected graphs, Baswana and Khan [7] presented an incremental algorithm for a DFS tree which takes overall $O(n^2)$ time for any sequence of edge insertions. Baswana and Choudhary [4] designed a randomized decremental algorithm for a DFS tree in a DAG with overall expected $O(mn \log n)$ time for any sequence of edge deletions.

None of the partially dynamic algorithms stated above achieves an $o(m)$ bound over the worst-case complexity of a single update. Moreover, there were no fully dynamic or fault tolerant algorithms for DFS in undirected graphs until recently. In 2016, the first fault tolerant algorithm was presented that takes $O(nk \log^4 n)$ time to report a DFS tree for any set of $k$ failed vertices or edges [3]. The time complexity was further improved by Chen et al. [17] to $O(nk \log^2 n)$. Both [3, 17] require a data structure occupying $O(m \log^2 n)$ bits. Nakamura and Sadakane [34] reduced the space occupied by the data structure to $O(m \log n)$ bits which is indeed optimal \(^1\). Using the standard technique of periodic rebuilding, the fault tolerant algorithms presented in [3, 17, 34] were also extended to $o(m)$ fully dynamic DFS algorithms (refer to Table 1 for comparison).

Recently, Chen et al. [18] designed an $O(n)$ time incremental algorithm for DFS tree in undirected graphs, which is optimal if it is required to output the DFS tree after each update. For the hardness results for the dynamic ordered DFS problem, the reader may refer to the work of Reif [36, 37] and Miltersen et al. [33].

1.2 Familiarizing with the Fault Tolerant DFS Problem

For an undirected graph, DFS traversal results in a spanning tree rooted at the vertex from where the DFS begins. The depth first nature of the traversal ensures the following property:

\begin{itemize}
  \item \textbf{Property 1.} (DFS Property) For each vertex $v \in V$, every neighbour of $v$ in the graph appears either as an ancestor of $v$ or as a descendant of $v$ in any DFS tree.
\end{itemize}

The DFS property implies that a non-tree edge is never a cross edge - an edge whose endpoints do not share an ancestor-descendant relationship. It is due to this reason that each non-tree edge is called a back edge. We now define ancestor-descendant paths.

\begin{itemize}
  \item \textbf{Definition 2.} (Ancestor-descendant Path) A path in a DFS tree is called an ancestor-descendant path if its endpoints have an ancestor-descendant relationship in the tree.
\end{itemize}

In order to familiarize with the problem of fault tolerant DFS tree, we discuss another related, but simpler problem, namely, rerooting of a DFS tree defined as follows.

\footnote{Precisely, their data structure occupies $(m + o(m)) \log n$ bits by using a wavelet tree.}
Problem 3. Preprocess an undirected graph $G = (V, E)$ to build a compact data structure so that given any vertex $v \in V$, we can report the DFS tree rooted at $v$ efficiently.

Let $T$ be an initial DFS tree, rooted at a vertex, say $r$ (see Figure 1(a)). We use $T(q)$ to represent the subtree of $T$ rooted at the vertex $q$. Also, let $path(a, b)$ denote the path from vertex $a$ to $b$ in $T$. For computing a DFS tree rooted at any vertex, the first natural idea is to just reverse the direction of the path from $r$ to $v$ in $T$. However, this may result in transforming many back edges to cross edges and hence a violation of the DFS property (see Figure 1(b)). To fix this problem, we may need to reroot various subtrees hanging from the reversed path. Along these lines, [3] presented an algorithm that takes $O(n \log^3 n)$ time to compute a DFS tree rooted at any vertex.

In order to see how rerooting a DFS tree is related to the problem of fault tolerant DFS tree, consider the failure of vertex $u$ in Figure 1(a). The subtree $T(v)$ is connected to the remaining tree through many back edges, and the back edge $(p, t)$ is incident closest to $u$ on the path $(r, u)$. If we reroot the subtree $T(v)$ at vertex $p$ and hang it from the remaining tree through the edge $(p, t)$, the resulting tree will indeed be a valid DFS tree of $G \setminus \{u\}$.

The fault tolerant DFS tree problem becomes more complex in the presence of multiple faults. However, the distribution of faults plays an important role as follows. If the failing vertices do not have any ancestor-descendant relationship in the DFS tree, they can be handled independently. For example, the simultaneous failure of vertices $u$ and $x$ in the DFS tree shown in Figure 1(a) requires rerooting the respective subtrees $T(v)$ and $T(y)$ at vertices $p$ and $z$ respectively. So, even for arbitrarily large number of faults, if no two of them appear on the same root to leaf path, we have to just reroot the corresponding disjoint subtrees of $T$ to report the DFS tree avoiding those failures. However, if two or more faults indeed appear on a single root to leaf path, the problem becomes more complex. For this case, [3] presents an algorithm which is quite different from their rerooting algorithm.

1.3 Overview of the Previous Results

We now begin with an overview of the existing algorithms for fault tolerant DFS tree. For any set of $k$ failures, the algorithm presented in [3] first partitions the original DFS tree into a pool of connected components. This pool consists of $k$ paths (specifically, ancestor-descendant paths) and potentially a large number of subtrees. The algorithm treats each
of these components as a super vertex and uses them to grow the DFS tree $T^*$ that avoids all the failures. At a high level, the algorithm can be visualized as a traversal on these super vertices. Each traversal extracts a path from the super vertex, attaches it to $T^*$, and places the remaining portion of the super vertex back into the pool. In order to pursue DFS traversal further in an efficient manner, the algorithm needs to compute minimal adjacency lists for the vertices of the traversed path (referred to as reduced adjacency lists). The algorithm makes use of the following crucial property of DFS traversal.

**Property 4. (Components Property [3])** Consider any DFS Traversal on any undirected graph $G = (V, E)$. When the traversal reaches a vertex $v \in V$, let the set of connected components induced by the unvisited vertices be $C$. If from any component $c \in C$, there exists two edges - $e$ to vertex $v$ and $e'$ to any of the visited vertices (including $v$), then for building a valid DFS tree, it is sufficient to consider only the edge $e$ during the rest of DFS traversal, that is, $e'$ can be ignored.

In order to use the above property to populate the reduced adjacency list, the algorithm needs a data structure to answer the following queries repeatedly.

- **Query** $(w, x, y)$: among all the edges from $w$ that are incident on the path $(x, y)$ in $T$, return an edge that is incident nearest to $x$ on the path $(x, y)$.
- **Query** $(T(w), x, y)$: among all the edges from $T(w)$ that are incident on the path $(x, y)$ in $T$, return an edge that is incident nearest to $x$ on the path $(x, y)$.

It is quite obvious from the description given above that these queries are quite non-trivial, and so a sophisticated data structure is designed in [3] to answer these queries efficiently. In addition to the complex data structure, the complete difference in the processing of a path and a subtree obfuscates the algorithm and its analysis.

The subsequent results [17, 34] keep the algorithm unchanged and replace the data structure used in [3] with alternate data structures. Chen *et al.* [17] model the two queries mentioned above as Orthogonal Range Successor/Predecessor (ORS/ORP) queries and this improves the query processing time. Nakamura and Sadakane [34] compressed the data structure used in [17] using Wavelet Trees [27] to achieve optimal space. Despite these improvements, the core of the fault tolerant algorithm remains intricate and the data structure still remains complex. Recently, in an empirical study [5], it was found that this algorithm, for incremental updates, performs even worse than the static DFS algorithm for certain classes of graphs. This naturally raises the question whether there exists a simpler algorithm for this fundamental problem.

In addition to being complex, all these algorithms fail to exploit the distribution of the faults to achieve efficiency. The running time of these algorithms is $O(nk \text{ polylog}(n))$ time irrespective of how the $k$ faults are distributed in the DFS tree. In the extreme case, when no two faults share the ancestor-descendant relationship, there is a simple $O(n \text{ polylog}(n))$ time algorithm as described in Section 1.2. This raises the question whether it is possible to have a fault tolerant DFS algorithm whose time complexity depends upon the maximum number of faults lying on any ancestor-descendant path instead of the total number of faults.

### 1.4 Our Contribution

We take a much simpler approach as compared to the previous algorithms. We first present a new and simple rerooting algorithm based on the following ideas. After building an initial DFS tree, say $T$, we decompose $T$ into a disjoint collection $\mathcal{P}$ of ancestor-descendant paths. Similar to [3], each of these paths are treated like super vertices. At a high level, the algorithm can still be viewed as a traversal on these super vertices. However, as the reader may also...
verify, the algorithm turns out to be lighter and quite different at the core. Interestingly, the original DFS tree alone acts as a powerful data structure to be used for rerooting or for the computation of another valid DFS tree in the presence of faults. The algorithm crucially exploits an implicit hierarchy among the ancestor-descendant paths in $P$. This hierarchy along with the DFS property of $T$ enables us to use much simpler queries. In particular, each query will ask only for an edge from a vertex to one of its ancestor paths in the hierarchy. The hierarchy allows us to represent $T$ as another tree structure, called shallow tree. In a nutshell, our algorithm can be viewed as an efficient DFS traversal guided by this shallow tree.

This rerooting algorithm extends to the fault tolerant algorithm with very little and obvious modifications. While preserving simplicity, the fault tolerant algorithm turns out to be faster than all the previous algorithms. Moreover, our algorithm is the first to implicitly incorporate the distribution of faults to gain efficiency. We summarize our result in the following theorem.

\begin{theorem}
An undirected graph $G$ can be preprocessed in $O(m + n)$ time to build a DFS tree, say $T$, and a data structure of $O(m + n)$ words\(^2\) such that for any set $F$ of $k$ failed vertices or edges, a DFS tree of $G \setminus F$ can be reported in $O(n (k' + \log n) \log n)$ time, where $k' \leq k$ is the maximum number of faults on any root-leaf path in the tree $T$.
\end{theorem}

We now present the highlights of our algorithm.

Drastically simpler algorithm: Our algorithm is drastically simpler and more intuitive than the previous algorithm. We feel confident to defend that it can be taught even in an undergraduate course on algorithms. The pseudo-codes in Algorithm 1 and Algorithm 2 are concise and very close to the corresponding implementations.

Faster time complexity: Our algorithm takes $O(n (k' + \log n) \log n)$ time, where $k'$ is the maximum number of failures on any ancestor-descendant path of the DFS tree when $k$ edges/vertices fail. In the worst-case $k'$ can be as large as $k$. However, $k'$ can be $o(k)$ as well. In the latter case, our result improves all the existing results significantly. Moreover, even in the case $k' = k$, our time complexity is superior to the previous best by a log factor.

Optimal preprocessing time: Our preprocessing relies upon DFS traversal only, taking $O(m + n)$ time. Given a graph, in order to report the initial DFS tree, one anyway has to run a static DFS. Hence, our preprocessing time is optimal.

Optimal space and elementary data structure: In contrast to the heavy data structure used by [3, 17], our algorithm makes use of very elementary data structures which are compact as well. Each vertex keeps an array storing edges incident on it from ancestors sorted according to their levels. This data structure uses just $m + n$ words and still achieves $O(n (k' + \log n) \log^2 n)$ time to report a DFS tree upon failure of any $k$ vertices or edges. By using fractional cascading [14], we get rid of one log factor while still keeping space requirement to be $O(m + n)$ words.

Faster Fully Dynamic Algorithm: Using Theorem 5 and periodic rebuilding technique used in [3], we also get the fastest algorithm for fully dynamic DFS.

\begin{theorem}
Given an undirected graph, one can maintain a DFS tree for any online sequence of insertions and deletions of vertices/edges in $O(\sqrt{mn \log n})$ worst-case time per update.
\end{theorem}

\(^2\) One word stores $\lceil \log m \rceil$ bits.
Table 1 Comparison of the existing and the new results. Note that $k$ is the total number of faults, whereas $k'$ is the maximum number of faults lying on any root-to-leaf path of the DFS tree.

<table>
<thead>
<tr>
<th></th>
<th>[3]</th>
<th>[17]</th>
<th>[34]</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space (in bits)</td>
<td>$O(m \log^2 n)$</td>
<td>$O(m \log^2 n)$</td>
<td>$O(m \log n)$</td>
<td>$O(m \log n)$</td>
</tr>
<tr>
<td>Preprocessing</td>
<td>$O(m \log n)$</td>
<td>$O(m \log n)$</td>
<td>$O(m \log \sqrt{n})$</td>
<td>$O(m + n)$</td>
</tr>
<tr>
<td>Fault tolerant</td>
<td>$O(nk \log^4 n)$</td>
<td>$O(nk \log^2 n)$</td>
<td>$O(nk \log^{3/2} \log n)$</td>
<td>$O(n(k' + \log n) \log n)$</td>
</tr>
<tr>
<td>Dynamic DFS</td>
<td>$O(\sqrt{mn} \log^{2.5} n)$</td>
<td>$O(\sqrt{mn} \log^{1.5} n)$</td>
<td>$O(\sqrt{mn} \log^{1.75} \log n)$</td>
<td>$O(\sqrt{mn} \log n)$</td>
</tr>
</tbody>
</table>

The new fully dynamic algorithm can be used to solve the dynamic subgraph problems discussed in [3] and improves upon their time complexity as well. Due to space constraint, we do not discuss these problems here. These can be accessed in the full version of the paper.

Table 1 offers a comparison of our results with all the previous results.

1.5 Organisation of the Paper

Section 2 introduces the notations and some well-known techniques/properties used throughout the paper. Section 3 defines the shallow tree representation, a concise structure which encapsulates the hierarchy of paths in the initial DFS tree. Section 4 is the core of our work. Here, we describe how a DFS tree can be rerooted efficiently. Section 5 describes how with some minor modifications to the data structure, rerooting procedure extends to a fault tolerant algorithm. We present the fully dynamic algorithm in Section 6.

2 Preliminaries

2.1 Notations

Following notations will be used throughout this paper.
- $T$: Any DFS tree of the original graph $G$.
- $dfn(x)$: The depth first number, i.e., the discovery time of the vertex $x$ during the DFS traversal.
- $v(i)$: the vertex $x \in V$ such that $dfn(x) = i$.
- $dist(x, y)$: distance between the vertices $x$ and $y$ in the DFS tree $T$.

For the sake of ease of explanation, we shall assume that the graph remains connected at all times. This assumption is without loss of generality because of the following standard way of transforming the original graph right in the beginning - Introduce a dummy vertex $r$ and connect it to all vertices of the graph. Henceforth, we maintain a DFS tree rooted at $r$ for this augmented graph. It is easy to observe that the augmented graph remains connected throughout and the DFS tree rooted at $r$ will be such that the subtrees rooted at the children of $r$ constitute a DFS forest of the original graph.

2.2 Heavy-Light Decomposition

Sleator and Tarjan, in their seminal result on dynamic trees [42] introduced a technique of partitioning any rooted tree called Heavy-light decomposition. Given any rooted tree, this technique splits it into a set of vertex-disjoint ancestor-descendant paths. It marks all the
tree edges either dashed or solid - a tree edge is marked solid iff the subtree of the child vertex is heaviest (in terms of number of vertices) among the subtrees of all its siblings and dashed otherwise. A maximal sequence of vertices connected through solid edges constitutes the required ancestor-descendant path. This decomposition can be carried out using DFS traversal in $O(n)$ time.

2.3 Fractional Cascading

Given $n$ sorted arrays and a value $x$, suppose we need to find the predecessor/successor of $x$ in each of them. A naive way is to make a binary search on each array. Chazelle and Guibas [14] introduced a novel tool called fractional cascading using which this problem can be solved more efficiently. Also, Chen et al. [18] used this tool for arriving at an $O(n)$ algorithm for incremental updates. We adapt a customized version of their method.

Lemma 7. Fractional cascading: Given $n$ sorted arrays $\{A_i\}_{i \in [n]}$ each with $l_i = |A_i|$ elements and total $\sum_{i} l_i = m$ elements. There exists a data structure of $O(m + n)$ words, which can be built in $O(m + n)$ time, such that for any given $x, i, k$ satisfying $i, k \in [n]$ and $i + k \leq n$, we can search for $x$ (or its predecessor/successor) in all arrays $A_i, \ldots, A_{i+k}$ using the data structure in $O(k + \log m)$ time.

3 Shallow Tree Representation

We now introduce the shallow tree representation for DFS tree $T$ that plays a key role in our algorithm. Using heavy-light decomposition, $T$ is broken down into a set of vertex-disjoint ancestor-descendant paths. Let’s denote this set with $\mathcal{P}$. Observe that these paths are connected through dashed edges in $T$. These dashed edges introduce a hierarchy among paths in $\mathcal{P}$ and the shallow tree defined below captures this hierarchy.

Definition 8. Given a DFS tree $T$ of an undirected graph $G$, let $\mathcal{P}$ be the set of paths obtained through heavy-light decomposition of $T$. Let $H$ be the set of edges marked dashed during the decomposition. For tree $T$, its shallow tree $S$ is a rooted tree formed by collapsing each element of $\mathcal{P}$ into a single node (super vertex). Note that, for each edge $(y, z) \in H$ with $y = \text{parent}(z)$, the node in $S$ that contains $y$ is the parent of the node containing $z$.

Figure 2 demonstrates how a DFS tree is decomposed to form a set of ancestor-descendant paths $\mathcal{P}$ which is subsequently used to form the shallow tree $S$.

To avoid ambiguity, we address the vertices in the shallow tree as “nodes” and the vertices in the DFS tree as “vertices”. node $(x)$ denotes the shallow tree node corresponding to the path in $\mathcal{P}$ containing vertex $x$.

The construction of $S$ described above ensures the following simple but crucial properties.

- As a result of heavy-light decomposition of a tree $T$ with $n$ vertices, there can be at most $\log n$ dashed edges on any root to leaf path in $T$. Recall that each edge in $S$ corresponds to a dashed edge. Thus, the depth of any node in tree $S$ can’t be larger than $\log n$. It is because of this small depth that we choose the name shallow tree for $S$.
- From the DFS property, neighbours of any vertex $v \in V$ in the graph are either ancestors or descendants of $v$ in $T$. Consider any such neighbour $u$. Let $p_1$ be the path in $\mathcal{P}$ containing $v$ and $p_2$ be the path in $\mathcal{P}$ containing $u$. $u$ and $v$ may also lie in the same path in $\mathcal{P}$. From the construction of $S$, the nodes corresponding to $p_1$ and $p_2$ will share an ancestor-descendant relation in $S$. So we can state the following lemma.
Figure 2 (a) A DFS tree \( T \). (b) Heavy light decomposition of \( T \) and the resulting paths in \( \mathcal{P} \). (c) The corresponding shallow tree.

Lemma 9. For a DFS tree \( T \) of an undirected graph \( G = (V, E) \) with shallow tree \( S \), any vertex \( v \in V \) which lies in node \( \mu \in S \) can have edges only to vertices lying in the nodes which are ancestors or descendants of \( \mu \) in \( S \).

We require that the vertices of each solid path have consecutive \( \text{dfn} \). This enables us to represent each path \( p \in \mathcal{P} \) in a compact manner using just the smallest and largest \( \text{dfn} \) of vertices on \( p \). For each path \( p \), we store this pair as \( \text{PathEndPoints} \) at the corresponding node of the shallow tree \( S \). Assigning consecutive \( \text{dfn} \) to each solid path can be accomplished easily - we carry out another DFS on \( T \) where for each vertex, the next vertex to be visited is its child hanging through a solid edge. Thus, the processing of \( T \) to make the shallow tree \( S \) requires only DFS on \( T \) and takes \( \mathcal{O}(n) \) time.

4 Rerooting DFS Tree \( T \)

Given a DFS tree \( T \) for a graph \( G = (V, E) \) and a vertex \( r' \in V \), the objective is to compute a DFS tree \( T^* \) rooted at \( r' \) for the same graph \( G \). First we compute the shallow tree representation \( S \) of \( T \). We now describe the rerooting procedure.

4.1 Reroot Procedure

The tree \( T^* \) is empty in the beginning and is grown gradually starting from \( r' \). To build this tree efficiently, we use the following two ideas. The first idea is to re-use the paths from \( \mathcal{P} \). Observe that while rerooting, if we use the original adjacency list for scanning the neighbours, we need \( \mathcal{O}(m + n) \) time. So the second idea is to populate for each vertex only a subset of its adjacency list which is small and yet sufficient to compute a DFS tree (this idea was used by [3] in the fault tolerant DFS problem). These lists we refer as reduced adjacency lists. We take a lazy and frugal approach to populate them.

The intuition sketched above is materialized by carrying out a DFS traversal \textit{guided} by the shallow tree \( S \). Note that a node of the shallow tree corresponds to a path in \( \mathcal{P} \). To compute \( T^* \), our algorithm performs a sequence of steps. Each step begins with entering a node of \( S \) through some vertex present on the path stored at the node and leaving it after traversing the path along one direction. The first node to be visited is \( \text{node}(r') \). We now
provide complete details of the computation involved in each step. Consider any node \( \nu \in S \). Let \( \text{path}(y,z) \) be the path corresponding to \( \nu \). When the DFS traversal enters \( \nu \) through a vertex, say \( x \), the following 3 simple operations are carried out.

1. **Move towards the farther end of the path.**

   We determine the vertex from \( \{y,z\} \) farther from \( x \). Let this vertex be \( y \). DFS traversal proceeds from \( x \) to \( y \) and \( \text{path}(x,y) \) is attached to the tree \( T^* \). Next, we update the \( \text{PathEndpoints}(\nu) \) such that it stores the endpoints of the untraversed part of \( \text{path}(y,z) \). This choice of direction ensures that at least half of the path is traversed (referred to as path halving technique, also used in [2, 3]).

2. **Populate the reduced adjacency list of the path just traversed.**

   For vertices on \( \text{path}(x,y) \), we populate the reduced adjacency list \( \mathcal{L} \) using ancestors and descendants of \( \nu \) in the shallow tree \( S \). This will be explained below in Section 4.1.1.

3. **Continue traversal.**

   Using the reduced adjacency list \( \mathcal{L} \) of the traversed path computed in step 2 above, we continue the DFS traversal along the unvisited neighbours of the vertices in the order from \( y \) to \( x \) (opposite to the direction of traversal, due to the recursive nature of DFS).

Algorithm 1 presents the complete pseudocode of the rerooting procedure based on the above 3 steps. All vertices are marked \( \text{unvisited} \) initially and the reduced adjacency list \( \mathcal{L} \) is empty. Invoking \( \text{Reroot}(r') \) produces the DFS tree rooted at \( r' \).

\begin{algorithm}
\caption{Recursive procedure to reroot the DFS Tree \( T \).}
1 Function \( \text{Reroot}(x) \)
2 \hspace{1em} (y,z) \leftarrow \text{PathEndpoints}(\text{node}(x)) ;
3 \hspace{1em} if \( \text{dist}(x,z) > \text{dist}(x,y) \) then \( \text{Swap}(y,z) ; \) /* compute \( \text{dist} \) using \( \text{dfns} \) */
4 \hspace{1em} \text{Attach path}(x,y) \) to \( T^* \);
5 \hspace{1em} if \( x \neq z \) then
6 \hspace{2em} \( w \leftarrow \text{Neighbour of } x \) on \( \text{path}(y,z) \) nearest to \( z \);
7 \hspace{2em} \text{PathEndpoints}(\text{node}(x)) \leftarrow (w,z) ; \) /* untraversed path */
8 \hspace{1em} end
9 \hspace{1em} \mathcal{L} \leftarrow \text{ReducedAL}(\mathcal{L}, (x,y)) ; \) /* Update \( \mathcal{L} \) for vertices on \( \text{path}(x,y) \) */
10 \hspace{1em} \text{for } i = \text{dfn}(y) \text{ to } \text{dfn}(x) \text{ do } \text{status}(v(i)) \leftarrow \text{visited} ;
11 \hspace{1em} \text{for } i = \text{dfn}(y) \text{ to } \text{dfn}(x) \text{ do }
12 \hspace{2em} \text{foreach vertex } u \in \mathcal{L}(v(i)) \text{ do }
13 \hspace{3em} \text{if } \text{status}(u) = \text{unvisited} \text{ then } \{ \text{add } (v(i), u) \text{ to } T^* ; \text{Reroot}(u) \} ;
14 \hspace{2em} \text{end}
15 \hspace{1em} \text{end}
\end{algorithm}

### 4.1.1 Populating Reduced Adjacency Lists

Here we define a query \( Q(u,(p_s,p_c)) \), where \( u, p_s, p_c \in V \) satisfy the following constraints:

- \( p_s \) and \( p_c \) are the endpoints of an ancestor-descendant path in \( T \).
- \( u \) is a descendant of the highest vertex on \( \text{path}(p_s,p_c) \), but does not lie on the path.

This query finds an edge from the vertex \( u \) that is incident to \( \text{path}(p_s,p_c) \) closest to \( p_c \). It returns \( \text{null} \) if no such edge exists, otherwise it returns the endpoint other than \( u \).

Consider any node \( \nu \in S \), and let \( \text{path}(y,z) \) be its corresponding path in \( \mathcal{P} \). When the DFS traversal enters \( \nu \) through \( x \) and proceeds towards \( y \), we populate the reduced adjacency lists of vertices on \( \text{path}(x,y) \) using Lemma 9 as follows.
Processing Ancestors of vertices on path(x, y).
For each u ∈ path(x, y) and for each ancestor µ of ν, we add Q(u, PathEndPoints(µ)) to \( L(u) \).

Processing Descendants of vertices on path(x, y).
Using the shallow tree \( S \), one can list all the vertices that are descendants of vertices on path(x, y) in \( T \). From all these vertices, we query for an edge to path(x, y) which is incident closest to \( y \). For any descendant \( u \), if the query \( Q(u, (x, y)) \) returns a non-null value, say \( w \), then we add \( u \) to \( L(w) \).

Remark 10. In Algorithm 1, after visiting \( x \), if we moved towards the leaf node, then the untraversed part of \( ν \) has to be treated as an ancestor path while populating the reduced adjacency lists of path(x, y), and as a descendant path, otherwise.

Algorithm 2 Populating reduced adjacency lists of vertices on path(x, y).

1. Function ReducedAL(\( L(x, y) \))
2. \( µ \leftarrow \text{parent}(\text{node}(x)) \);
3. while \( µ \neq \text{NULL} \) do /* edges to ancestor paths */
4.  for \( i = dfn(y) \) to \( dfn(x) \) do
5.    \( L(v(i)) \leftarrow L(v(i)) \cup \{Q(v(i), \text{PathEndPoints}(µ))\} \);
6.  end
7.  \( µ \leftarrow \text{parent}(µ) \);
8. end
9. if \( dfn(x) < dfn(y) \) then \( z \leftarrow x \) else \( z \leftarrow y \); /* \( z \) is ancestor among two */
10. \( C \leftarrow \text{Desc}_T(z) \setminus \{v(dfn(x)), \ldots, v(dfn(y))\} \);
11. foreach \( u \in C \) do /* each descendant of path(x, y) */
12.    \( w \leftarrow Q(u, (x, y)) \); /* ensure \( w \) is closest to \( y \) */
13. end
14. if \( w \neq \text{NULL} \) then \( L(w) \leftarrow L(w) \cup \{u\} \);
15. end
16. return \( L \)

In Algorithm 2, we used query \( Q(u, (p_s, p_e)) \) as a black box. This query can be answered very easily as follows. The constraints of \( Q(u, (p_s, p_e)) \) imply that the returned edge is always an edge from \( u \) to an ancestor of \( u \). For answering this query, we use a data structure \( D \) which stores the following information for each vertex \( u \).

Definition 11. \( D(u) \) is an array that stores each ancestor of \( u \) in \( T \) to which \( u \) is a neighbour and it stores them in the increasing order of distance from the root of \( T \).

This data structure enables us to answer query \( Q(u, (p_s, p_e)) \) using a binary search on array \( D(u) \) and it takes \( O(\log n) \) time only. Interestingly, the data structure \( D \) can be preprocessed very easily in \( O(m + n) \) time as follows. We visit vertices in the increasing order of their depth first numbers. Note that for each vertex \( v \), the neighbours of \( v \) which have \( dfn \) larger than that of \( v \) are its descendants. For each such descendant \( u \), we append \( v \) to \( D(u) \). Iterating in increasing order of \( dfn \) ensures that all arrays in \( D \) are sorted as needed.

Theorem 12. Given a DFS Tree \( T \) of an undirected graph \( G \), it takes \( O(m + n) \) time to build a data structure consisting of exactly \( (m + n) \) words which can answer the query \( Q(u, (p_s, p_e)) \) in \( O(\log n) \) time.
4.2 Time Complexity Analysis

During preprocessing, we construct the shallow tree $S$ for the DFS tree $T$ and build $D$. This processing as shown earlier can be completed in $O(m + n)$ time. The time complexity of the Reroot procedure is bounded by the time required to populate the reduced adjacency lists $L$. This in turn, is bounded by the number of calls to query $Q$. To analyse the number of calls made from any vertex $w$, let $v$ be the node in the shallow tree $S$, containing $w$. In general, if the bound on height of $S$ is $d$, ReducedAL makes worst-case $d$ queries from $w$ to ancestors of $v$. Also when an ancestor of $v$ in $S$, say $\mu$, is visited during Reroot procedure, ReducedAL makes a query from $w$ to $\mu$. The path halving technique (line 3 in Algorithm 1) ensures that any node in $S$ (or a path in $P$) is visited at most $\log n$ times. This implies that any such ancestor $\mu$ may be visited at most $\log n$ times. Thus, we can have worst-case $d(\log n + 1)$ queries from $w$ to its ancestor paths throughout the Reroot procedure. Summing over all the vertices, there can be at most $nd(\log n + 1)$ calls to query $Q$. Therefore, populating the reduced adjacency lists $L$ takes overall $O(nd\log^2 n)$ time. From Section 3, we know $d \leq \log n$. Thus, using Theorem 12, we can state the following lemma.

\textbf{Lemma 13.} Given a DFS Tree $T$ of an undirected graph $G$, it takes $O(m + n)$ time to build a data structure of $(m + n)$ words using which we can compute a DFS Tree of $G$ rooted at any given vertex in $O(n\log^3 n)$ time.

4.2.1 Getting rid of a log factor

Consider the moment when the Reroot procedure enters a path $(y, z)$ through the vertex $x$ and reaches the endpoint $y$. In Algorithm 2, for each descendant $w$ of path $(x,y)$, we perform query $Q(w, (x,y))$ separately. Instead, using Fractional Cascading, we can perform all these queries together in an efficient manner. Among $x$ and $y$, let $x$ be the vertex closer to the root of $T$. As described earlier, all the vertices of path $(x, y)$ have consecutive $dfn$, and so do the vertices of subtree $T(x)$. Let $last(x)$ be the vertex in $T(x)$ with the largest $dfn$. Since vertices on path $(x, y)$ have already been visited, we need to query for edges only from vertices with $dfn$ between $dfn(y) + 1$ and $dfn(last(x))$. Finding edges to path $(x, y)$ from these vertices can be done with a single query to the fractionally cascaded $D$ (Lemma 7). It takes $O(\log m + dfn(last(x)) − dfn(y))$ time to execute this query. We charge the $\log m$ part of the query time to the vertices on path $(x, y)$ and the $dfn(last(x)) − dfn(y)$ is distributed among the descendant vertices. Thus, each descendant vertex incurs a constant charge.

Note that, queries to the ancestors of path $(x, y)$ are answered using the original $D$ itself. Therefore, in a shallow tree of height $d$, each vertex $v \in V$ incurs following charges during Reroot procedure - $O(d \log n)$ when $v$ acts as a descendant in the queries made while visiting ancestors of $v$, and $O(d \log n + \log m)$ while visiting $v$ itself. Overall the charge on any vertex is $O(d \log n)$ and, therefore, the time complexity of Reroot procedure reduces to $O(nd \log n)$. So we can state the following lemma.

\textbf{Lemma 14.} Given an undirected graph $G$, the DFS tree $T$ and corresponding shallow tree of height $d$, it takes $O(m + n)$ time to build a data structure occupying $O(m + n)$ words using which the Reroot algorithm executes in $O(nd \log n)$ time.

4.3 Correctness of Reroot Procedure

As is evident from the pseudo code of Algorithm 1, besides populating the reduced adjacency lists, Reroot imitates a usual DFS traversal. So in order to show that Reroot computes a valid DFS tree, we need to prove the following. Each edge $e$ that is not added to reduced
adjacency list is indeed redundant and will appear as a back edge in the resulting DFS tree. We can show this as follows.

During Reroot procedure, consider the moment when we attach some path, say $\text{path}(x, y)$ to $T^*$. Lemma 9 implies that all the neighbours of vertices on the $\text{path}(x, y)$ will lie either in ancestors or in descendants of $\text{path}(x, y)$ in the shallow tree. For each descendant vertex $v$ of $\text{path}(x, y)$, we add an edge from $v$ incident on $\text{path}(x, y)$ closest to $y$ (lines 12, 13 of Algorithm 2). For each ancestor path of $\text{path}(x, y)$, say $\mu$, we add an edge from each vertex of $\text{path}(x, y)$ to $\mu$ (line 5 of Algorithm 2). These ensure that at the moment $\text{path}(x, y)$ is visited and attached to $T^*$, from each connected component in the graph induced by unvisited vertices, the edge incident on $\text{path}(x, y)$ closest to vertex $y$ is added to $L$ (if exists). Using Components Property, it follows that all those edges incident on $\text{path}(x, y)$, that are not added to $L$, will appear as back edges in the resulting DFS tree. Hence, the traversal performed using the reduced adjacency lists indeed results in a valid DFS tree.

5 Extension to Fault Tolerant DFS Tree

Given an undirected graph $G$, we build a DFS tree of $G$, say $T$, the corresponding shallow tree $S$, and the fractionally cascaded $D$. Let $F$ be the set of failures (edges/vertices) in $G$, with $|F| = k$. Here, we describe how with some elementary modifications to the shallow tree of $T$, procedure Reroot can be utilized to report a DFS tree of $G \setminus F$. For the given set $F$, we update the set $\mathcal{P}$ and the shallow tree $S$ as follows.

1. Each vertex maintains a state: active or failed. For a failed vertex $x \in F$, we toggle $x$’s state to failed. Let $p \in \mathcal{P}$ be the path containing $x$. We remove $x$ from $p$. The resulting smaller paths are added to $\mathcal{P}$ and $p$ is removed from $\mathcal{P}$.
2. For each failed edge $e = (u, v)$, we mark the corresponding entries in the adjacency lists of $u$ and $v$ as failed. Here, we do not make any changes to fractionally cascaded $D$. If failed edge $e$ is a tree-edge and was marked solid during heavy-light decomposition, $\mathcal{P}$ is updated as follows. Path $p \in \mathcal{P}$ containing $e$ is split into two smaller paths. These smaller paths are added to $\mathcal{P}$ and $p$ is removed from $\mathcal{P}$.
3. After updating $\mathcal{P}$, shallow tree $S$ is updated as follows. For any path $p \in \mathcal{P}$, let the vertex in $p$ closest to root of $T$ be $x$. The parent of node corresponding to $p$ in $S$ will be the node containing the nearest active ancestor of $x$.

Given the set $F$, we can update $\mathcal{P}$ and $S$ in $\mathcal{O}(n)$ time as follows - Make a DFS traversal through $T$ while ensuring vertices are visited in increasing order of their dfn. Update $\mathcal{P}$ and $S$ as discussed above. For each failure, the number of nodes in $S$ increases by at most one. After $k$ updates, $S$ can have at most $k$ new nodes and may not be shallow anymore. However, the depth of a node $\nu$ in $S$ doesn’t increase due to any failure which does not lie on the path from $\nu$ to root of $S$. Thus, if the maximum number of failures on any root-leaf path in $T$ is $k'$, the height of the shallow tree will be at most $k' + \log n$.

To ensure that no deleted vertices/edges are traversed during Reroot procedure, we make the following modifications. Since no changes are made to fractionally cascaded $D$, the result of some query $Q(v, (p_s, p_e))$ may be an endpoint of a failed edge. In such a case, we iterate in $D(v)$ starting from the failed edge towards $p_s$ until we find an edge present in $G \setminus F$. However, if we cross $p_s$ in doing so, it implies that there is no edge between vertex $v$ and $\text{path}(p_s, p_e)$, we stop and return null. During the procedure Reroot, we spend overall $\mathcal{O}(k)$ time in such iterations. Also note that each node of the updated $S$ corresponds to a path consisting of active vertices only. So $Q(v, (p_s, p_e))$ never returns a failed vertex. It follows from the changes in the query and the shallow tree described above that Reroot procedure reports a valid DFS tree of $G \setminus F$. Using Lemma 14, we can conclude the following theorem.
Theorem 15. An undirected graph $G$ can be preprocessed in $O(m + n)$ time to build a DFS tree, say $T$, and a data structure of $O(m + n)$ words such that for any set $F$ of $k$ failed vertices or edges, a DFS tree of $G \setminus F$ can be reported in $O(n (k' + \log n) \log n)$ time, where $k' \leq k$ is the maximum number of faults on any root-leaf path in the tree $T$.

## 6 Fully Dynamic DFS

We first describe how the fault tolerant DFS algorithm can handle incremental updates. Following that, we use the overlapped periodic rebuilding technique to arrive at a fully dynamic DFS algorithm. The ideas utilized in both of these steps were used by Baswana et al. [3].

Let $U$ be the set of updates in any undirected graph $G$. For the edge insertions, we directly add them to the reduced adjacency lists of the endpoints of the edges. To handle insertion of a vertex, we add the new vertices in $V$ and treat their edges as edge insertions. Observe that these modifications are sufficient to handle incremental updates. The size of the reduced adjacency lists after $|U|$ updates is at most $n (|U| + \log n) (\log n + 1)$ (from Section 4.2). So the worst-case time complexity of our algorithm is $O(n (|U| + \log n) \log n)$.

We can use the Reroot procedure to report a DFS tree after every update. But as $|U|$ increases, the algorithm slows down. Therefore, we need to rebuild the data structures (denoted collectively using $A$) periodically to maintain the efficiency. We rebuild them after, let’s say, every $c$ updates. It takes $O(m + n)$ time to rebuild the data structures. The cost of rebuilding is amortized among the $c$ updates. This results in an amortized runtime fully dynamic algorithm.

To get a worst-case bound, the rebuilding is done in an overlapped fashion as follows. The original data structures $A_0$ are used till first $2c$ updates. At the end of $kc$ updates (for $k \geq 1$), we start building $A_k$, the data structures for the graph after incorporating the first $kc$ updates. The computation of building $A_k$ is distributed evenly over the next $c$ updates to the graph (from $(kc + 1)$ to $(k + 1)c$ updates). However, during these updates we use $A_{k-1}$ to report the updated DFS Tree. This ensures $|U|$ is never larger than $2c$. After $(k + 1)c$ updates, we have $A_k$ ready for use and we can discard $A_{k-1}$. This helps us manage our data structures in $O(m + n)$ space. For this overlapped periodic rebuilding framework, the following lemma from [3] provides the worst-case bound on the update time we can derive.

Lemma 16. (Overlapped Periodic Rebuilding: Lemma 6.1 in [3]) Let $D$ be a data structure that can be used to report the solution of a graph problem after a set of $U$ updates on an input graph $G$. If $D$ can be built in $O(f)$ time and the solution for graph $G + U$ can be reported in $O(h + |U| \cdot g)$ time, then $D$ can be used to report the solution after every update in worst-case $O(\sqrt{fg} + h)$ update time, given that $f/g \leq n$.

Substituting $f = m$, $g = n \log n$ and $h = n \log^2 n$, we obtain, $O(\sqrt{fg} + h) = O(\sqrt{mn \log n})$. Hence we conclude with the following theorem.

Theorem 17. An undirected graph can be preprocessed in $O(m + n)$ time to build a data structure occupying $O(m + n)$ words, using which one can maintain a DFS tree for any online sequence of insertions and deletions of vertices/edges in $O(\sqrt{mn \log n})$ worst-case time per update.
References


