SIAM J. COMPUT.
Vol. 48, No. 4, pp. 1335–1363

DYNAMIC DFS IN UNDIRECTED GRAPHS: BREAKING THE $O(m)$ BARRIER*

SURENDER BASWANA†, SHREEJIT RAY CHAUDHURY†, KEERTI CHOUDHARY†, AND SHAHBAZ KHAN†

Abstract. Depth first search (DFS) tree is a fundamental data structure for solving various problems in graphs. It is well known that it takes $O(m+n)$ time to build a DFS tree for a given undirected graph $G = (V, E)$ on n vertices and m edges. We address the problem of maintaining a DFS tree when the graph is undergoing *updates* (insertion and deletion of vertices or edges). We present the following results for this problem: (1) *Fault tolerant DFS tree*: There exists a data structure of size $\tilde{O}(m)$ (where $\tilde{O}()$ hides the polylogarithmic factors) which can be preprocessed in $\tilde{O}(m)$ time such that given any set $\mathcal{F}$ of failed vertices or edges, a DFS tree of the graph $G \setminus \mathcal{F}$ can be reported in $\tilde{O}(n|\mathcal{F}|)$ time. (2) *Fully dynamic DFS tree*: There exists a fully dynamic algorithm for maintaining a DFS tree that takes $\tilde{O}(m)$ time for preprocessing and worst case $\tilde{O}(\sqrt{mn})$ time per update for any arbitrary online sequence of updates. (3) *Incremental DFS tree*: There exists an incremental algorithm for maintaining a DFS tree that takes $\tilde{O}(m)$ time for preprocessing and worst case $\tilde{O}(n)$ time per update for any arbitrary online sequence of edge insertion. These are the first $o(m)$ worst case time results for maintaining a DFS tree of a dense graph in a dynamic environment. Moreover, our fully dynamic algorithm provides, in a seamless manner, the first deterministic algorithm for dense graphs with $O(1)$ query time and $o(m)$ worst case update time for connectivity, biconnectivity, and 2-edge connectivity in the dynamic subgraph model.

Key words. depth first search, DFS, dynamic graph algorithm

AMS subject classifications. 05C85, 68W40, 68Q25

DOI. 10.1137/17M114306X

1. Introduction.

Depth first search (DFS) is a well-known graph traversal technique. Right from the seminal work of Tarjan [57], DFS traversal has played the central role in the design of efficient algorithms for many fundamental graph problems, namely biconnected components [57], strongly connected components [57], topological sorting, bipartite matching [38], dominators in directed graph [58], and planarity testing [39]. Interestingly, the role of DFS traversal is not confined to merely the design of efficient algorithms. For example, consider the classical result of Erdős and Rényi [25] for the phase transition phenomena in random graphs. There exist many proofs of this result which are intricate and based on highly sophisticated probability tools. However, recently, Krivelevich and Sudakov [42] designed a truly simple, short, and elegant proof for this result based on the insights from a DFS traversal in a graph.

Let $G = (V, E)$ be an undirected connected graph on n vertices and m edges. A DFS traversal of G starting from any vertex $r \in V$ produces a rooted spanning tree, called a DFS tree, with r as its root. It takes $O(m + n)$ time to perform a DFS

*Received by the editors August 10, 2017; accepted for publication (in revised form) May 24, 2019; published electronically July 23, 2019. The preliminary version of this paper appeared in *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA 2016), ACM, New York; SIAM, Philadelphia, 2016, pp. 730–739.

https://doi.org/10.1137/17M114306X

Funding: The first author's research was partially supported by UGC-ISF (the University Grants Commission of India and the Israel Science Foundation) and IMPECS (the Indo-German Max Planck Center for Computer Science). The third author's research was partially supported by Google India under the Google India Ph.D. Fellowship Award.

†Department of Computer Science and Engineering, I.I.T. Kanpur, Kanpur, 208016 Uttar Pradesh, India (sbaswana@cse.iitk.ac.in, shreejit.1@gmail.com, keerti@cse.iitk.ac.in, shahbazk@cse.iitk.ac.in).

traversal and generate a DFS tree. Given any rooted spanning tree of graph G, all nontree edges of the graph can be classified into two categories, namely back edges and cross edges, as follows. A nontree edge is called a *back edge* if one of its endpoints is an ancestor of the other in the tree. Otherwise, it is called a *cross edge*. A necessary and sufficient condition for any rooted spanning tree to be a DFS tree is that every nontree edge is a back edge. Thus, it can be seen that many DFS trees are possible for any given graph. However, if the traversal of the graph is performed according to the order specified by the adjacency lists of the graph, the resulting DFS tree will be unique. The ordered DFS tree problem is to compute the order in which the vertices get visited when the traversal is performed strictly according to the adjacency lists.

Several graph applications in the real world deal with graphs that keep changing with time. These changes/updates can be in the form of insertion or deletion of vertices or edges. An algorithmic graph problem is modeled in a dynamic environment as follows. There is an online sequence of updates on the graph, and the objective is to update the solution of the problem efficiently after each update. In particular, the time taken to update the solution has to be much smaller than that of the best static algorithm for the problem. In the last two decades, many dynamic algorithms have been designed for various graph problems, such as connectivity [36, 37, 41, 47], reachability [51, 53, 20], shortest path [19, 59, 52], spanners [9, 30, 50], and matching [32, 6, 56] and min-cut [60]. Another, and more restricted, variant of a dynamic environment is the fault tolerant environment. Here the aim is to build a compact data structure for a given problem that is resilient to failure of vertices/edges and can efficiently report the solution of the problem for any given set of failures. There has been a lot of work in the last two decades on fault tolerant algorithms for connectivity [13, 22, 28], shortest paths [8, 16, 21], and spanners [10, 15].

A dynamic graph algorithm is said to be *fully dynamic* if it handles both insertion as well as deletion updates. A partially dynamic algorithm is said to be *incremental* or *decremental* if it handles only insertion or only deletion updates, respectively. In this paper, we address the problem of maintaining a DFS tree efficiently in any dynamic environment.

1.1. Existing results on dynamic DFS. In spite of the simplicity of a DFS tree, designing any efficient parallel or dynamic algorithm for a DFS tree has turned out to be quite challenging. Reif [48] showed that the ordered DFS tree problem is a P-Complete problem. For many years, this result seemed to imply that the general DFS tree problem, that is, the computation of any DFS tree, is also inherently sequential. However, Aggarwal and Anderson [2] proved that the general DFS tree problem is in RNC by designing a parallel randomized algorithm that takes $O(\log^3 n)$ expected time. Further, the fastest parallel deterministic algorithm for the general DFS tree still takes $O(\sqrt{n})$ time [3, 29]. Whether the general DFS tree problem is in NC for directed (or undirected) graphs is still a longstanding open problem.

Reif [49] and later Miltersen et al. [43] proved that P-Completeness of a problem also implies hardness of the problem in the dynamic setting. The work of Miltersen et al. [43] shows that if the ordered DFS tree is updateable in $O(\mathbf{t}(m,n))$ time, then the solution of every problem in class P is updateable in $\tilde{O}(\mathbf{t}(m,n))$ time. In other words, maintaining the ordered DFS tree is indeed the hardest among all the problems in class P. In our view, this hardness result, which is actually for only the ordered DFS tree problem, has proved to be quite discouraging for the researchers working in the area of dynamic algorithms. This is evident from the fact that for all the static graph problems that were solved using a DFS traversal in the 1970's, none of their dynamic

counterparts used a dynamic DFS tree [36, 37, 41, 51, 12, 13, 22].

Apart from the hardness of the ordered DFS tree problem in a dynamic environment, very little progress has been achieved even for the problem of maintaining any DFS tree. Franciosa, Gambosi, and Nanni [26] designed an incremental algorithm for a DFS tree in a directed acyclic graph (DAG). For any arbitrary sequence of edge insertions, this algorithm takes $O(mn)$ total time to maintain a DFS tree from a given source. Recently, Baswana and Choudhary [5] designed a decremental algorithm for a DFS tree in a DAG that requires expected $O(mn \log n)$ total time. For undirected graphs, recently, Baswana and Khan [7] designed an incremental algorithm for maintaining a DFS tree requiring $O(n^2)$ total time. These algorithms are the only results known for the dynamic DFS tree problem. Moreover, none of these existing algorithms, though designed for only a partially dynamic environment, achieves a worst case bound of $o(m)$ on the update time. Furthermore, none of these results proves that general DFS is not as hard as ordered DFS in the dynamic environment. This is because the speculations of having to incur a complete recomputation in the worst case after an update is not disproved by amortized bounds resulting in the perceived $O(m)$ *barrier* for general DFS as well. So the following intriguing questions remain unanswered to date:

- Does there exist any nontrivial fully dynamic algorithm for maintaining a DFS tree?
- Is it possible to achieve worst case $o(m)$ update time for maintaining a DFS tree in a dynamic environment?

Not only do we answer these open questions affirmatively for undirected graphs, but we also use our dynamic algorithm for a DFS tree to provide efficient solutions for a couple of well-studied dynamic graph problems. Moreover, our results also handle vertex updates (insertion and deletion of vertices along with the incident edges), which are generally considered harder than edge updates. This feature enables us to use our results for applications in the dynamic subgraph model. Furthermore, assuming strongly believed conjectures [1], our results finally imply that general DFS is indeed not as hard as ordered DFS in the dynamic setting, as was the case in the parallel setting. This is because even an $\tilde{O}(\sqrt{mn})$ worst case update time dynamic algorithm for ordered DFS would imply the same bound for every problem in class P [43], refuting the conditional lower bounds [1].

1.2. Our results. We consider a generalized notion of updates wherein an update could be either insertion/deletion of a vertex (along with incident edges) or insertion/deletion of an edge. For any set U of such updates, let $G + U$ denote the graph obtained after performing the updates U on the graph G. Our main result can be succinctly described in the following theorem.

THEOREM 1. *An undirected graph can be preprocessed in $O(m \log n)$ time to build a data structure of size $O(m \log n)$, such that for any set U of $k \leq n$ updates a DFS tree of $G + U$ can be reported in $O(nk \log^4 n)$ time.*

With this result at the core, we easily obtain the following results for a dynamic DFS tree in an undirected graph:

1. *Fault tolerant DFS tree.* Given any set of k failed vertices or edges, we can report a DFS tree for the resulting graph in $O(nk \log^4 n)$ time.
2. *Fully dynamic DFS tree.* Given any arbitrary online sequence of vertex or edge updates, we can maintain a DFS tree in $O(\sqrt{mn} \log^{2.5} n)$ worst case time per update.

3. *Incremental DFS tree.* Given any arbitrary online sequence of edge insertions, we can maintain a DFS tree in $O(n \log^3 n)$ worst case time per edge insertion.

These are the first $o(m)$ worst case update time algorithms for dense graphs[1] for maintaining a DFS tree in a dynamic environment. Recently, there has been significant work [1, 33] on establishing conditional lower bounds on the time complexity of various dynamic graph problems. A simple reduction from [1], based on the strong exponential time hypothesis (SETH) [11, 40], implies a conditional lower bound of $\Omega(n)$ on the update time of any fully/partially dynamic algorithm for a DFS tree under vertex updates. We also present an unconditional lower bound of $\Omega(n)$ for maintaining a fully dynamic DFS tree explicitly under edge updates.

1.3. Applications of fully dynamic DFS. In the static setting, a DFS tree can be easily used to answer connectivity, 2-edge connectivity, and biconnectivity queries. Our fully dynamic algorithm for a DFS tree thus seamlessly solves these problems for both the vertex and the edge updates. Further, our result gives the first deterministic algorithm with $O(1)$ query time and $o(m)$ worst case update time for dense graphs for several well-studied variants of these problems in the dynamic setting. These problems include dynamic subgraph connectivity [13, 22, 24, 27, 37, 41] and vertex update versions of dynamic biconnectivity [35, 34, 37] and dynamic 2-edge connectivity [37, 24, 27]. The existing results offer different trade-offs between the update time and the query time and differ on the types (amortized or worst case) of update time and the types (deterministic or randomized) of query time. Our algorithm, in particular, improves the deterministic worst case bounds for these problems, thus demonstrating the relevance of DFS trees in solving dynamic graph problems.

Remark 1.1. After the preliminary version of this article [4] was published, there were several followup papers that improved the bounds proved in this article by polylogarithmic factors. These results essentially improved the data structure used by our algorithm. Chen et al. [17] reduced our data structure queries to the *orthogonal range search* problem. They showed that this data structure along with *fractional cascading* [14] reduces the update time of the incremental DFS algorithm to $O(n)$. Later, Nakamura and Sadakane [45] improved the space required by the data structure from $O(m \log n)$ to $O(m)$ words using *wavelet trees* [31]. Finally, Nakamura [44] presented a fully dynamic connectivity structure under vertex updates, which improves our presented bounds in some cases.

1.4. Main idea. Let T be a DFS tree of G. To compute a DFS tree of $G+U$ for a given set U of updates, the main idea is to make use of the original tree T itself. We preprocess the graph G using tree T to build a data structure $\mathcal{D}$. In order to achieve $o(m)$ update time for dense graphs, our algorithm makes use of $\mathcal{D}$ to create a *reduced* adjacency list for each vertex such that performing a DFS traversal using these lists gives a DFS tree for $G + U$. In fact, these reduced adjacency lists are generated on the fly and are guaranteed to have only $\tilde{O}(n|U|)$ edges.

We now give an outline of the paper. In section 2, we describe various notations used throughout the paper. Section 3 describes an algorithm to report the DFS tree after a single update in the graph. The details of the required data structure $\mathcal{D}$ are described in section 4. Then, in section 5, we provide an overview of our algorithm for handling multiple updates, highlighting the main intuition behind our approach.

[1] Our bounds are not $o(m)$ when $m = \tilde{O}(n)$, where even the trivial algorithm is close to the $\Omega(n)$ lower bound.

Our main algorithm (Theorem 1), which reports a DFS tree after any set of updates in the graph, is described in section 7. In section 8, we convert this algorithm to fully dynamic and incremental algorithms for maintaining a DFS tree using the *overlapped periodic rebuilding* technique. Finally, in sections 9 and 10 we describe the applications and lower bounds of dynamic DFS trees.

2. Preliminaries. Let U be any given set of updates. We add a dummy vertex r to the given graph in the beginning and connect it to all the vertices. Our algorithm starts with any arbitrary DFS tree T rooted at r in the augmented graph, and it maintains a DFS tree rooted at r at each stage. It can be observed easily that each subtree rooted at any child of r is a DFS tree of a connected component of the graph $G+U$. The following notations will be used throughout the paper:

- $T(x)$: The subtree of T rooted at vertex x.
- $path(x, y)$: The path from the vertex x to the vertex y in T.
- $dist_T(x, y)$: The number of edges on the path from x to y in T.
- $LCA(x, y)$: The lowest common ancestor of x and y in tree T.
- $N(w)$: The adjacency list of vertex w in the graph $G+U$.
- $L(w)$: The reduced adjacency list of vertex w in the graph $G+U$.
- T^*: The DFS tree rooted at r computed by our algorithm for the graph $G+U$.
- $par(w)$: Parent of w in T^*.

A subtree T' is said to be *hanging* from a path p if the root r' of T' is a child of some vertex on the path p and r' does not belong to the path p. Unless stated otherwise, every reference to a path refers to an ancestor-descendant path defined as follows.

DEFINITION 2 (ancestor-descendant path). *A path p in a DFS tree T is said to be an ancestor-descendant path if its endpoints have an ancestor-descendant relationship in T.*

We now state the operations supported by the data structure $\mathcal{D}$ (complete details of $\mathcal{D}$ are in section 4). Let U below refer to a set of updates that consists of vertex and edge deletions only. For any three vertices $w, x, y \in T$, where $path(x, y)$ is an ancestor-descendant path in T, the following two queries can be answered using $\mathcal{D}$ in $O(\log^3 n)$ time:

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

We now describe an important property of a DFS traversal that will be crucially used in our algorithm.

2.1. Properties of a DFS tree. A DFS traversal has the following flexibility: when the traversal reaches a vertex, say v, the next vertex to be traversed can be *any* unvisited neighbor of v. In order to compute a DFS tree for $G+U$ efficiently, our algorithm exploits this flexibility, the original DFS tree T, and the following property of a DFS traversal.

LEMMA 3 (components property). *Let T^* be the partially grown DFS tree and v be the vertex currently being visited. Let C be any connected component in the subgraph induced by the unvisited vertices. Suppose two edges e and e' from C are incident, respectively, on v and some ancestor (not necessarily proper) w of v in T^*. Then it is sufficient to consider only e during the rest of the DFS traversal; i.e., the edge e' need not be scanned. (Refer to Figure 1.)*

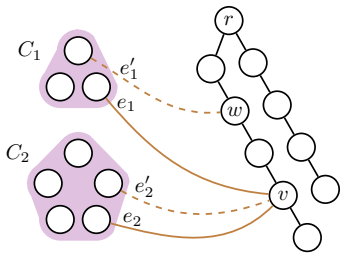

FIG. 1. *Edges e'_1 as well as e'_2 can be ignored during the DFS traversal.*

Skipping e' during the DFS traversal, as stated in the components property, is justified because e' will appear as a back edge in the resulting DFS tree. A similar property describing the *inessential* edges of a DFS trees was used by Smith [55] for computing a DFS tree of a planar graph in the parallel setting. In order to highlight the importance of the components property, and to motivate the requirement of data structure $\mathcal{D}$, we first consider a simpler case which deals with reporting a DFS tree after a single update in the graph.

3. Handling a single update. Consider the failure of a single edge (b, f) (refer to Figure 2(i)). Exploiting the flexibility of a DFS traversal, we can assume a stage in the DFS traversal of $G\backslash\{(b, f)\}$ where the partial DFS tree T^* is $T\backslash T(f)$ and vertex b is currently being visited. Thus, the unvisited graph is a single connected component containing the vertices of $T(f)$. Now, according to the components property, we need to process only the lowest edge from $T(f)$ to $path(b, r)$ ((k, b) in Figure 2(ii)). Hence, the DFS traversal enters this component using the edge (k, b) and performs a traversal of the subgraph induced by the vertices of $T(f)$. The resulting DFS tree of this subgraph would now be rooted at k. Rebuilding the DFS tree after the failure of edge (b, f) thus reduces to finding the lowest edge from $T(f)$ to $path(b, r)$ and then rerooting a subtree $T(f)$ of T at the new root k. We now describe how this rerooting can be performed in $\tilde{O}(n)$ time in the following section.

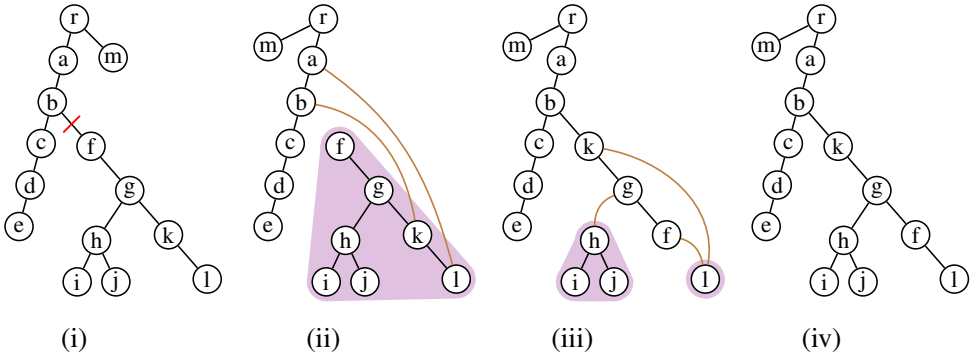

FIG. 2. (i) *Failure of edge (b, f).* (ii) *Partial DFS tree T^* with unvisited graph $T(f)$; components property allows us to neglect (a, l).* (iii) *Augmented path(k, f) to T^*; the components property allows us to neglect (l, k).* (iv) *Final DFS tree of $G\backslash\{(b, f)\}$.*

3.1. Rerooting a DFS tree. Given a DFS tree T originally rooted at r_0 and a vertex r', the aim is to compute a DFS tree of the graph that is rooted at r'. Note that any subtree $T(x)$ of the DFS tree T is also a DFS tree of the subgraph induced by the vertices of $T(x)$. Hence, the same procedure can be applied to reroot a subtree $T(x)$ of the DFS tree T. Thus, in general our aim is to reroot $T(r_0)$ at a new root $r' \in T(r_0)$ (see Figure 2(ii), where the subtree $T(f)$ would be rerooted at its new root k).

Procedure Reroot($T(r_0), r'$): Reroots the subtree $T(r_0)$ of T to be rooted at the vertex $r' \in T(r_0)$.

```
1  foreach (a, b) on path(r_0, r') do  /* a = par(b) in original tree T(r_0). */
2  |  par(a) ← b;
3  |  foreach child c of b not on path(r_0, r') do
4  |  |  (u, v) ← Query(T(c), r_0, b) ;        /* where u ∈ path(r_0, r') and
   |  |    v ∈ T(c). */
5  |  |  if (u, v) is nonnull then
6  |  |  |  Reroot(T(c), v);
7  |  |  |  par(v) ← u;
8  |  |  end
9  |  end
10 end
```

FIG. 3. *The recursive algorithm to reroot a DFS tree $T(r_0)$ at the new root r'.*

Our algorithm (refer to Procedure Reroot) essentially performs the DFS traversal (exploiting the flexibility of DFS) in such a way that components of the unvisited graph can be easily identified. The components property can then be applied to each such component, processing only $O(n)$ edges to compute the rerooted DFS tree. The DFS traversal first visits the path from r' to the root of tree $T(r_0)$. This reverses $path(r_0, r')$ in the new DFS tree T^*, as now r' would be an ancestor of r_0 (see Figure 2(iii)). Now, each subtree hanging from $path(r', r_0)$ in T forms a component of the unvisited graph. This is because the presence of any edge between these subtrees would imply a cross edge in the original DFS tree. Using the components property, we know that for each of these subtrees, say T_i, we only need to process the lowest edge from T_i on the new path from r' to r_0 in T^*. Since $path(r', r_0)$ is reversed in T^*, it is equivalent to processing the highest edge e_i from T_i to the $path(r_0, r')$ in T. Recall that this query can be answered by our data structure $\mathcal{D}$ in $O(\log^3 n)$ time (refer to section 2). Now, let v_i be the end vertex of e_i in T_i. The DFS traversal will thus visit the component induced by the vertices of T_i through e_i and produce its DFS tree, which is rooted at v_i. This rerooting can be performed by invoking the rerooting procedure recursively on the subtree T_i with the new root v_i.

We now analyze the total time required by Procedure Reroot to reroot a subtree T' of the DFS tree T. The total time taken by our algorithm is proportional to the number of edges processed by the algorithm. These edges include the *tree edges* that were a part of the original tree T' and the *added edges* that are returned by the data structure $\mathcal{D}$. Clearly, the number of tree edges in T' are $O(|T'|)$. Also, since the added edges eventually become a part of the new DFS tree T^*, they too are bounded

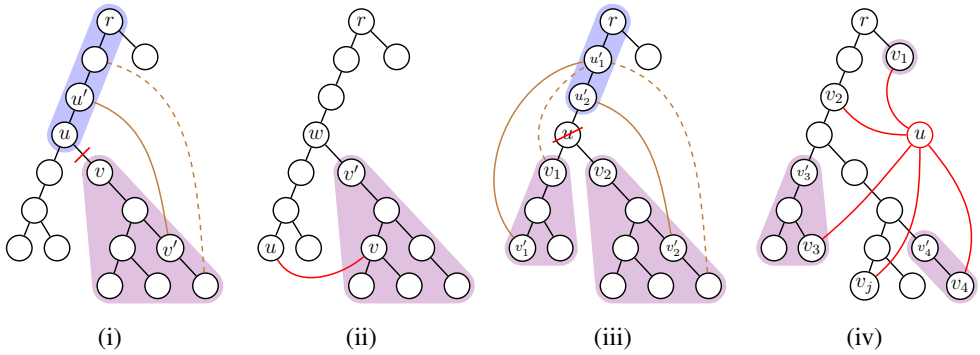

FIG. 4. *Updating the DFS tree after a single update:* (i) *deletion of an edge,* (ii) *insertion of an edge,* (iii) *deletion of a vertex, and* (iv) *insertion of a vertex. The reduction algorithm reroots the marked subtrees (shown in violet) and hangs it from the inserted edge (in case of insertion) or the lowest edge (in case of deletion) on the marked path (shown in blue) from the marked subtree. Color is available online only.*

by the size of the tree T'. Further, the data structure $\mathcal{D}$ takes $O(\log^3 n)$ time to report each added edge. Hence, the total time taken by our algorithm to rebuild T' is $O(|T'| \log^3 n)$ time. Since $\mathcal{D}$ can be built in $O(m \log n)$ time (refer to Theorem 6 in section 4), we have the following theorem.

THEOREM 4. *An undirected graph can be preprocessed to build a data structure in $O(m \log n)$ time, such that any subtree T' of the DFS tree can be rerooted at any vertex in T', in $O(|T'| \log^3 n)$ time.*

We now formally describe how rebuilding a DFS tree after an update can be reduced to this simple rerooting procedure (see Figure 4).

1. **Deletion of an edge (u, v):**
   In case (u, v) is a back edge in T, simply delete it from the graph. Otherwise, let $u = par(v)$ in T. The algorithm finds the lowest edge (u', v') on the $path(u, r)$ from $T(v)$, where $v' \in T(v)$. The subtree $T(v)$ is then rerooted at its new root v' and hanged from u' using (u', v') in the final tree T^*.
2. **Insertion of an edge (u, v):**
   In case (u, v) is a back edge, simply insert it in the graph. Otherwise, let w be the LCA of u and v in T and v' be the child of w such that $v \in T(v')$. The subtree $T(v')$ is then rerooted at its new root v and hanged from u using (u, v) in the final tree T^*.
3. **Deletion of a vertex u:**
   Let $v_1, \ldots, v_c$ be the children of u in T. For each subtree $T(v_i)$, the algorithm finds the lowest edge (u'_i, v'_i) on the $path(par(u), r)$ from $T(v_i)$, where $v'_i \in T(v_i)$. Each subtree $T(v_i)$ is then rerooted at its new root v'_i and hanged from u'_i using (u'_i, v'_i) in the final tree T^*.
4. **Insertion of a vertex u:**
   Let $v_1, \ldots, v_c$ be the neighbors of u in the graph. Make u a child of some v_j in T^*. For each v_i, such that $v_i \notin path(v_j, r)$, let $T(v'_i)$ be the subtree hanging from $path(v_j, r)$ such that $v_i \in T(v'_i)$. In case $T(v'_i)$ is same for multiple v_i's, choose any one. Each subtree $T(v'_i)$ is then rerooted at its new root v_i and hanged from u using (u, v_i) in the final tree T^*.

In case of vertex updates, multiple subtrees may be rerooted by the algorithm. Let these subtrees be $T_1, \ldots, T_c$. Thus, the total time taken by our algorithm is equal to the time taken to reroot the subtrees $T_1, \ldots, T_c$. Using Theorem 4, we know that a subtree T' can be rerooted in $\tilde{O}(|T'|)$ time. Since these subtrees are disjoint, the total time taken by our algorithm to build the resulting DFS tree is $\tilde{O}(|T_1| + \cdots + |T_c|) = \tilde{O}(n)$. Thus, we have the following theorem.

THEOREM 5. *An undirected graph can be preprocessed to build a data structure in $O(m \log n)$ time such that after a single update in the graph, the DFS tree can be reported in $O(n \log^3 n)$ time.*

4. Data structure. The efficiency of our algorithm heavily relies on the data structure $\mathcal{D}$. For any three vertices $w, x, y \in T$, where $path(x, y)$ is an ancestor-descendant path in T, we need to answer the following two kinds of queries:

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

We now describe construction of the data structure $\mathcal{D}$. It employs a combination of two well-known techniques, namely heavy-light decomposition [54] and suitable augmentation of a binary tree (segment tree), as follows:

1. Perform a preorder traversal of tree T with the following restriction: Upon visiting a vertex $v \in T$, the child of v that is visited first is the one storing the largest subtree. Let $\mathcal{L}$ be the list of vertices ordered by this traversal.
2. Build a segment tree $\mathcal{T}_\mathcal{B}$ whose leaf nodes from left to right represent the vertices in list $\mathcal{L}$.
3. Augment each node z of $\mathcal{T}_\mathcal{B}$ with a binary search tree $\mathcal{E}(z)$, storing all the edges $(u, v) \in E$, where u is a leaf node in the subtree rooted at z in $\mathcal{T}_\mathcal{B}$. These edges are sorted according to the position of the second endpoint in $\mathcal{L}$.

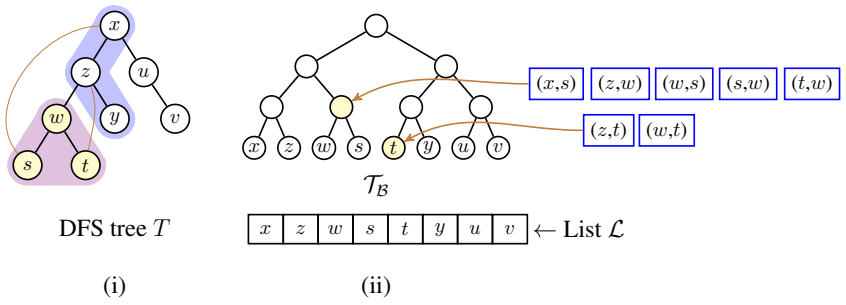

FIG. 5. (i) *The highest edge from subtree $T(w)$ on $path(x, y)$ is edge (x, s), and the lowest edges are edge (z, w) and (z, t).* (ii) *The vertices of $T(w)$ are represented as the union of two subtrees in segment tree $\mathcal{T}_\mathcal{B}$.*

The construction of $\mathcal{D}$ described above ensures the following properties, which are helpful in answering a query $Query(T(w), x, y)$ (see Figure 5).

- $T(w)$ is present as an interval of vertices in $\mathcal{L}$ (by step 1). Moreover, this interval can be expressed as a union of $O(\log n)$ disjoint subtrees in $\mathcal{T}_\mathcal{B}$ (by step 2). Let these subtrees be $\mathcal{T}_\mathcal{B}(z_1), \ldots, \mathcal{T}_\mathcal{B}(z_q)$.
- It follows from the heavy-light decomposition used in step 1 that the path $path(x, y)$ can be divided into $O(\log n)$ subpaths $path(x_1, y_1), \ldots, path(x_\ell, y_\ell)$

such that each subpath $path(x_i, y_i)$ is an interval in $\mathcal{L}$.

- Let query $Q(z, x, y)$ return the edge on $path(x, y)$ from the vertices in the subtree $\mathcal{T}_{\mathcal{B}}(z)$ that is closest to vertex x. Then it follows from step 3 that any query $Q(z_j, x_i, y_i)$ can be answered by a single predecessor or successor query on binary search tree $\mathcal{E}(z_j)$ in $O(\log n)$ time.

To answer $Query(T(w), x, y)$, we thus find the edge closest to x among all the edges reported by the queries $\{Q(z_j, x_i, y_i) | 1 \leq j \leq q \text{ and } 1 \leq i \leq \ell\}$. Thus, $Query(T(w), x, y)$ can be answered in $O(\log^3 n)$ time. Notice that $Query(w, x, y)$ can be considered as a special case where $q = 1$ and $\mathcal{T}_{\mathcal{B}}(z_1)$ is the leaf node of $\mathcal{T}_{\mathcal{B}}$ representing w, i.e., $z_1 = w$. The space required by $\mathcal{D}$ is $O(m \log n)$, as each edge is stored at $O(\log n)$ levels in $\mathcal{T}_{\mathcal{B}}$. Now, the segment tree $\mathcal{T}_{\mathcal{B}}$ can be built in linear time. Further, for every node $u \in \mathcal{T}_{\mathcal{B}}$, the sorted list of edges in $\mathcal{E}(u)$ can be computed in linear time by merging the sorted lists of its children. Thus, the binary search tree $\mathcal{E}(u)$ for each node $u \in \mathcal{T}_{\mathcal{B}}$ can be built in time linear in the number of edges in $\mathcal{E}(u)$. Hence, the total time required to build this data structure is $O(m \log n)$. Thus, we have the following theorem.

THEOREM 6. *The queries $Query(T(w), x, y)$, $Query(w, x, y)$ on T can be answered in $O(\log^3 n)$ worst case time using a data structure $\mathcal{D}$ of size $O(m \log n)$, which can be built in $O(m \log n)$ time.*

Note. Procedure Reroot can also use a simpler version of $\mathcal{D}$ which requires a smaller query time. However, our *generic* algorithm (described in section 7) would require these additional features of $\mathcal{D}$ as follows:

1. For Procedure Reroot, the binary search tree $\mathcal{E}(u)$ stored at each node u of $\mathcal{T}_{\mathcal{B}}$ can be replaced by an array storing the sorted list of edges, making it simpler to implement. However, our *generic* algorithm also requires deletion of edges from $\mathcal{D}$. An edge can be deleted from $\mathcal{D}$ by deleting the edge from the binary search trees stored at its endpoints and their ancestors in $\mathcal{T}_{\mathcal{B}}$. Since a deletion in a binary search tree takes $O(\log n)$ time, an edge can be deleted from $\mathcal{D}$ in $O(\log^2 n)$ time.
2. Procedure Reroot only performs the second type of query on the data structure $\mathcal{D}$, i.e., $Query(T(w), x, y)$. Thus, it would essentially be querying only the part of $path(x, y)$ comprising of the ancestors of w in $path(x, y)$. This is thus equivalent to $Query(T(w), LCA(x, w), LCA(y, w))$ (see Figure 5), which will answer the required query, as the only edges from $T(w)$ in this interval are incident on $path(x, y)$. In such a case, heavy-light decomposition and hence division of $path(x, y)$ to $O(\log n)$ subpaths would not be required. Hence, on each node in $u \in \mathcal{T}_{\mathcal{B}}$, the query is performed for a single path, requiring total $O(\log^2 n)$ time. However, our *generic* algorithm also uses the first type of query, i.e., $Query(w, x, y)$, where w can be an ancestor of x and y. In such a case, we need to perform the query only on contiguous intervals of $\mathcal{L}$, as the interval between x and y in $\mathcal{L}$ would have several other edges from w that are not incident on $path(x, y)$. This necessitates the use of heavy-light decomposition, and hence each query requires $O(\log^3 n)$ time.

Remark 4.1. Recently, Nakamura and Sadakane [45] reduced the space required by our data structure to $O(m)$. Some of their observations can also be used to address the above concerns as follows:

1. The requirement of binary search trees stated above can be relaxed if we implement an edge deletion using a vertex deletion and a vertex insertion as follows. We delete one of its endpoints and reinsert it without the corre-

sponding deleted edge.

2. The *heavy-light decomposition* is required only by query $Query(w, x, y)$. Thus, if we use different data structures for $Query(w, x, y)$ and $Query(T(w), x, y)$ requiring $O(\log^3 n)$ and $O(\log^2 n)$ time, respectively, the running time of the fault tolerant algorithm reduces by a $\log n$ factor. This is because the algorithm uses $O(\log n)$ times fewer queries of type $Query(w, x, y)$ as compared to $Query(T(w), x, y)$ (see section 7.3). Hence, the algorithms for fully dynamic DFS and the incremental DFS can be improved by $O(\sqrt{\log n})$ factors.

5. Handling multiple updates: Overview. A DFS tree can be computed in $\tilde{O}(n)$ time after a single update in the graph, by reducing it to Procedure Reroot. However, the same procedure cannot be directly applied to handle a sequence of updates for the following reason. The efficiency of Procedure Reroot crucially depends on the data structure $\mathcal{D}$ which is built using the DFS tree T of the original graph. Thus, when the DFS tree is updated, we are required to rebuild $\mathcal{D}$ for the updated tree. Now, rebuilding $\mathcal{D}$ is highly inefficient because it requires $O(m \log n)$ time. Thus, in order to handle a sequence of updates, our aim is to use the same $\mathcal{D}$ for handling multiple updates, without having to rebuild it after every update. We now give an overview of the algorithm that reports the DFS tree after a set U of updates.

In case of a single update, all the edges reported by $\mathcal{D}$ are added to the final DFS tree T^*. However, while handling multiple updates, we use $\mathcal{D}$ to build *reduced adjacency lists* for vertices of the graph, such that the DFS traversal of the graph using these *sparser* lists gives the DFS tree of the updated graph. Now, the data structure $\mathcal{D}$ finds the lowest/highest edge from a subtree of T to an ancestor-descendant path of T. Thus, in order to employ $\mathcal{D}$ to report a DFS tree of $G+U$, we need to ensure that the queried subtrees and paths do not contain any failed edges or vertices from U. Hence, for any set U of updates, we compute a partitioning of T into a disjoint collection of ancestor-descendant paths and subtrees such that none of these subtrees and paths contains any failed edge or vertex. An important property of this partitioning is that there are no edges from G lying between any two subtrees in this partitioning. We refer to this partitioning as a *disjoint tree partitioning*. Note that this partitioning depends only upon the vertex and edge failures present in the set U.

Recall that during the DFS traversal we need to find the lowest edge from each component C of the unvisited graph. It turns out that any component C can be represented as a union of subtrees and ancestor-descendant paths of the original DFS tree T. The components property can now be employed to compute the reduced adjacency lists of the vertices of the graph as follows. We just find the lowest edge from each of the subtrees and the ancestor-descendant paths to T^* by querying the data structure $\mathcal{D}$. Let this edge be (x, y), where $x \in T^*$ and $y \in C$. We can just add y to the reduced adjacency list $L(x)$ of x. Since the components property ensures the remaining edges adjacent to T^* can be ignored, the DFS traversal would thus consider all possible candidates for the lowest edge from every component C to T^*. Let the initial disjoint tree partitioning consist of a set of ancestor-descendant paths $\mathcal{P}$ and a set of subtrees $\mathcal{T}$. The algorithm for computing a DFS tree of $G+U$ can be summarized as follows.

Perform the static DFS traversal on the graph with the elements of $\mathcal{P} \cup \mathcal{T}$ as the super *vertices. Visiting a super vertex v^* by the algorithm involves extracting an ancestor-descendant path p_0 from v^* and attaching it to the partially grown DFS tree T^*. The remaining part of v^* is added back to $\mathcal{P} \cup \mathcal{T}$ as* new *super vertices. Thereafter, the reduced adjacency lists of the vertices on path p_0 are computed using*

the data structure $\mathcal{D}$. The algorithm then continues to find the next super vertex using the reduced adjacency lists, and so on.

6. Disjoint tree partitioning. We formally define disjoint tree partitioning as follows.

DEFINITION 7. *Given a DFS tree T of an undirected graph G and a set U of failed vertices and edges, let A be a vertex set in $G+U$. The disjoint tree partitioning defined by A is a partition of the subgraph of T induced by A into*

1. *a set of paths $\mathcal{P}$ such that* (i) *each path in $\mathcal{P}$ is an ancestor-descendant path in T and does not contain any deleted edge or vertex, and* (ii) $|\mathcal{P}| \leq |U|$*; and*
2. *a set of trees $\mathcal{T}$ such that each tree $\tau \in \mathcal{T}$ is a subtree of T which does not contain any deleted edge or vertex.*

Note that for any distinct $\tau_1, \tau_2 \in \mathcal{T}$, there is no edge between τ_1 and τ_2, as T is a DFS tree.

Let V_U be the vertex set of the updated graph $G+U$. The disjoint tree partitioning for set $A = V_U \setminus \{r\}$ can be computed as follows. Let V_f and E_f denote, respectively, the set of failed vertices and edges associated with the updates U. We initialize $\mathcal{P} = \emptyset$ and $\mathcal{T} = \{T(w) \mid w \text{ is a child of } r\}$. We refine the partitioning by processing each vertex $v \in V_f$ as follows (see Figure 6(i)):

- If v is present in some $T' \in \mathcal{T}$, we add the path from $par(v)$ to the root of T' to $\mathcal{P}$. We remove T' from $\mathcal{T}$ and add all the subtrees hanging from this path and v to $\mathcal{T}$.
- If v is present in some path $p \in \mathcal{P}$, we split p at v into two paths. We remove p from $\mathcal{P}$ and add these two paths to $\mathcal{P}$.

Edge deletions are handled as follows. We first remove edges from E_f that do not appear in T. Processing of the remaining edges from E_f is quite similar to the processing of V_f as described above. For each edge $e \in E_f$, just visualize deleting an imaginary vertex lying at the midpoint of the edge e (see Figure 6(ii)). It takes $O(n)$ time to process any $v \in V_f$ and any $e \in E_f$.

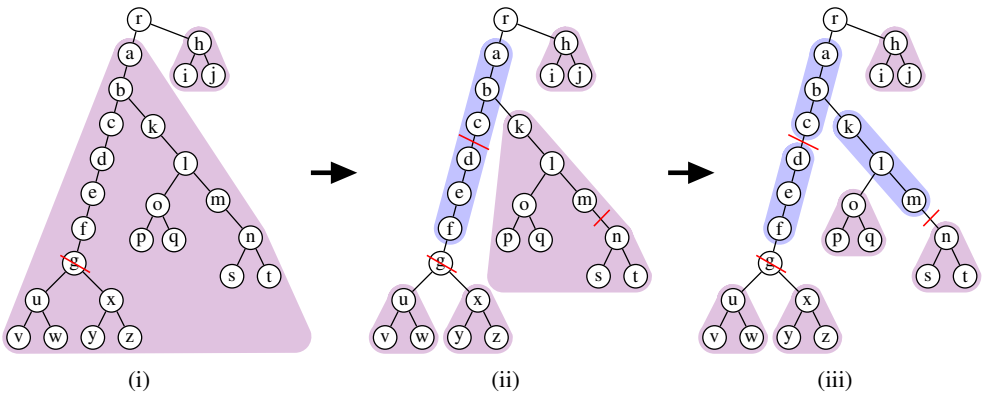

FIG. 6. *Disjoint tree partitioning for $V_U \setminus \{r\}$:* (i) *Initializing $\mathcal{T} = \{T(a), T(h)\}$ and $\mathcal{P} = \emptyset$.* (ii) *Disjoint tree partition obtained after deleting the vertex g.* (iii) *Final disjoint tree partition obtained after deleting the edges (c, d) and (m, n).*

Note that each update can add at most one path to $\mathcal{P}$. So the size of $\mathcal{P}$ is bounded by $|U|$. The fact that T is a DFS tree of G ensures that no two subtrees in $\mathcal{T}$ will have

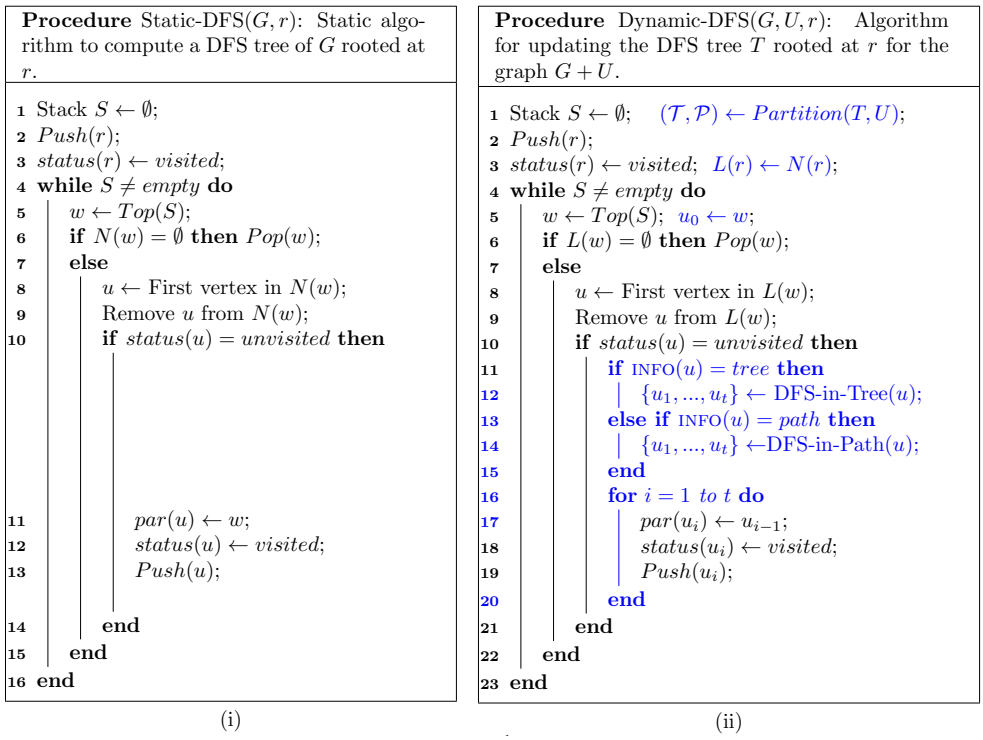

FIG. 7. *The static (and dynamic) algorithm for computing (updating) a DFS tree. The key differences are shown in blue. Color is available online only.*

an edge between them. So $\mathcal{P} \cup \mathcal{T}$ satisfies all the conditions stated in Definition 7.

LEMMA 8. *Given an undirected graph G with a DFS tree T and a set U of failing vertices and edges, we can find a disjoint tree partition of set $V_U \setminus \{r\}$ in $O(n|U|)$ time.*

7. Fault tolerant DFS tree. We first present a fault tolerant algorithm for a DFS tree. Let U be any given set of failed vertices or edges in G. In order to compute the DFS tree T^* for $G + U$, our algorithm first constructs a disjoint tree partition $(\mathcal{T}, \mathcal{P})$ for $V_U \setminus \{r\}$ defined by the updates U (see Lemma 8). Thereafter, it can be visualized as the static DFS traversal on the graph whose (*super*) vertices are the elements of $\mathcal{P} \cup \mathcal{T}$. Note that our notion of super vertices is for the sake of understanding only.

Consider the stack-based implementation of the static algorithm for computing a DFS tree rooted at a vertex r in graph G (refer to Figure 7(i)). Our algorithm for computing a DFS tree for $G + U$ (refer to Figure 7(ii)) is quite similar to the static algorithm. The only points of difference are the following:

- In the static DFS algorithm whenever a vertex is visited, it is attached to the DFS tree and pushed into the stack S. In our algorithm when a vertex u in some super vertex $v_s \in \mathcal{P} \cup \mathcal{T}$ is visited, a path starting from u is extracted from v_s and attached to the DFS tree, and this entire path is pushed into the stack S.

- Instead of scanning the entire adjacency list $N(w)$ of a vertex w, the reduced adjacency list $L(w)$ is scanned.

When a path is extracted from a super vertex v_s, the remaining unvisited part of v_s is added back to $\mathcal{T} \cup \mathcal{P}$. However, we need to ensure that the properties of the disjoint tree partitioning are satisfied in the updated $\mathcal{T} \cup \mathcal{P}$. This is achieved using Procedure DFS-in-Path and Procedure DFS-in-Tree, which also build the reduced adjacency list for the vertices on the path. The construction of a sparse reduced adjacency list is inspired by the components property, which can be easily *adapted* in the context of our algorithm as follows.

LEMMA 9 (adapted components property). *When a path p is attached to the partially constructed DFS tree T^* during the algorithm, for every edge (x, y), where $x \in p$ and y belongs to the unvisited graph, the following condition holds. Either y is added to $L(x)$ or y' is added to $L(x')$ for some edge (x', y'), where x' is a descendant (not necessarily proper) of x in p and y' is connected to y in the unvisited graph.*

We now describe how the properties of disjoint tree partitioning (and hence the adapted components property) are maintained by our algorithm when a vertex $v \in v_s$ is visited by the traversal:

1. Let $v_s = path(x, y) \in \mathcal{P}$. Exploiting the flexibility of DFS, we traverse from v to the farther end of $path(x, y)$. Now, $path(x, y)$ is removed from $\mathcal{P}$ and the untraversed part of $path(x, y)$ (with length at most half of $|path(x, y)|$) is added back to $\mathcal{P}$. We refer to this as *path halving*. This technique was also used by Aggarwal and Anderson [2] in their parallel algorithm for computing a DFS tree in undirected graphs. Notice that $|\mathcal{P}|$ remains unchanged or decreases by 1 after this step.
2. Let $v_s = \tau \in \mathcal{T}$. Exploiting the flexibility of a DFS traversal, we traverse the path from v to the root of τ, say x, and add it to T^*. Thereafter, τ is removed from $\mathcal{T}$ and all the subtrees hanging from this path are added to $\mathcal{T}$. Observe that every newly added subtree is also a subtree of the original DFS tree T. So the properties of disjoint tree partitioning are satisfied after this step as well.

Remark 7.1. In both cases, where $v_s \in \mathcal{T}$ or $v_s \in \mathcal{P}$, the algorithm crucially exploits the fact that the underlying graph is an undirected graph.

Let $path(v, x)$ be the path extracted from v_s. For each vertex w in this newly added path, we compute $L(w)$, ensuring the adapted components property as follows:

(i) For each path $p \in \mathcal{P}$, among potentially many edges incident on w from p, we just add any one edge.

(ii) For each tree $\tau' \in \mathcal{T}$, we add at most one edge to L as follows. Among all edges incident on τ' from $path(v, x)$, if (w, z) is the edge such that w is nearest to x on $path(v, x)$, then we add z to $L(w)$. However, for the case $v_s \in \mathcal{T}$, we have to consider only the newly added subtrees in $\mathcal{T}$ for this step. This is because the disjoint tree partitioning ensures the absence of edges between v_s and any other tree in $\mathcal{T}$.

Figure 8 provides an illustration of how $\mathcal{T} \cup \mathcal{P}$ is updated when a super vertex in $\mathcal{T} \cup \mathcal{P}$ is visited.

7.1. Implementation of our algorithm. We now describe our algorithm in full detail. First, we delete all the failed edges in U from the data structure $\mathcal{D}$. Now, the algorithm begins with a disjoint tree partition $(\mathcal{T}, \mathcal{P})$ which evolves as the algorithm proceeds. The state of any unvisited vertex in this partition is captured by

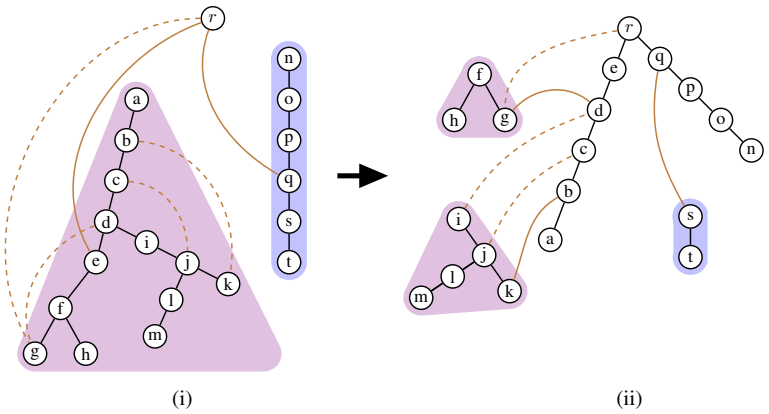

FIG. 8. *Visiting a super vertex from* $\mathcal{T} \cup \mathcal{P}$. (i) *The algorithm visits* $T(a) \in \mathcal{T}$ *using the edge* (r, e) *and the* $path(n, t) \in \mathcal{P}$ *using the edge* (r, q). (ii) *The traversal extracts* $path(e, a)$ *and* $path(q, n)$ *and augments them to* T^*. *The unvisited segments are added back to* $\mathcal{T}$ *and* $\mathcal{P}$.

the following three variables:

INFO(u): this variable is set to *tree* if u belongs to a tree in $\mathcal{T}$ and set to *path* otherwise.

ISROOT(v): this variable is set to $True$ if v is the root of a tree in $\mathcal{T}$ and $False$ otherwise.

PATHPARAM(v): if v belongs to some path, say $path(x, y)$, in $\mathcal{P}$, then this variable stores the pair (x, y) and is *null* otherwise.

Remark 7.2. The initialization and maintenance of variables INFO(u), ISROOT(v), and PATHPARAM(v) are not explicitly described (except in pseudocodes) for better readability.

Procedure Dynamic-DFS: For each vertex v, $status(v)$ is initially set as *unvisited*, and $L(v)$ is initialized to $\emptyset$. First, a disjoint tree partition is computed for the DFS tree T based on the updates U. Procedure Dynamic-DFS then inserts the root vertex r into the stack S. While the stack is nonempty, the procedure repeats the following steps. It reads the top vertex from the stack. Let this vertex be w. If $L(w)$ is empty, then w is popped out from the stack; otherwise, let u be the first vertex in $L(w)$. If vertex u is unvisited until now, then depending upon whether u belongs to some tree in $\mathcal{T}$ or some path in $\mathcal{P}$, Procedure DFS-in-Tree or DFS-in-Path is executed. A path p_0 is then returned to Procedure Dynamic-DFS, where for each vertex of p_0 a parent is assigned and the status is marked visited. The whole of this path is then pushed into the stack. The procedure proceeds to the next iteration of the while loop with the updated stack.

Procedure DFS-in-Tree: Let vertex u be present in some tree, say $T(v)$, in $\mathcal{T}$ (the vertex v can be found easily by scanning the ancestors of u and checking their value of ISROOT). The DFS traversal enters the tree from u and leaves from the vertex v. Let $path(u, v) = \langle w_1 = u, w_2, \ldots, w_t = v \rangle$. The $path(u, v)$ is pushed into the stack and attached to the partially constructed DFS tree T^*. We now update the partition $(\mathcal{P}, \mathcal{T})$ and also update the reduced adjacency list for each w_i present on $path(u, v)$ as follows:

1. For each vertex w_i and every path $path(x, y) \in \mathcal{P}$, we perform $Query(w_i, x, y)$ on the data structure $\mathcal{D}$ that returns an edge (w_i, z) such that $z \in path(x, y)$.

Procedure DFS-in-Tree(u): DFS traversal enters from node u and exits from v, the root of the tree containing node u in set $\mathcal{T}$.

```
1  v ← u;
2  while IsRoot(v) ≠ True do
3  |  v ← par(v)
4  end
5  IsRoot(v) ← False;
6  T ← T\T(v);
7  (w_1, ..., w_t) ← path(u, v);
8  for i = 1 to t do
9  |  foreach path(x, y) ∈ P do
10 |  |  if Query(w_i, x, y) ≠ ∅ then
11 |  |  |  (w_i, z) ← Query(w_i, x, y);
12 |  |  |  L(w_i) ← L(w_i) ∪ {z};
13 |  |  end
14 |  end
15 |  foreach child w of w_i except w_{i-1} do
16 |  |  (y, z) ← Query(T(w), v, u);
   |  |  /* where y ∈ path(u, v)          */
17 |  |  L(y) ← L(y) ∪ {z};
18 |  |  T ← T ∪ T(w);
19 |  |  IsRoot(w) ← True;
20 |  end
21 end
22 Return path(u, v);
```

Procedure DFS-in-Path(u): DFS traversal enters from node u and exits from v, the farther end of path containing node u in set $\mathcal{P}$.

```
1  (v, d) ← PathParam(u);
2  if dist_T(u, d) > dist_T(u, v) then Swap(v, d);
3  c ← Neighbor of u on path(v, d) nearer to d;
4  P ← (P\path(v, d)) ∪ path(c, d);
5  for c' ∈ path(c, d) do
6  |  PathParam(c') ← (c, d);
7  end
8  (w_1, ..., w_t) ← path(u, v);
9  for i = 1 to t do
10 |  foreach path(x, y) ∈ P do
11 |  |  if Query(w_i, x, y) ≠ ∅ then
12 |  |  |  (w_i, z) ← Query(w_i, x, y);
13 |  |  |  L(w_i) ← L(w_i) ∪ {z};
14 |  |  end
15 |  end
16 end
17 foreach T(w) ∈ T do
18 |  if Query(T(w), v, u) ≠ ∅ then
19 |  |  (y, z) ← Query(T(w), v, u);
   |  |  /* where y ∈ path(u, v)          */
20 |  |  L(y) ← L(y) ∪ {z};
21 |  end
22 end
23 Return path(u, v);
```

FIG. 9. *The pseudocode of Procedures DFS-in-Tree and Procedures DFS-in-Path.*

We add z to $L(w_i)$.

2. Recall that since subtrees in T do not have any cross edge between them, therefore there cannot be any edge incident on $path(u, v)$ from trees which are already present in $\mathcal{T}$. An edge can be incident only from the subtrees which were hanging from $path(u, v)$. $T(v)$ is removed from $\mathcal{T}$, and all the subtrees of $T(v)$ hanging from $path(u, v)$ are inserted into $\mathcal{T}$. For each such subtree, say τ, inserted into $\mathcal{T}$, we perform $Query(\tau, v, u)$ on the data structure $\mathcal{D}$ that returns an edge, say (y, z), such that $z \in \tau$ and y is nearest to v on $path(u, v)$. We insert z into $L(y)$.

Procedure DFS-in-Path: Let vertex u visited by the DFS traversal lie on a $path(v, y) \in \mathcal{P}$. Assume $dist_T(u, v) > dist_T(u, y)$. The DFS traversal travels from u to v (the farther end of the path). The path $path(v, y)$ in set $\mathcal{P}$ is replaced by its subpath, which remains unvisited. The reduced adjacency list of each $w \in path(u, v)$ is updated in a way similar to that in Procedure DFS-in-Tree, except that in step 2, we perform $Query(\tau, v, u)$ for each $\tau \in \mathcal{T}$. Note that while performing step 1, the vertex w_i can be an ancestor of the vertices of $path(x, y)$. This is because the vertices of a path in $\mathcal{P}$ can be ancestors of the vertices of another path in $\mathcal{P}$. This was not true for Procedure DFS-in-Tree because vertices of a subtree in $\mathcal{T}$ cannot be ancestors of vertices of any path in $\mathcal{P}$. Thus, our data structure $\mathcal{D}$ needs to support queries where w_i is an ancestor of the queried path (refer to the note at the end of section 4).

The reader may refer to Figure 9 for pseudocode of Procedures DFS-in-Tree and DFS-in-Path. This completes the description of the fault tolerant algorithm for a DFS tree. This algorithm maintains the adapted components property at each stage

by construction, given that the properties of disjoint tree partitioning are satisfied.

7.2. Correctness. It can be seen that the following two invariants hold for the while loop in the Procedure Static-DFS described in Figure 7(i). It is easy to see that these invariants imply the correctness of the algorithm, i.e., the generated tree is a rooted spanning tree where every nontree edge is a back edge:

I_1: The sequence of vertices in the stack from bottom to top constitutes an ancestor-descendant path from r in the DFS tree computed.

I_2: For each vertex v that is popped out, all vertices in the set $N(v)$ have already been visited.

Invariant I_1 also holds for Procedure Dynamic-DFS by construction. The following lemma proves that invariant I_2 is also maintained by the procedure. This proof uses the fact that our algorithm maintains the adapted components property, which holds by construction.

LEMMA 10. *The invariant I_2 holds true at each stage of Procedure Dynamic-DFS.*

Proof. We give a proof by contradiction as follows. Assume that x is the first vertex that is popped out of the stack before some vertex $y \in N(x)$ is visited. Consider the time when a path p containing x was pushed in the stack. Clearly, $y \notin L(x)$, and hence using the adapted components property we know that some $y' \in L(x')$ is connected to y in the unvisited graph, where x' is a descendant (not necessarily proper) of x in p. Let p^* be a path between y' and y in the unvisited graph.

Now, consider the time when x is popped out of the stack. Clearly, all its descendants including x' have been popped out, and so using invariant I_2 for x', y' has been visited by the traversal. Thus, p^* can be divided into two nonempty sets A and B, denoting visited and unvisited vertices of p^*, respectively. Here $y' \in A$ and $y \in B$, and thus clearly for the last vertex of p^* that is present in A, the invariant I_2 is not satisfied. This contradicts our assumption that x is the first vertex that is popped out of the stack for which I_2 is not satisfied. Thus, maintenance of the adapted components property ensures the invariant I_2 in our algorithm. ∎

Hence, our algorithm indeed computes a valid DFS tree for $G + U$.

7.3. Time complexity analysis. As described earlier, the disjoint tree partitioning and the components property play a key role in the efficiency of our algorithm. They allow us to limit the size of the reduced adjacency lists L, which are built during the algorithm. Our algorithm computes T^* by performing a DFS traversal on the reduced adjacency list L. Thus, the time complexity of our algorithm is $O(n + |L|)$, excluding the time required to compute L.

We first establish a bound on the size of L. In each step, our algorithm extracts a path from $v_s \in \mathcal{P} \cup \mathcal{T}$ and attaches it to T^*. Let P_t and P_p denote the set of such paths that originally belonged to some tree in $\mathcal{T}$ and some path in $\mathcal{P}$, respectively. For every path $p_0 \in P_t \cup P_p$, our algorithm performs the following queries on $\mathcal{D}$:

(i) For each vertex w in p_0, we query each path in $\mathcal{P}$ for an edge incident on the vertex w. Thus, the total number of edges added to L by these queries is $O(n|\mathcal{P}|)$.

(ii) If p_0 belongs to P_p, then we query for an edge from each $\tau \in \mathcal{T}$ to p_0. It follows from the path halving technique that each path in $\mathcal{P}$ reduces to at most half of its length whenever some path is extracted from it and attached to T^*. Hence, the size of P_p is bounded by $|\mathcal{P}| \log n$.

(iii) If p_0 belongs to P_t, then we query for an edge from only those subtrees which were hanging from p_0. Note that these subtrees will now be added to set $\mathcal{T}$.

Hence, the total number of trees queried for this case will be bounded by the number of trees inserted to $\mathcal{T}$. Since each subtree can be added to $\mathcal{T}$ only once, these edges are bounded by $O(n)$ throughout the algorithm.

Thus, the size of L is bounded by $O\big(n(1+|\mathcal{P}|)\log n\big)$. Since each edge added to L requires querying the data structure $\mathcal{D}$ which takes $O(\log^3 n)$ time, the total time taken to compute L is $O\big(n(1+|\mathcal{P}|\log n)\log^3 n\big)$. Thus, we have the following lemma.

LEMMA 11. *An undirected graph can be preprocessed in $O(m\log n)$ time to build a data structure of $O(m\log n)$ size such that for any set U of k failed vertices or edges (where $k \le n$), the DFS tree of $G+U$ can be reported in $O(n(1+|\mathcal{P}|\log n)\log^3 n)$ time.*

From Definition 7 we have that $|\mathcal{P}|$ is bounded by $|U|$. Thus, we have the following theorem.

THEOREM 12. *An undirected graph can be preprocessed in $O(m\log n)$ time to build a data structure of $O(m\log n)$ size such that for any set U of k failed vertices or edges (where $k \le n$), the DFS tree of $G+U$ can be reported in $O(nk\log^4 n)$ time.*

It can be observed that Theorem 12 directly implies a data structure for a fault tolerant DFS tree.

7.4. Extending the algorithm to handle insertions. In order to update the DFS tree, our focus has been to restrict the number of edges that are processed. For the case when the updates are deletions only, we have been able to restrict this number to $O(nk\log n)$ for a given set of k updates (failure of vertices or edges). We now describe the procedure to handle vertex and edge insertions. Let V_I be the set of vertices inserted, and let E_I be the set of edges inserted (including the edges incident to the vertices in V_I). If there are k vertex insertions, the size of E_I is bounded by nk. So even if we add all the edges in E_I to the reduced adjacency lists, the size of L would still be bounded by $O(nk\log n)$. Hence, we perform the following two additional steps before starting the DFS traversal:

- Initialize $L(v)$ to store the edges in E_I instead of $\emptyset$. That is, $L(v) \leftarrow \{y \mid (y,v) \in E_I\}$.
- Each newly inserted vertex is treated as a singleton tree and added to $\mathcal{T}$. That is, $\mathcal{T} \leftarrow \mathcal{T} \cup \{x | x \in V_I\}$.

In order to establish that our algorithm, after incorporating the insertions, correctly computes a DFS tree of $G+U$, we need to ensure that all the edges *essential* for the DFS traversal as described in the adapted components property are added to L. All the essential edges from G are added to L during the algorithm itself. In case an essential edge belongs to E_I, the edge has already been added to L during its initialization. Note that the time taken by our algorithm remains unchanged since the size of L remains bounded by $O(nk\log n)$. This completes the proof of our main result, stated in Theorem 1.

Let us consider the case when U consists of insertions only. In this case, $\mathcal{P}$ will be an empty set. As discussed above, we initialize the reduced adjacency lists using E_I, whose size is equal to $|U|$. Additionally, since the vertices in V_I would be added to the set of trees, $|V_I|$ would be added to n. Hence, Lemma 11 implies the following theorem.

THEOREM 13. *An undirected graph can be preprocessed in $O(m\log n)$ time to build a data structure of $O(m\log n)$ size such that for any set U of k vertex insertions and m' edge insertions, a DFS tree of $G+U$ can be reported in $O(m'+(n+k)\log^3 n)$ time.*

Note. In Theorem 13, the size of the input is $k+m'$. Also, even a single insertion may change $\Omega(n)$ edges of the DFS tree. Hence, our algorithm is optimal up to $\tilde{O}(1)$ factors for processing edge or vertex insertions if the DFS tree has to be maintained explicitly.

8. Fully dynamic DFS. We now describe the overlapped periodic rebuilding technique to convert our algorithm for computing a DFS tree after k updates to fully dynamic and incremental algorithms for maintaining a DFS tree. A similar technique was previously used by Thorup [59] for maintaining fully dynamic all pairs shortest paths.

In the fully dynamic model, we need to report the DFS tree after every update in the graph. Given the data structure $\mathcal{D}$ built using the DFS tree of the graph G, we are able to report the DFS tree of $G+U$ after $|U|=k$ updates in $\tilde{O}(nk)$ time. This becomes inefficient if k becomes large. Rebuilding $\mathcal{D}$ after every update is also inefficient, as it takes $\tilde{O}(m)$ time to build $\mathcal{D}$. Thus, it is better to rebuild $\mathcal{D}$ after every $|U'|=c$ updates for a carefully chosen c. Let $\mathcal{D}'$ be the data structure built using the DFS tree of the updated graph $G+U'$, with $|U'|=c$. $\mathcal{D}'$ can thus be used to process the next c updates efficiently (see Figure 10(a)). The cost of building $\mathcal{D}'$ can thus be amortized over these c updates.

To achieve an efficient worst case update time, we divide the building of $\mathcal{D}'$ over the first c updates. This $\mathcal{D}'$ is then used by our algorithm in the next c updates, during which a new $\mathcal{D}''$ is built in a similar manner, and so on (see Figure 10(b)). The following lemma describes how this technique can be used in general for any dynamic graph problem. For notational convenience, we denote any function $f(m,n)$ as f.

LEMMA 14. *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 $O(h+|U|\times g)$ time, then D can be used to report the solution after every update in worst case $O(\sqrt{fg}+h)$ update time, given that $\sqrt{f/g}\le n$.*

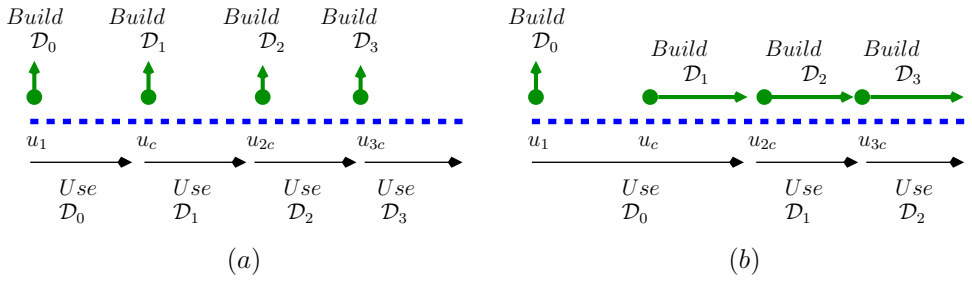

FIG. 10. (a) *Fully dynamic algorithm with amortized update time.* (b) *De-amortization of the algorithm.*

Proof. We first present an algorithm that achieves amortized $O(\sqrt{fg}+h)$ update time. It is based on the simple idea of periodic rebuilding. Given the input graph G_0, we preprocess it to compute the data structure D_0 over it. Now, let $u_1,\ldots,u_c$ ($c\le n$) be the sequence of first c updates on G_0. To report the solution after the ith update, we use D_0 to compute the solution for $G_0+\{u_1,\ldots,u_i\}$. This takes

$O(h + (i \times g))$ time. So the total time for preprocessing and handling the first c updates is $O(f + \sum_{i=1}^{c} h + (i \times g))$. Therefore, the average time for the first c updates is $O(f/c + c \times g + h)$. Minimizing this quantity over c gives the optimal value $c_0 = \sqrt{f/g}$, which is bounded by n. So after every c_0 updates we rebuild our data structure and use it for the next c_0 updates (see Figure 10(a)). Substituting the value of c_0 gives the amortized time complexity as $O(\sqrt{fg} + h)$.

The above algorithm can be de-amortized as follows. Let $G_1, G_2, G_3, \ldots$ be the sequence of graphs obtained after $c_0, 2c_0, 3c_0, \ldots$ updates. We use the data structure D_0 built during preprocessing to handle the first $2c_0$ updates. Also, after the first c_0 updates we start building the data structure D_1 over G_1. This D_1 is built in c_0 steps, and thus the extra time spent per update is $f/c_0 = O(\sqrt{fg})$ only. We use D_1 to handle the next c_0 updates on graph G_2 and also in parallel to compute the data structure D_2 over the graph G_2. (See Figure 10(b).) Since the time for building each data structure is now divided in c_0 steps, we have that the worst case update time is $O(\sqrt{fg} + h)$. ∎

The above lemma combined with Theorems 1 and 13 directly implies the following results for the fully dynamic DFS tree problem and the incremental DFS tree problem, respectively.

(For the following theorem, we use Theorem 1, implying $f = m \log n$, $g = n \log^4 n$ and $h = 0$.)

THEOREM 15. *There exists a fully dynamic algorithm for maintaining a DFS tree in an undirected graph that uses $O(m \log n)$ preprocessing time and can report a DFS tree after each update in worst case $O(\sqrt{mn} \log^{2.5} n)$ time. An update in the graph can be insertion/deletion of an edge as well as a vertex.*

(For the following theorem, we use Theorem 13, implying $f = m \log n$, $g = \log^3 n$ and $h = n \log^3 n$.)

THEOREM 16. *There exists an incremental algorithm for maintaining a DFS tree in an undirected graph that uses $O(m \log n)$ preprocessing time and can report a DFS tree after each edge insertion in worst case $O(n \log^3 n)$ time.*

9. Applications. Our fully dynamic algorithm for maintaining a DFS tree can be used to solve various dynamic graph problems, such as dynamic subgraph connectivity, biconnectivity, and 2-edge connectivity. Note that these problems are solved trivially using a DFS tree in the static setting. Let us now describe the importance of our result in light of the existing results for these problems.

9.1. Existing results. The dynamic subgraph connectivity problem is defined as follows. Given an undirected graph, the status of any vertex can be switched between *active* and *inactive* in an update. For any online sequence of updates interspersed with queries, the goal is to efficiently answer each connectivity query on the subgraph induced by the active vertices. This problem can be solved by using dynamic connectivity data structures [24, 27, 37, 41] that answer connectivity queries under an online sequence of edge updates. This is because switching the state of a vertex is equivalent to $O(n)$ edge updates. Chan [12] introduced this problem and showed that it can be solved more efficiently. He gave an algorithm using fast matrix multiplication (FMM) that achieves $O(m^{0.94})$ amortized update time and $\tilde{O}(m^{1/3})$ query time. Later, Chan, Patrascu, and Roditty [13] presented a new algorithm that improves the amortized update time to $\tilde{O}(m^{2/3})$. They also mentioned the following among the open problems:

1. Is it possible to achieve constant query time with worst case sublinear ($o(m)$) update time?
2. Can nontrivial updates be obtained for richer queries such as counting the number of connected components?

Duan [22] partially answered the first question affirmatively but at the expense of a much higher update time and nonconstant query time. He presented an algorithm with $O(m^{4/5})$ worst case update time and $O(m^{1/5})$ query time, improving the worst case bounds for the problem. Kapron, King, and Mountjoy [41] presented a randomized algorithm for fully dynamic connectivity which takes $\tilde{O}(1)$ time per update and answers the query correctly with high probability in $\tilde{O}(1)$ time, giving a Monte Carlo algorithm for subgraph connectivity with worst case $\tilde{O}(n)$ update time. Thus, their result answered the first question in a randomized setting. However, in the deterministic setting both of these questions were still open. Our result answers both of these questions affirmatively for the deterministic setting as well. Our fully dynamic algorithm directly provides an $\tilde{O}(\sqrt{mn})$ update time and $O(1)$ query time algorithm for the dynamic subgraph connectivity problem. Our algorithm maintains the number of connected components simply as a by-product. In fact, our fully dynamic algorithm for a DFS tree solves a generalization of dynamic subgraph connectivity—in addition to just switching the status of vertices, it allows insertion of new vertices as well. Hence, the existing results offer different trade-offs between the update time and the query time, and differ on the types (amortized or worst case) of update time and the types (deterministic or randomized) of query time. Our algorithm, in particular, improves the deterministic worst case bounds for the problem (see Figure 11). Further, unlike all the previous algorithms for dynamic subgraph connectivity, which use the heavy machinery of existing dynamic algorithms, our algorithm is arguably much simpler and self contained.

References	Update time	Query time
Frederickson [27] (1985), Eppstein et al. [24] (1997)	$O(n\sqrt{n})$	$O(1)$
Holm, de Lichtenberg, and Thorup [37] (2001)	$\tilde{O}(n)$ amortized	$\tilde{O}(1)$
Chan [12] (2006)	$\tilde{O}(m^{0.94})$ amortized	$\tilde{O}(m^{1/3})$
Chan, Patrascu, and Roditty [13] (2008)	$\tilde{O}(m^{2/3})$ amortized	$\tilde{O}(m^{1/3})$
Duan [22] (2010)	$\tilde{O}(m^{4/5})$	$\tilde{O}(m^{1/5})$
Kapron, King, and Mountjoy [41] (2013)	$\tilde{O}(n)$	$\tilde{O}(1)$ (Monte Carlo)
New	$\tilde{O}(\sqrt{mn})$	$O(1)$

FIG. 11. *Current state-of-the-art of the algorithms for the dynamic subgraph connectivity.*

Remark 9.1. After the preliminary version of this article [4] was published, there were several results that improved the bounds for dynamic subgraph connectivity under various settings. Duan and Zhang [23] improved the worst case update time using

a randomized algorithm (Monte Carlo) requiring $\tilde{O}(m^{3/4})$ update time answering each query correctly with high probability in $\tilde{O}(m^{1/4})$ time. Further, a series of new results [61, 46, 47] presented randomized (Las Vegas) worst case bounds for dynamic connectivity [47], eventually implying $O(n^{1+o(1)})$ update time with high probability for dynamic subgraph connectivity.

Exploiting the rich structure of DFS trees, we also obtain $\tilde{O}(\sqrt{mn})$ update time algorithms for dynamic biconnectivity and dynamic 2-edge connectivity under vertex updates in a seamless manner. These problems have mainly been studied in the dynamic setting under edge updates. Some of these results also allow insertion and deletion of isolated vertices. Our result, on the other hand, does not impose any such restriction on insertion or deletion of vertices. Figure 12 illustrates our results and the existing results in the right perspective. We now describe how our algorithm can be used to solve these problems.

References	Update time	Query time
Frederickson [27] (1985), Eppstein et al. [24] (1997)†	$O(n\sqrt{n})$	$O(1)$
Henzinger [35] (2000) ∗	$\tilde{O}(n\sqrt{n})$	$O(1)$
Holm, de Lichtenberg, and Thorup [37] (2001) ∗†	$\tilde{O}(n)$ amortized	$\tilde{O}(1)$
New ∗†	$\tilde{O}(\sqrt{mn})$	$O(1)$

FIG. 12. *Current state-of-the-art of the algorithms for the dynamic biconnectivity* (∗) *and dynamic* 2*-edge connectivity* (†) *under vertex updates.*

9.2. Algorithm. The solution of dynamic subgraph connectivity follows seamlessly from our fully dynamic algorithm as follows. As mentioned in section 2, we maintain a DFS tree rooted at a dummy vertex r, such that the subtrees hanging from its children correspond to the connected components of the graph. Hence, the connectivity query for any two vertices can be answered by comparing their ancestors at depth two (i.e., children of r). This information can be stored for each vertex and updated whenever the DFS tree is updated. Thus, we have a data structure for subgraph connectivity with worst case $\tilde{O}(\sqrt{mn})$ update time and $O(1)$ query time. Our fully dynamic DFS algorithm can be extended to solve fully dynamic biconnectivity and 2-edge connectivity under vertex updates as follows.

A set S of vertices in a graph is called a *biconnected component* if it is a maximal set of vertices such that on failure of any vertex w in S, the vertices of $S \setminus \{w\}$ remain connected. Similarly, a set S is said to be 2*-edge connected component* if it is a maximal set of vertices such that on failure of any edge with both endpoints in S, the vertices of S remain connected. The biconnectivity and 2-edge connectivity queries can be answered easily by finding *articulation points* and *bridges* of the graph. It can be shown [18] that two vertices belong to the same biconnected component if and only if the path connecting them in a DFS tree of the graph does not pass through any *articulation point*. Similarly, two vertices belong to the same 2-edge connected component if and only if the path connecting them in a DFS tree of the graph does not have a *bridge*. An articulation point and a bridge of a graph can be defined as follows.

DEFINITION 17. *Given a graph $G = (V, E)$, a vertex $v \in V$ is called an articula-*

tion point of G if there exist a pair of vertices $x, y \in V$ such that every path between x and y in G passes through v.

DEFINITION 18. *Given a graph $G = (V, E)$, an edge $e \in E$ is called a bridge of G if there exist a pair of vertices $x, y \in V$ such that every path between x and y in G passes through e.*

The articulation points and bridges of a graph can be easily computed by using a DFS traversal of the graph. Given a DFS tree T of an undirected graph G, we can index the vertices in the order they are visited by the DFS traversal. This index is called the *DFN number* of the vertex. The *high number* of a vertex v is defined as the lowest DFN number among the vertices from which there is an edge incident to $T(v)$. Now, any nonroot vertex v will be an articulation point of the graph if the high number of at least one of its children is equal to $DFN(v)$. The root of the DFS tree T will be an articulation point if it has more than one child. An edge (x, y) of the DFS tree, where $x = par(y)$, will be a bridge if the high number of y is $DFN(x)$ and the high number of each child of y (if any) is equal to $DFN(y)$. Thus, given the high number of each vertex in the DFS tree, the articulation points and bridges can be determined in $O(n)$ time.

Remark 9.2. Adding the dummy vertex r with edges incident to all vertices affects articulation points and bridges (and hence biconnectivity and 2-edge connectivity) of the graph. Hence, the dummy vertex r and the edges incident on it have to be ignored for the computation of the high number of a vertex. To address this issue, the subtrees hanging from r are treated as individual DFS trees with the children of r being the corresponding roots during the computation of high numbers.

We can augment our fully dynamic DFS algorithm with an additional procedure to compute the high number of each vertex using the same time bounds. For this, we show that given any set of k updates to graph G, while computing the new tree T^* we also compute the high number of each vertex in $O(nk \log^4 n)$ time. For each vertex x, let $a(x)$ denote the highest ancestor of x in T^* such that $(x, a(x))$ is an edge in $G+U$. Note that if $(x, a(x))$ is a newly added edge, then it can be easily computed by scanning all the new edges added to the graph. This is due to the fact that the total number of new edges added to G is bounded by nk. So we restrict ourselves to the case when $(x, a(x))$ was originally present in the graph G. Recall that our algorithm computes T^* by attaching paths to the partially grown tree. Let P_t and P_p be the set of paths attached to T^* (during its construction) that originally belonged to $\mathcal{T}$ and $\mathcal{P}$, respectively. Further, path halving ensures that the size of P_p is bounded by $k \log n$. For each path $p_0 \in P_t \cup P_p$, let $H(p_0)$ denote the vertex in p_0 that is closest to r in T^*.

We now present the procedure for constructing a subset $A(x)$ of neighbors of x while computing T^* in $O(nk \log^4 n)$ time, such that the following condition holds:

- For a vertex x, if $a(x) \notin A(x)$, then there is some descendant y of x in T^* such that $a(x) \in A(y)$.

It is easy to see that if we get such an $A(x)$ for each x, then the high number of each vertex can be computed easily by processing the vertices of T^* in bottom-up manner. Now, depending upon whether paths containing x and $a(x)$ belong to set P_p or P_t, we can have different cases described as follows:

1. Vertex $a(x)$ lies on a path in P_p.
   For every vertex $v \in V$ and each path $p_0 \in P_p$, we query $\mathcal{D}$ to compute the edge (u, v), where u is closest to $H(p_0)$ on path p_0, and add u to $A(v)$. Note

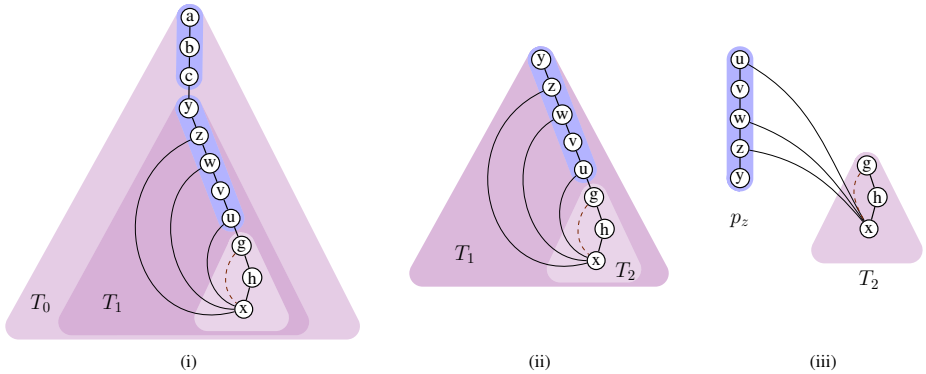

FIG. 13. (i) *Before the beginning of the algorithm vertex x belongs to tree $T_0 \in \mathcal{T}$, z is the highest ancestor of x in T_0 such that (x, z) is an edge.* (ii) *The partitioning changes as the algorithm proceeds, and $T_1 (\in \mathcal{T})$ is the tree containing vertex z just before it is attached to T^*.* (iii) *A path containing vertex z (i.e., p_z) is extracted from T_1 and attached to T^*. If $a(x)$ belongs to T_0, then it is the highest neighbor of x in p_z.*

that if $a(x)$ lies on p_0, for $v = x$ the computed vertex u will be the same as $a(x)$.

2. Vertex x lies on a path in P_p.
   For each $u \in T^*$ and $p_0 \in P_p$, we query $\mathcal{D}$ for an edge (u, y) such that the endpoint y is farthest from $H(p_0)$ on path p_0. We add u to $A(y)$. Now, consider a vertex x on p_0 such that $a(x) = u$. If x is equal to y, then we have added $a(x)$ (i.e., u) to $A(x)$. If x is not equal to y, then we have added $a(x)$ (i.e., u) to $A(y)$, where y is a descendant of x in T^*.
3. Vertices x and $a(x)$ lie on the same path in P_t.
   For every vertex $v \in p_0$ for a path $p_0 \in P_t$, we query $\mathcal{D}$ to compute the edge (u, v), where u is closest to $H(p_0)$ on path p_0, and add u to $A(v)$. Note that for $x = v$, if $a(x)$ also lies on p_0, then u will be the same as $a(x)$.
4. Vertices x and $a(x)$ lie on different paths in P_t.
   Let x belong to T_0 in the initial disjoint tree partitioning $\mathcal{T} \cup \mathcal{P}$. We claim that $a(x)$ would also belong to the same tree T_0. This is because disjoint tree partitioning ensures the absence of edges between two subtrees in $\mathcal{T}$. Let z be the highest ancestor of x in T_0 such that (x, z) is an edge in $G + U$. Let p_z be the path in P_t containing vertex z.
   We now prove that $a(x)$ belongs to p_z. Recall that as the algorithm proceeds, our partitioning $\mathcal{P} \cup \mathcal{T}$ evolves with time. Let T_1 be the tree in $\mathcal{T}$ containing vertex z just before p_z is attached to T^*. Then T_1 is either the same as T_0 or a subtree of T_0 (see Figure 13(i)). Also, $a(x)$ must lie in tree T_1 since it cannot be an ancestor of z in T_0. Now, let T_2 be the tree containing x which is obtained on removal of p_z from T_1. Since z is an ancestor of x in T_0, the vertices in T_2 will eventually hang from some descendant of z (not necessarily proper) in T^*. For $a(x)$ to be the highest neighbor of x in T^*, it should be an ancestor of z in T^*, which is only possible if $a(x) \in p_z$.
   Therefore, for each vertex x belonging to a tree T_0 in $\mathcal{T}$, we calculate the highest ancestor z of x in T_0 such that (x, z) is an edge in $G + U$. We compute a list $l(z)$ that consist of all the vertices x whose highest ancestor in T_0 to which x has an edge is z. Now, when p_z is added to T^*, we process

$l(z)$ as follows. For every $v \in l(z)$, we query $\mathcal{D}$ for an edge (u, v), where u is closest to $H(p_z)$ on path p_z, and add u to $A(v)$. Note that if $a(x)$ also lies in T_0, then u must be the same as $a(x)$ (see Figure 13(iii)).

Now, in the first two steps the total time taken is dominated by the number of queries between each path in P_p and the vertices in T, i.e., $|P_p| \times n \times \log^3 n = O(nk \log^4 n)$. In the last two steps, the total time taken is dominated by a single query for each vertex in T, i.e., $n \times \log^3 n = O(n \log^3 n)$. Thus, we have the following theorem.

THEOREM 19. *Given an undirected graph $G(V, E)$, with $|V| = n$ and $|E| = m$, we can maintain a data structure for answering queries of biconnected components and 2-edge connectivity in a dynamic graph which takes $O(\sqrt{mn} \log^{2.5} n)$ update time, $O(1)$ query time, and $O(m \log n)$ time for preprocessing.*

10. Lower bounds. We now prove two conditional lower bounds for maintaining a DFS tree under fully/partially dynamic vertex or fully dynamic edge updates.

10.1. Vertex updates. The lower bound for maintaining a DFS tree under vertex updates is based on the strong exponential time hypothesis (SETH) as defined below.

DEFINITION 20 (SETH [11, 40]). *For every $\epsilon > 0$, there exists a positive integer k, such that SAT on k-CNF formulas on n variables cannot be solved in $\tilde{O}(2^{(1-\epsilon)n})$ time.*

Given an undirected graph G on n vertices and m edges in a dynamic environment (fully dynamic or partially dynamic, i.e., incremental/decremental) under vertex updates, the status of any vertex can be switched between *active* and *inactive* in an update. The goal of subgraph connectedness is to efficiently answer whether the subgraph induced by the active vertices is connected. Abboud and Vassilevska Williams [1] proved a conditional lower bound of $\Omega(n)$ per update based on SETH for answering dynamic subgraph connectedness queries. They proved that any algorithm for answering dynamic subgraph connectedness queries using arbitrary polynomial preprocessing time and $O(n^{1-\epsilon})$ amortized update and query time would essentially refute SETH. They also proved that any algorithm for maintaining partially dynamic (incremental/decremental) subgraph connectedness using arbitrary polynomial preprocessing time and $O(n^{1-\epsilon})$ worst case update and query time would essentially refute SETH.

We present a simple reduction from subgraph connectedness to maintaining a DFS tree under vertex updates requiring the algorithm to report whether the number of children of the root in any DFS tree of the subgraph is greater than 1. Thus, we establish the following.

THEOREM 21. *Given an undirected graph G with n vertices and m edges undergoing vertex updates, an algorithm for maintaining a DFS tree that can report the number of children of the root in the DFS tree with preprocessing time $p(m, n)$, update time $u(m, n)$, and query time $q(m, n)$ would imply an algorithm for subgraph connectedness with preprocessing time $p(m + n, n)$, update time $u(m + n, n)$, and query time $q(m + n, n)$.*

Proof. Given the graph G, for which we need to query for subgraph connectedness, we make a graph G' as follows. We add all vertices and edges of G to G'. Further, add another vertex r called as *pseudoroot* and connect it to all other vertices of G'. Thus, G' has $n+1$ vertices and $m+n$ edges. Now, in any DFS tree T of G' rooted at

r, the number of children of r will be equal to the number of components in G. Here, the subtree rooted at each child of r represents a component of G. Any change on G can be performed on G', and querying for subgraph connectedness in G is equivalent to querying whether r has more than one child in T. ∎

Thus, any algorithm for maintaining *fully dynamic* DFS under vertex updates with arbitrary preprocessing time and $O(n^{1-\epsilon})$ amortized update time would imply the same bound for dynamic subgraph connectedness, refuting SETH [1]. Similarly, any algorithm for maintaining *partially dynamic* DFS under vertex updates with arbitrary preprocessing time and $O(n^{1-\epsilon})$ worst case update time would imply the same bound for dynamic subgraph connectedness, refuting SETH [1].

10.2. Edge updates. We now present a lower bound for maintaining a DFS tree under fully dynamic edge updates that holds for any algorithm which maintains tree edges of the DFS tree explicitly. In the following example, we prove that there exist a graph G and a sequence of edge updates U, such that any DFS tree of the graph would require a conversion of $\Omega(n)$ edges from tree edges to back edges and vice versa after every pair of updates in U.

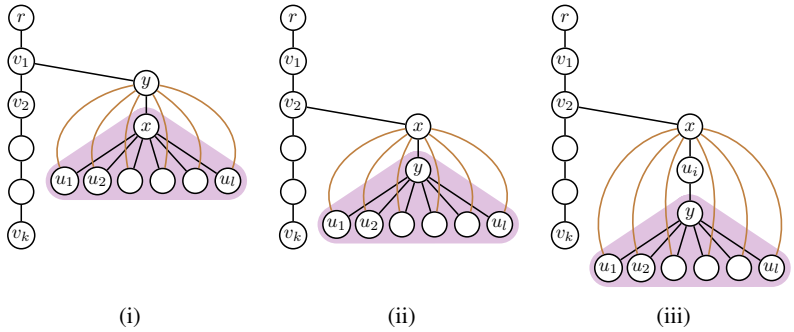

FIG. 14. *Worst case example for a lower bound on maintaining a DFS tree under fully dynamic edge updates.*

Consider the following graph for which a DFS tree rooted at r is to be maintained under fully dynamic edge updates. There are $n/2$ vertices $u_1,\ldots,u_l$ that have edges to vertices x and y. The remaining $n/2-3$ vertices $v_1,\ldots,v_k$ are connected in the form of a line as shown in Figure 14. At any point of time, one of $v_1,\ldots,v_k$ (say v_1) is connected to either x or y. The DFS tree for the graph is shown in Figure 14(i). Now, upon insertion of edge (v_i,x) (say $i=2$) and deletion of edge (v_1,y), the DFS tree will transform to either Figure 14(ii) or Figure 14(iii). Clearly, $\Omega(n)$ edges are converted from tree edges to back edges and vice versa. This can be repeated alternating between x and y, ensuring that the new DFS tree requires $\Omega(n)$ after every two edge updates. Further, we repeat this for different v_i's, ensuring that the new DFS tree is not exactly the same as some previous DFS tree. Note that the same procedure can be applied to both of the possible trees shown in Figure 14(ii) and Figure 14(iii). Hence, any algorithm maintaining tree edges explicitly takes $\Omega(n)$ time to handle such a pair of edge updates.

Note. Choosing different v_i's is not required to prove the bound when tree edges are maintained *explicitly*. However, in case we always use the same v_i, one can possibly relax explicit maintenance slightly to achieve significantly better bounds using *memorization* (i.e., store the complete DFS tree for the different cases). This par-

ticular example shows that the lower bound holds even when *explicit* maintenance is *slightly* relaxed as described above.

11. Conclusion. We have presented a fully dynamic algorithm for maintaining a DFS tree that takes worst case $\tilde{O}(\sqrt{mn})$ update time. This is the first fully dynamic algorithm that achieves $o(m)$ update time for dense graphs. In the fault tolerant setting, our algorithm takes $\tilde{O}(nk)$ time to report a DFS tree, where k is the number of vertex or edge failures in the graph. We show the immediate applications of fully dynamic DFS for solving various problems, such as dynamic subgraph connectivity, biconnectivity, and 2-edge connectivity. We also prove a conditional lower bound of $\Omega(n)$ for maintaining DFS trees under fully/partially dynamic vertex updates or fully dynamic edge updates.

DFS trees have been extensively used for solving various graph problems in the static setting. Most of these problems are also solved efficiently in the dynamic environment. However, their solutions have not used dynamic DFS trees. Furthermore, solutions to most dynamic graph problems under edge updates require $o(n)$ update time. However, this is not true for the vertex update variants of these problems. In light of the $\Omega(n)$ lower bound for updating DFS under both fully dynamic edge and fully/partially dynamic vertex updates, it becomes clear that dynamic DFS trees would be more applicable in dynamic graph problems under vertex updates. The applications of our fully dynamic algorithm follow from the fact that it handles vertex updates, which was not the case with the existing algorithms for maintaining a DFS tree in any dynamic setting. This paper is thus an attempt to restore the glory of DFS trees for solving graph problems in the dynamic setting, as was the case in the static setting. We believe that our dynamic algorithm for DFS, on its own or after further improvements/modifications, would encourage other researchers to use it in solving various other dynamic graph problems.

REFERENCES

[1] A. Abboud and V. Vassilevska Williams, *Popular conjectures imply strong lower bounds for dynamic problems*, in Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2014), 2014, pp. 434–443.

[2] A. Aggarwal and R. J. Anderson, *A random NC algorithm for depth first search*, Combinatorica, 8 (1988), pp. 1–12.

[3] A. Aggarwal, R. J. Anderson, and M.-Y. Kao, *Parallel depth-first search in general directed graphs*, SIAM J. Comput., 19 (1990), pp. 397–409, https://doi.org/10.1137/0219025.

[4] S. Baswana, S. R. Chaudhury, K. Choudhary, and S. Khan, *Dynamic DFS in undirected graphs: breaking the $O(m)$ barrier*, in Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2016), SIAM, Philadelphia; ACM, New York, 2016, pp. 730–739.

[5] S. Baswana and K. Choudhary, *On dynamic DFS tree in directed graphs*, in Mathematical Foundations of Computer Science, Part II, Springer, Heidelberg, 2015, pp. 102–114.

[6] S. Baswana, M. Gupta, and S. Sen, *Fully dynamic maximal matching in $O(\log n)$ update time (corrected version)*, SIAM J. Comput., 47 (2018), pp. 617–650, https://doi.org/10.1137/16M1106158.

[7] S. Baswana and S. Khan, *Incremental algorithm for maintaining DFS tree for undirected graphs*, in International Colloquium on Automata, Languages and Programming, Part I, Springer, Heidelberg, 2014, pp. 138–149.

[8] S. Baswana and N. Khanna, *Approximate shortest paths avoiding a failed vertex: Near optimal data structures for undirected unweighted graphs*, Algorithmica, 66 (2013), pp. 18–50.

[9] S. Baswana, S. Khurana, and S. Sarkar, *Fully dynamic randomized algorithms for graph spanners*, ACM Trans. Algorithms, 8 (2012), 35.

[10] G. Braunschvig, S. Chechik, D. Peleg, and A. Sealfon, *Fault tolerant additive and $((\mu), (\alpha))$-spanners*, Theoret. Comput. Sci., 580 (2015), pp. 94–100.

[11] C. Calabro, R. Impagliazzo, and R. Paturi, *The complexity of satisfiability of small depth circuits*, in Parameterized and Exact Computation, Springer, Berlin, 2009, pp. 75–85.
[12] T. M. Chan, *Dynamic subgraph connectivity with geometric applications*, SIAM J. Comput., 36 (2006), pp. 681–694, https://doi.org/10.1137/S009753970343912X.
[13] T. M. Chan, M. Patrascu, and L. Roditty, *Dynamic connectivity: Connecting to networks and geometry*, in Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2008), 2008, pp. 95–104.
[14] B. Chazelle and L. J. Guibas, *Fractional cascading:* I. *A data structuring technique*, Algorithmica, 1 (1986), pp. 133–162.
[15] S. Chechik, M. Langberg, D. Peleg, and L. Roditty, *Fault tolerant spanners for general graphs*, SIAM J. Comput., 39 (2010), pp. 3403–3423, https://doi.org/10.1137/090758039.
[16] S. Chechik, M. Langberg, D. Peleg, and L. Roditty, *f-sensitivity distance oracles and routing schemes*, Algorithmica, 63 (2012), pp. 861–882.
[17] L. Chen, R. Duan, R. Wang, H. Zhang, and T. Zhang, *An improved algorithm for incremental DFS tree in undirected graphs*, in Proceedings of the 16th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2018), Malmö, Sweden, 2018, 16.
[18] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed., MIT Press, Cambridge, MA, 2009.
[19] C. Demetrescu and G. F. Italiano, *A new approach to dynamic all pairs shortest paths*, J. ACM, 51 (2004), pp. 968–992.
[20] C. Demetrescu and G. F. Italiano, *Maintaining dynamic matrices for fully dynamic transitive closure*, Algorithmica, 51 (2008), pp. 387–427.
[21] C. Demetrescu, M. Thorup, R. A. Chowdhury, and V. Ramachandran, *Oracles for distances avoiding a failed node or link*, SIAM J. Comput., 37 (2008), pp. 1299–1318, https://doi.org/10.1137/S0097539705429847.
[22] R. Duan, *New data structures for subgraph connectivity*, in International Colloquium on Automata, Languages and Programming, Part I, Springer, Berlin, 2010, pp. 201–212.
[23] R. Duan and L. Zhang, *Faster randomized worst-case update time for dynamic subgraph connectivity*, in Algorithms and Data Structures, Springer, Cham, 2017, pp. 337–348.
[24] D. Eppstein, Z. Galil, G. F. Italiano, and A. Nissenzweig, *Sparsification—a technique for speeding up dynamic graph algorithms*, J. ACM, 44 (1997), pp. 669–696.
[25] P. Erdős and A. Rényi, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl, 5 (1960), pp. 17–61.
[26] P. G. Franciosa, G. Gambosi, and U. Nanni, *The incremental maintenance of a depth-first-search tree in directed acyclic graphs*, Inform. Process. Lett., 61 (1997), pp. 113–120.
[27] G. N. Frederickson, *Data structures for on-line updating of minimum spanning trees, with applications*, SIAM J. Comput., 14 (1985), pp. 781–798, https://doi.org/10.1137/0214055.
[28] D. Frigioni and G. F. Italiano, *Dynamically switching vertices in planar graphs*, Algorithmica, 28 (2000), pp. 76–103.
[29] A. V. Goldberg, S. A. Plotkin, and P. M. Vaidya, *Sublinear-time parallel algorithms for matching and related problems*, in Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1988), 1988, pp. 174–185.
[30] L. Gottlieb and L. Roditty, *Improved algorithms for fully dynamic geometric spanners and geometric routing*, in Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2008), SIAM, Philadelphia; ACM, New York, 2008, pp. 591–600.
[31] R. Grossi, A. Gupta, and J. S. Vitter, *High-order entropy-compressed text indexes*, in Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2003), SIAM, Philadelphia; ACM, New York, 2003, pp. 841–850.
[32] M. Gupta and R. Peng, *Fully dynamic* $(1 + \epsilon)$*-approximate matchings*, in Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2013), Berkeley, CA, 2013, pp. 548–557.
[33] M. Henzinger, S. Krinninger, D. Nanongkai, and T. Saranurak, *Unifying and strengthening hardness for dynamic problems via the online matrix-vector multiplication conjecture*, in Proceedings of the 47th Annual ACM Symposium on Theory of Computing, 2015, pp. 21–30.
[34] M. R. Henzinger, *Fully dynamic biconnectivity in graphs*, Algorithmica, 13 (1995), pp. 503–538.
[35] M. R. Henzinger, *Improved data structures for fully dynamic biconnectivity*, SIAM J. Comput., 29 (2000), pp. 1761–1815, https://doi.org/10.1137/S0097539794263907.
[36] M. R. Henzinger and V. King, *Randomized fully dynamic graph algorithms with polylogarithmic time per operation*, J. ACM, 46 (1999), pp. 502–516.

[37] J. Holm, K. de Lichtenberg, and M. Thorup, *Poly-logarithmic deterministic fully-dynamic algorithms for connectivity, minimum spanning tree,* 2*-edge, and biconnectivity*, J. ACM, 48 (2001), pp. 723–760.
[38] J. E. Hopcroft and R. M. Karp, *An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs*, SIAM J. Comput., 2 (1973), pp. 225–231, https://doi.org/10.1137/0202019.
[39] J. E. Hopcroft and R. E. Tarjan, *Efficient planarity testing*, J. Assoc. Comput. Mach., 21 (1974), pp. 549–568.
[40] R. Impagliazzo and R. Paturi, *On the complexity of k-sat*, J. Comput. System Sci., 62 (2001), pp. 367–375.
[41] B. M. Kapron, V. King, and B. Mountjoy, *Dynamic graph connectivity in polylogarithmic worst case time*, in Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2013), SIAM, Philadelphia; ACM, New York, 2013, pp. 1131–1142, https://doi.org/10.1137/1.9781611973105.81.
[42] M. Krivelevich and B. Sudakov, *The phase transition in random graphs: A simple proof*, Random Structures Algorithms, 43 (2013), pp. 131–138.
[43] P. B. Miltersen, S. Subramanian, J. S. Vitter, and R. Tamassia, *Complexity models for incremental computation*, Theoret. Comput. Sci., 130 (1994), pp. 203–236.
[44] K. Nakamura, *Fully dynamic connectivity oracles under general vertex updates*, in Proceedings of the 28th International Symposium on Algorithms and Computation (ISAAC 2017), Phuket, Thailand, 2017, 59.
[45] K. Nakamura and K. Sadakane, *A space-efficient algorithm for the dynamic DFS problem in undirected graphs*, in WALCOM: Algorithms and Computation, Springer, Cham, 2017, pp. 295–307.
[46] D. Nanongkai and T. Saranurak, *Dynamic spanning forest with worst-case update time: Adaptive, Las Vegas, and $O(n^{1/2-\epsilon})$-time*, in Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2017), Montreal, QC, Canada, 2017, pp. 1122–1129.
[47] D. Nanongkai, T. Saranurak, and C. Wulff-Nilsen, *Dynamic minimum spanning forest with subpolynomial worst-case update time*, in Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2017), Berkeley, CA, 2017, pp. 950–961.
[48] J. H. Reif, *Depth-first search is inherently sequential*, Inform. Process. Lett., 20 (1985), pp. 229–234.
[49] J. H. Reif, *A topological approach to dynamic graph connectivity*, Inform. Process. Lett., 25 (1987), pp. 65–70.
[50] L. Roditty, *Fully dynamic geometric spanners*, Algorithmica, 62 (2012), pp. 1073–1087.
[51] L. Roditty and U. Zwick, *Improved dynamic reachability algorithms for directed graphs*, SIAM J. Comput., 37 (2008), pp. 1455–1471, https://doi.org/10.1137/060650271.
[52] L. Roditty and U. Zwick, *On dynamic shortest paths problems*, Algorithmica, 61 (2011), pp. 389–401.
[53] P. Sankowski, *Dynamic transitive closure via dynamic matrix inverse (extended abstract)*, in Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2004), 2004, pp. 509–517.
[54] D. D. Sleator and R. E. Tarjan, *A data structure for dynamic trees*, J. Comput. System Sci., 26 (1983), pp. 362–391.
[55] J. R. Smith, *Parallel algorithms for depth-first searches* I. *Planar graphs*, SIAM J. Comput., 15 (1986), pp. 814–830, https://doi.org/10.1137/0215058.
[56] S. Solomon, *Fully dynamic maximal matching in constant update time*, in Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2016), New Brunswick, NJ, 2016, pp. 325–334.
[57] R. Tarjan, *Depth-first search and linear graph algorithms*, SIAM J. Comput., 1 (1972), pp. 146–160, https://doi.org/10.1137/0201010.
[58] R. Tarjan, *Finding dominators in directed graphs*, SIAM J. Comput., 3 (1974), pp. 62–89, https://doi.org/10.1137/0203006.
[59] M. Thorup, *Worst-case update times for fully-dynamic all-pairs shortest paths*, in Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC 2005), 2005, pp. 112–119.
[60] M. Thorup, *Fully-dynamic min-cut*, Combinatorica, 27 (2007), pp. 91–127.
[61] C. Wulff-Nilsen, *Fully-dynamic minimum spanning forest with improved worst-case update time*, in Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2017), Montreal, QC, Canada, 2017, pp. 1130–1143.