International Journal of Foundations of Computer Science

FINDING SMALLEST SUPERTREES UNDER MINOR CONTAINMENT *†

NAOMI NISHIMURA

Department of Computer Science, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada.

PRABHAKAR RAGDE

Department of Computer Science, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada.

and

DIMITRIOS M. THILIKOS‡

Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, Campus Nord – Mòdul C5 – Desp. 211b, c/Jordi Girona Salgado, 1-3. E-08034, Barcelona, Spain

Received (received date)
Revised (revised date)
Communicated by Editor's name

ABSTRACT

The diversity of application areas relying on tree-structured data results in wide interest in algorithms which determine differences or similarities among trees. One way of measuring the similarity between trees is to find the smallest common superstructure or supertree, where common elements are typically defined in terms of a mapping or embedding. In the simplest case, a supertree will contain exact copies of each input tree, so that for each input tree, each vertex of a tree can be mapped to a vertex in the supertree such that each edge maps to the corresponding edge. More general mappings allow for the extraction of more subtle common elements captured by looser definitions of similarity.

We consider supertrees under the general mapping of minor containment. Minor containment generalizes both subgraph isomorphism and topological embedding; as a consequence of this generality, however, it is NP-complete to determine whether or not G is a minor of H, even for general trees. By focusing on trees of bounded degree, we obtain an $O(n^3)$ algorithm which determines the smallest tree T such that both of the input trees are minors of T, even when the trees are assumed to be unrooted and unordered.

Keywords: Supertrees, Minor Containment, Algorithms.

---

*Research supported by the Natural Sciences and Engineering Research Council of Canada and Communications and Information Technology Ontario.

†E-mail: nishi@uwaterloo.ca, plragde@uwaterloo.ca, sedthilk@lsi.upc.es

‡Research supported by the Ministry of Education and Culture of Spain, Grant number MEC-DGES SB98 0K148809.

1. Introduction

The breadth of algorithmic research on trees stems from both the simplicity of the structure and the variety of application domains. When information about a data set can be derived from its tree structure, comparisons among two or more data sets can entail determining similarities among two or more trees. Algorithms of this type have been developed in areas such as compiler design, structured text databases, theory of natural languages, computer vision [18], and computational biology (the reader is directed to a previous paper on trees [10] for further references).

Comparisons of trees range from the classical tree pattern matching problem (finding an exact copy of one tree in another) to numerous variants, including problems on multiple trees and inexact matches. Each problem can be viewed as finding a way to relate trees by mappings, where trees are related if it is possible to map vertices to sets of vertices and edges to sets of edges subject to certain constraints. Researchers have considered different types of trees (ordered, unordered, labeled, unlabeled) and different mappings between pairs of trees (exact matching, approximate matching, subgraph isomorphism, topological embedding, minor containment) [3, 5, 9, 13, 14]. In addition, researchers have measured the similarity between trees by finding the largest common subtree or smallest common supertree under various constraints [1, 4, 7, 8, 10, 12, 19].

In this paper we consider the problem of finding the smallest common supertree under minor containment. Concisely, a graph G is a minor of a graph H if it is possible to map all the vertices in G to mutually disjoint connected subgraphs in H and there exists a bijection, from the edges of G to the edges of H that are not in any of these subgraphs, such that the images of the endpoints of any edge e in G contain the endpoints of the image of e through this bijection; equivalently we can view the mapping as taking edges to paths. Minor containment is of interest due to its generality; it encompasses both subgraph isomorphism and topological embedding and is fundamental in the work of Robertson and Seymour on graph minors [17]. However, due in large part to the generality, many problems which are tractable under subgraph isomorphism and topological embedding are NP-complete for minor containment. In particular, it is NP-complete to determine whether or not one tree is a minor of another [6], but this can be determined in polynomial time when there is a degree bound of $O(\log n/\log\log n)$ [9]. We thus restrict our attention to trees of bounded degree, noting that the resultant supertree will also be of bounded degree (in contrast, a common subtree of two bounded degree trees may not have bounded degree).

Interest in supertrees under minor containment arises from their applications to editing, image clustering, genetics, chemical structure analysis, and evolution [12, 19]. Previous algorithms to find supertrees have been limited to special cases: in ordered minor containment, there is an order imposed on the children of each vertex in each input tree, and this order must be preserved by the mapping [12]; for evolutionary trees, the leaves have distinct labels and are constrained to map to other leaves [19].

2. Preliminaries

Each input to our algorithm is a bounded-degree tree (a connected undirected graph with no cycles). $V(T)$ denotes the vertices of T and $E(T)$ the edges of T. A tree T may be rooted at a distinguished vertex r; in this case we can view the rooted tree as a directed graph, with children and parents defined as in standard graph-theoretic references [2]. When processing rooted trees we will consider a *subtree* T_v of T, defined to be the subgraph of T induced by v and all its descendants. More generally, for A a subset of the children of some vertex v, we define T_A to be the subgraph induced by v, the vertices in A, and all descendants of vertices in A.

For A an arbitrary subset of vertices, $T[A]$ is defined to be the subgraph of T induced by A. Our algorithm will rely on relationships between neighborhoods of sets. We use $N_G(v)$ to denote the neighborhood of the vertex v in the graph G. We say that two subsets S_1, S_2 of the vertex set of a graph G are *touching* if either $S_1 \cap S_2 \neq \emptyset$ or there exists an edge $(v_1, v_2) \in E(G)$ for $v_i \in S_i, i = 1, 2$.

Given input trees Q and R, we wish to find a tree T such that both Q and R are minors of T and T is as small as possible. There are several equivalent definitions of minors; the most relevant one for our purposes is given below. Intuitively, a graph G is a minor of a graph H (or H is a major of G) if G can be obtained from H by a series of vertex and edge deletions and edge contractions, where a contraction of an edge (u, v) in G is the operation that replaces u and v by a new vertex whose neighbors are the vertices that were adjacent to u or v. It is not difficult to see that, for trees, the following definition is equivalent:

Definition 1 *A tree Q is a* minor *of a tree T if and only if there exists a surjection $f : V(T) \to V(Q)$ such that*

1. *for each $a \in V(Q)$, $T[f^{-1}(a)]$ is connected;*
2. *for each pair $a, b \in V(Q)$, $a \neq b$, $f^{-1}(a) \cap f^{-1}(b) = \emptyset$; and*
3. *for $S = \{(u, v) \in E(T) \mid f(u) \neq f(v)\}$, there exists a bijection $\xi : S \to E(Q)$ such that for each $e = (s, t) \in S, \xi(e) = (f(s), f(t))$.*

We call f a minor embedding *of T into Q. Intuitively, $f^{-1}(a)$ is the set of vertices of T contracted into a; (2) captures the notion that each vertex of T corresponds to exactly one vertex of Q; and (3) ensures that uncontracted edges of T are preserved in Q.*

The problem we wish to solve is that of determining the smallest common acyclic major of Q and R, henceforth called the *smallest common tree major*. For $\mathsf{sctmj}(Q, R)$ the minimum number of vertices in a common tree major of Q and R, it is not difficult to see that $\max\{|V(Q)|, |V(R)|\} \leq \mathsf{sctmj}(Q, R) \leq |V(Q)| + |V(R)|$. We observe that $\mathsf{sctmj}(Q, R) = |V(Q)|$ if and only if R is a minor of Q. Duchet [6] proved that it is NP-complete to determine whether one tree is a minor of another. It is thus easy to prove that deciding whether $\mathsf{sctmj}(Q, R) \leq k$ for two general trees Q, R is NP-complete. In view of this, we will restrict our attention to the case where the input graphs are both trees with maximum degree bounded by a fixed constant.

In the remainder of the paper we will make use of the following notational conventions. Since we will be finding a graph T such that Q and R are both minors of T, we will use f to denote the minor embedding of T into Q and g to denote the minor embedding of T into R. We will use letters near the beginning of the alphabet for vertices of Q and letters near the end of the alphabet for vertices of R.

3. Expansions

Our algorithm makes use of an *expansion*, a graph $\mathcal{E}$ representing the correlations between vertices and edges of Q and R. In the remainder of the section we establish various properties of expansions, culminating in Theorem 1, which explicitly links expansions and smallest common tree majors.

To facilitate understanding of the algorithm, it is beneficial to consider the mappings between Q, R, and a common tree major T. The edges of T correspond to edges in the input trees Q and R; we distinguish between *strong* edges, which correspond to edges in both Q and R, and *weak* edges, each of which corresponds to an edge in only one of Q and R. For f and g the minor embeddings of T into Q and R, respectively, $f^{-1}(a)$ and $g^{-1}(u)$ describe connected subgraphs of T. Since for $a \in V(Q)$ each vertex in $f^{-1}(a)$ is in $g^{-1}(u)$ for some $u \in V(R)$, we can associate a with a set of vertices in $V(R)$ whose preimages in T overlap the preimage of a. This notion of association can be formalized in a graph with bipartition $(V(Q), V(R))$ and edge set $\{(a,u) \mid f^{-1}(a) \cap g^{-1}(u) \neq \phi\}$. However, since we are searching for an unknown minimum T, we instead define (below) an *expansion* $\mathcal{E}$ of Q and R solely in terms of the properties of this kind of graph. In Lemma 2 we demonstrate that a common tree major can be extracted from any expansion; in Lemma 3 we show that there exists an expansion isomorphic to a smallest common tree major.

Definition 2 *For Q and R trees on disjoint sets of vertices, an* expansion *of Q and R is a bipartite graph $\mathcal{E} = (V(\mathcal{E}), E(\mathcal{E}))$ with bipartition $(V(Q), V(R))$ such that*

1. *the neighborhood in $\mathcal{E}$ of any vertex of $V(R)$ (respectively, $V(Q)$) induces a connected subgraph of Q (respectively, R);*
2. *$\mathcal{E}$ has no isolated vertices;*
3. *the neighborhoods in $\mathcal{E}$ of two vertices in $V(Q)$ (respectively, $V(R)$) intersect in at most one vertex; and*
4. *for every edge (a,b) in $E(Q)$, either there are edges (a,u) and (b,u) in $\mathcal{E}$ for some $u \in V(R)$, or there are edges (a,u) and (b,v) in $\mathcal{E}$ for some edge $(u,v) \in E(R)$ (and symmetrically for edges in R).*

Given an expansion $\mathcal{E}$ of Q and R, we define $T_{\mathcal{E}}$ to be a graph whose vertices are edges in $\mathcal{E}$ and whose edges are formed by condition 4 in the definition above. For an edge $(a,b) \in E(Q)$, if there are edges (a,u) and (b,u) in $\mathcal{E}$, then $\{(a,u),(b,u)\}$ is an edge in $T_{\mathcal{E}}$, and if there are edges (a,u) and (b,v) in $\mathcal{E}$ for some $(u,v) \in E(R)$, and neither (a,v) nor (b,u) is in $\mathcal{E}$, then $\{(a,u),(b,v)\}$ is in $T_{\mathcal{E}}$. Edges $(u,v) \in E(R)$ define edges in $T_{\mathcal{E}}$ in a similar fashion. In the former case we call the edge (a,b)

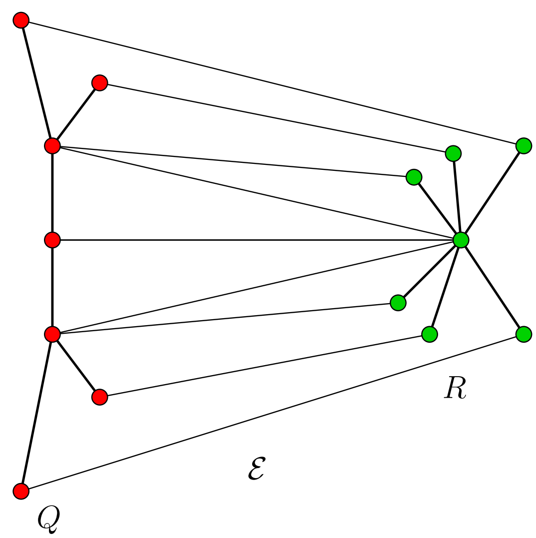

Fig. 1. An expansion $\mathcal{E}$ of Q and R.

weak; in the latter case, (a, b) and (u, v) are *strong*. Figure 1 gives an example of an expansion $\mathcal{E}$ between two trees Q and R; Figure 2 shows the corresponding $T_{\mathcal{E}}$, which by inspection is a common tree major of Q and R.

We will denote the weak (strong) edges of Q as weak(Q) (strong(Q)) and we will use analogous notation for R. We call the edges of $T_{\mathcal{E}}$ that are defined on the basis of weak edges of Q (R), *Q-weak* (*R-weak*). If an edge of $T_{\mathcal{E}}$ is not Q or R-weak, then we call it *strong*. There exist natural bijections between the weak edges of Q (R) and the Q-weak (R-weak) edges of $T_{\mathcal{E}}$, between the strong edges of Q (R) and the strong edges of $T_{\mathcal{E}}$, and between strong edges in $E(Q)$ and strong edges in $E(R)$, the last of which we denote $f_{\mathcal{E}}$. As direct consequences of the definition of weak and strong edges, $|\text{strong}(Q)| = |\text{strong}(R)|$ and $|E(T_{\mathcal{E}})| = |\text{weak}(Q)| + |\text{weak}(R)| + |\text{strong}(Q)|$. We define $|E(T_{\mathcal{E}})|$ to be the *size of the expansion $\mathcal{E}$*.

For convenience, if $\mathcal{E}$ is an expansion of two trees Q and R, (a, b) is a strong edge of Q, and $(u, v) = f_{\mathcal{E}}((a, b))$, we will say that (a, b) and (u, v) are *$\mathcal{E}$-counterparts* of each other and conclude that $(a, u), (b, v) \in \mathcal{E}$. Given a vertex t in $T_{\mathcal{E}}$ which corresponds to an edge $(a, u) \in E(\mathcal{E})$ where $a \in V(Q)$ and $u \in V(R)$, a is the *Q-side* of t and u is the *R-side* of t. Finally, we use $P_G(p_1, p_2)$ to denote the set of vertices in the (unique) path between two vertices p_1 and p_2 in the tree G.

The next lemmas are essential tools used in the proof of Theorem 1.

Lemma 1 *For any expansion $\mathcal{E}$ of Q and R and any edge $e = (a, b) \in E(Q)$ ($e \in E(R)$), $N_{\mathcal{E}}(a)$ and $N_{\mathcal{E}}(b)$ are touching in R.*

Proof. By condition 4 of the definition of $\mathcal{E}$, for any edge $(a, b) \in E(Q)$ either there is a vertex w in R such that $(a, w), (b, w) \in E(\mathcal{E})$ or there is an edge $(u, v) \in E(R)$ such that $(a, u), (b, v) \in E(\mathcal{E})$. In the first case the connected graphs $R[N_{\mathcal{E}}(a)]$ and $R[N_{\mathcal{E}}(b)]$ have a common point w, and in the second, they contain u and v respectively, and $(u, v) \in E(R)$. Therefore, in both cases, their vertex sets are

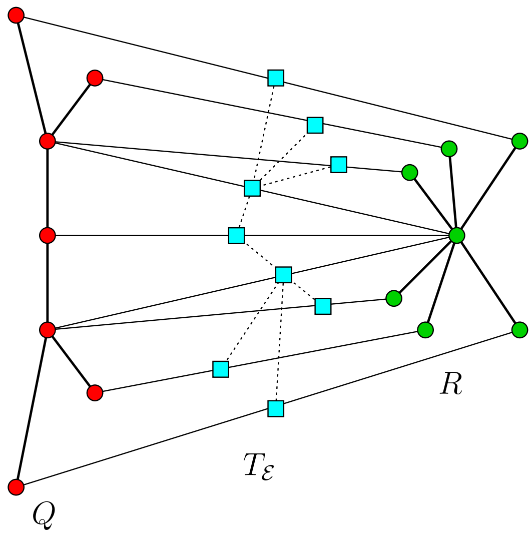

Fig. 2. A graph $T_{\mathcal{E}}$ derived from $\mathcal{E}$.

touching. □

Lemma 2 *If $\mathcal{E}$ is an expansion of two trees Q and R, then $T_{\mathcal{E}}$ is a common tree major of Q and R.*

Proof. We will prove first that $T_{\mathcal{E}}$ is a tree. By property 1 of the definition of an expansion, for any vertex a in Q, $N_{\mathcal{E}}(a)$ induces in R a tree T^a, and hence the number of edges of $\mathcal{E}$ with a as endpoint is equal to $|E(T^a)|+1$. Moreover, all the edges in T^a are weak edges of R and any weak edge e of R is in exactly one tree T^b where b is the vertex of Q adjacent to both endpoints of e. As a consequence of the above observations,

$$\begin{aligned}
|V(T_{\mathcal{E}})| &= |E(\mathcal{E})| = \sum_{a\in V(Q)} (|\text{weak edges in } R[N_{\mathcal{E}}(a)]|+1) \\
&= |V(Q)| + |\text{weak}(R)| = 1 + |E(Q)| + |\text{weak}(R)| \\
&= 1 + |\text{weak}(Q)| + |\text{strong}(Q)| + |\text{weak}(R)| \\
&= 1 + |E(T_{\mathcal{E}})|.
\end{aligned}$$

Showing that $T_{\mathcal{E}}$ is connected will prove that it is a tree. Let t_1, t_2 be two vertices in $T_{\mathcal{E}}$ and let a_1 and a_2 be their Q-sides (recall the definition of Q-side in the discussion following the definition of an expansion). We will use induction on $j = |P_Q(a_1, a_2)|$ to show that there exists a path connecting t_1 and t_2. Since a path exists trivially for $j = 1$, we suppose first that $j = 2$ and let (a_1, u_1) and (a_2, u_2) be the edges of $\mathcal{E}$ corresponding to t_1 and t_2 respectively. By Lemma 1, $N_{\mathcal{E}}(a_1)$ and $N_{\mathcal{E}}(a_2)$ are touching. Therefore, either $P_R(u_1, u_2)$ contains a vertex $u \in N_{\mathcal{E}}(a_1) \cap N_{\mathcal{E}}(a_2)$ or it contains an edge (u, u') where $u \in N_{\mathcal{E}}(a_1)$ and $u' \in N_{\mathcal{E}}(a_2)$. In the first case (a_1, u) and (a_2, u), and in the second, (a_1, u) and (a_2, u') define two adjacent vertices t and t' of $T_{\mathcal{E}}$. For any pair of edges $(v_1, v_2), (v_2, v_3)$ of $P_R(u_1, u)$, there is a pair of edges (r_1, r_2) and (r_2, r_3) in $E(T_\mathcal{E})$ where r_1, r_2, r_3 correspond to (a_1, v_1), (a_1, v_2), and (a_1, v_3) respectively. Using this observation, it is easy to see that t_1 and t are connected in $T_\mathcal{E}$. The proof of the existence of a path connecting t and t_2 in $T_\mathcal{E}$ is similar. Since t_1 and t_2 are connected, the base case of the induction holds.

Suppose now that the claim holds for $j < k$, $k \geq 3$ and let t_1 and t_2 be two vertices in $T_\mathcal{E}$ whose Q-sides are a_1 and a_2 and $|P_Q(a_1, a_2)| = k$. Let a' be the vertex in $P_Q(a_1, a_2)$ that is adjacent to a_1. There are two cases: (1) (a_1, a') is a strong edge with $\mathcal{E}$-counterpart (u^*, u') and thus $T_\mathcal{E}$ contains two adjacent vertices r, t' corresponding to the edges (a_1, u^*) and (a', u') respectively, or (2) (a_1, a') is a weak edge whose endpoints are both connected to some vertex u^* in R, and $T_\mathcal{E}$ contains two adjacent vertices r, t' corresponding to the edges (a_1, u^*) and (a', u^*), respectively.

In either case, since $|P_Q(a', a_2)| < k$, we can apply the induction hypothesis for t' and t_2 to show that t' and t_2 are connected in $T_\mathcal{E}$. The edge (t', r) shows that r and t_2 are connected. We need only consider the case where r is different from t_1. The crucial property of r and t_1 is that the edges of $\mathcal{E}$ corresponding to them, (a_1, u_1) and (a_1, u^*), both contain a_1 as the Q-side. Since the neighborhood of a_1 induces a tree R, u_1 and u^* are connected in R. Using the same arguments on t_1 and r as we did for t_1 and t in the base case, we can prove that t_1 and r are connected in $T_\mathcal{E}$ and therefore t_1 and t_2 are connected. Thus $T_\mathcal{E}$ is connected and is a tree.

In order to prove that $T_\mathcal{E}$ is a common major of Q and R we have to provide functions f and ξ as in Definition 1. We define $f : V(T_\mathcal{E}) \to V(Q)$, such that f maps every vertex of $T_\mathcal{E}$ to its Q-side and define ξ to map any edge in $T_\mathcal{E}$ whose endpoints have different Q-sides to the edge of Q that connects them. Condition 1 of Definition 1 holds because the vertices in $T_\mathcal{E}$ with the same Q-side induce a connected subgraph of $T_\mathcal{E}$. Conditions 2 and 3 are direct consequences of the way $T_\mathcal{E}$ is defined. The intuition behind the above definition of f is that a graph isomorphic to Q can be obtained from $T_\mathcal{E}$ if we contract all the R-weak edges of $T_\mathcal{E}$. This proves that Q is a minor of $T_\mathcal{E}$. The proof that R is a minor of $T_\mathcal{E}$ is symmetric. □

Lemma 3 *For T a smallest common tree major of Q and R, there exists an expansion $\mathcal{E}$ such that $T_\mathcal{E}$ is isomorphic to T.*

Proof. Given minor embeddings f and g of T into Q and R, for each $a \in V(Q)$ and each $u \in V(R)$, $|f^{-1}(a) \cap g^{-1}(u)| \leq 1$, since otherwise the minor of T obtained after contracting the edges in the graph induced by $\{f^{-1}(a) \cap g^{-1}(u)\}$ would be a smaller common tree major of Q and R. We define the expansion $\mathcal{E}$ to be the set $\{(a, u) \ : \ |f^{-1}(a) \cap g^{-1}(u)| = 1\}$. It is straightforward to verify that $\mathcal{E}$ is an expansion of Q and R. □

As a corollary of Lemmas 2 and 3, we can conclude that $\mathsf{sctmj}(Q, R)$ is the number of edges in the minimum expansion of Q and R, resulting in Theorem 1, which reduces the problem to the computation of the rooted version of expansions.

Theorem 1 *For trees Q and R and for any $a \in V(Q)$, $\mathsf{sctmj}(Q, R)$ is the minimum over all $u \in V(R)$ of the number of edges in the smallest expansion $\mathcal{E}$ of Q and R*

such that (a, u) is an edge in $\mathcal{E}$.

The next six lemmas establish properties of expansions; these are used in proving necessary technical lemmas in Section 4. The proof of the following lemma is a direct consequence of the definition of an expansion and is omitted.

Lemma 4 *Let $\mathcal{E}_i$ be a minimum size expansion of two trees Q_i and R_i for $i = 1, 2$, $V(Q_1) \cap V(Q_2) = \{a\}$, $V(R_1) \cap V(R_2) = \{u\}$, and $(a, u) \in \mathcal{E}_1 \cap \mathcal{E}_2$. Then $\mathcal{E}_1 \cup \mathcal{E}_2$ is an expansion of minimum size (among those containing (a, u)) of $Q_1 \cup Q_2$ and $R_1 \cup R_2$.*

The following two lemmas are direct applications of Lemma 4.

Lemma 5 *Let $\mathcal{E}_i$ be a minimum size expansion of two trees Q_i and R_i, $a_i \in Q_i$ for $i = 1, 2$, $V(Q_1) \cap V(Q_2) = \emptyset$, $V(R_1) \cap V(R_2) = \{u\}$, $(a_1, u) \in \mathcal{E}_1$, and $(a_2, u) \in \mathcal{E}_2$. Then $\mathcal{E}_1 \cup \mathcal{E}_2$ is an expansion of minimum size (among those containing at least one of (a_1, u) and (a_2, u)) of the graph with vertex set $V(Q_1) \cup V(Q_2)$ and edge set $E(Q_1) \cup E(Q_2) \cup \{(a_1, a_2)\}$ and the graph $R_1 \cup R_2$.*

Lemma 6 *Let $\mathcal{E}_i$ be a minimum size expansion of two disjoint trees Q_i and R_i, $a_i \in V(Q_i), u_i \in V(R_i)$, and $(a_i, u_i) \in \mathcal{E}_i$ for $i = 1, 2$. Then $\mathcal{E}_1 \cup \mathcal{E}_2$ is an expansion of minimum size (among those containing at least one of (a_1, u_1) and (a_2, u_2)) of the graph with vertex set $V(Q_1) \cup V(Q_2)$ and edge set $E(Q_1) \cup E(Q_2) \cup \{(a_1, a_2)\}$ and the graph with vertex set $V(R_1) \cup V(R_2)$ and edge set $E(R_1) \cup E(R_2) \cup \{(u_1, u_2)\}$.*

The lemma below is a useful tool in proving properties of expansions; it shows that if two pairs of vertices are related by an expansion, the paths joining the vertices are also related.

Lemma 7 *For any expansion $\mathcal{E}$ of Q and R, if $(a_i, u_i) \in \mathcal{E}, i = 1, 2$, then every vertex in $P_Q(a_1, a_2)$ is adjacent in $\mathcal{E}$ to a vertex in $P_R(u_1, u_2)$.*

Proof. We will prove the lemma by contradiction, using induction on j, the size of $P_Q(a_1, a_2)$. Since the lemma holds trivially for $j \leq 2$, it suffices to show that the lemma holds for $j = k$, assuming that it holds for all values $j < k$.

We call a vertex b in in $P_Q(a_1, a_2)$ *bad* if $N_{\mathcal{E}}(b) \cap P_R(u_1, u_2) = \phi$, and *good* otherwise. If any interior vertex in $P_Q(a_1, a_2)$ is a good vertex, then we can show that every vertex on the path has a neighbor in $P_R(u_1, u_2)$. That is, if b is a good vertex with neighbor v in $P_R(u_1, u_2)$, then we can apply the induction hypothesis on the smaller problem $P_Q(a_1, b)$ and $P_R(u_1, v)$ and also the smaller problem $P_Q(b, u_2)$ and $P_R(v, u_2)$ to reach our conclusion. We can now assume that every interior vertex in $P_Q(a_1, a_2)$ is bad.

Furthermore, we can assume that there is no vertex v in $P_R(u_1, u_2)$ which is a neighbor of both a_1 and a_2, since if there were, then by property 1 in the definition of an expansion every vertex in $P_Q(a_1, a_2)$ would also be in the neighborhood of v. Thus $N_{\mathcal{E}}(a_1) \cap N_{\mathcal{E}}(a_2) \cap P_R(u_1, u_2)$ is empty.

For each bad vertex a, we can define a vertex $v(a)$ in $P_R(u_1, u_2)$ which is the vertex in $P_R(u_1, u_2)$ closest to $N_{\mathcal{E}}(a)$ in R; this vertex is unique due to property 1 in the definition of $\mathcal{E}$. We let b_i be the neighbor of a_i in $P_Q(a_1, a_2)$ and show that $v(b_i) \in N_{\mathcal{E}}(a_i) \cap P_R(u_1, u_2)$. Suppose instead $v(b_i) \notin N_{\mathcal{E}}(a_i) \cap P_R(u_1, u_2)$. As R is a tree, we can partition the vertices of $R \setminus P_R(u_1, u_2)$ into connected subgraphs on the basis of the closest vertex in $P_R(u_1, u_2)$. Since $N_{\mathcal{E}}(b_i) \cap P_R(u_1, u_2) = \emptyset$, $N_{\mathcal{E}}(b_i)$ must be contained entirely in one partition, namely that associated with $v(b_i)$. We observe that $v(b_i)$ is a cutset separating $N_{\mathcal{E}}(a_i)$ and $N_{\mathcal{E}}(b_i)$ and not contained in either set. This contradicts Lemma 1, which states that since $(a_i, b_i) \in E(Q)$, $N_{\mathcal{E}}(a_i)$ and $N_{\mathcal{E}}(b_i)$ are touching.

By a similar argument we can show that if a and b are bad neighbors in Q, then $v(a) = v(b)$. Since there is a path from b_1 to b_2, $v(b_1) = v(b_2)$. Since for $i = 1, 2$, $v(b_i) \in N_{\mathcal{E}}(a_i) \cap P_R(u_1, u_2)$, then $v(b_1) \in N_{\mathcal{E}}(a_1) \cap N_{\mathcal{E}}(a_2) \cap P_R(u_1, u_2)$, which we proved to be empty. □

We finish this section with the following observation, which will prove useful when justifying the recurrence used by our algorithm.

Lemma 8 *For any trees Q and R where $|E(Q)|, |E(R)| \geq 1$ and for any $a \in V(Q)$ and $u \in V(R)$, the smallest expansion of Q and R that contains (a, u) as an edge has size smaller than $|E(Q)| + |E(R)|$.*

Proof. As $|E(Q)|, |E(R)| \geq 1$, there exist edges (a, b) and (u, v) with a and u as endpoints. Let Q_1 and Q_2 (R_1 and R_2) be the connected components of the graph formed by removing the edge (a, b) from Q (the graph formed by removing the edge (u, v) from R) that contain a and b (u and v) respectively. It is easy to verify that $\mathcal{E} = (V(Q) \cup V(R), E)$ where

$$\begin{aligned} E &= \{(c, u) \mid c \in V(Q_1)\} \cup \{(c, v) \mid c \in V(Q_2)\} \cup \\ &\quad \{(a, w) \mid w \in V(R_1)\} \cup \{(b, w) \mid w \in V(R_2)\} \end{aligned}$$

is an expansion of Q and R containing (a, u). Since $|E| = |E(Q)| + |E(R)|$, $|E(T_{\mathcal{E}})| = |E(Q)| + |E(R)| - 1$. □

4. Smallest common tree major algorithm

4.1. Algorithm overview

For algorithmic convenience, we construct a rooted tree major, where any vertex of either input tree could be associated with the root. We fix a root for one tree and then try all possible rootings of the other tree; the following description concerns one possible choice of a root.

Our algorithm proceeds by dynamic programming, at each stage building tree majors of various subtrees of the inputs. After topologically sorting each tree with respect to the chosen root, we process each vertex a in $V(Q)$ in order from leaves to root, pairing a with each u in $V(R)$ in order from leaves to root.

For a given pair (a, u) we wish to determine the size of the largest common tree major T such that Q_a is a minor of T and R_u is a minor of T where for r the root of T, $f(r) = a$ and $g(r) = u$ (recall the definitions of Q_a and R_u in the first paragraph of Section 2). We solve this problem using subproblems involving children of a and u, where in each subproblem we specify not only the roots of the subtrees of Q and R, but also the subsets of the children included thus far in the mapping.

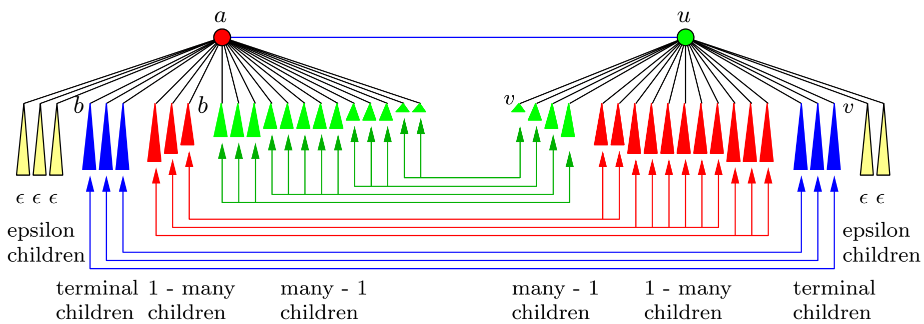

Fig. 3. The different ways possible smaller expansions involving subtrees rooted at the children of a and u can be combined in a general expansion.

Expansions, as defined in the previous section, give a convenient framework for expressing the progress of the algorithm, where expansions involving subgraphs of Q and R are augmented to form expansions of larger subgraphs of Q and R. The dynamic programming formulation of the problem relies on a set of subproblems at $a \in V(Q)$ and $u \in V(R)$, where each subproblem corresponds to one choice of how the children of a and the children of u are related, assuming that (a, u) is to be an edge in the expansion and that all subproblems rooted at children have already been solved.

4.2. Technical lemmas

When processing (a, u), we are assuming that $(a, u) \in E(\mathcal{E})$ and attempting to see where subsets of the children of a and u can map. To build our intuition, we consider the process from the point of view of Q (viewing from R is symmetric and hence the reasoning is identical). Each child b of a must eventually be involved in $\mathcal{E}$. There are four different cases for a child b of a, reflecting four different possible smaller expansions involving subtrees rooted at the children of a and u (for an illustration of the case analysis that follows, see Figure 3). The definitions of weak edges and $\mathcal{E}$-counterparts follow the definitions of expansion in Section 3.

1. (epsilon child) The subtree rooted at b is not involved in any previous expansion. It will be included by creating an edge in $\mathcal{E}$ from each vertex in the subtree to u.

2. (terminal child) The subtree rooted at b has been mapped to a subtree rooted at a child v of u, where (a, v) is not an edge in any previous expansion. In this case the edges (a, b) and (u, v) will be strong edges that are $\mathcal{E}$-counterparts.

3. (one-many child) The subtree rooted at b is mapped to subtrees rooted at a set of children of u, where (b, u) is an edge in a previous expansion. In this case (a, b) is a weak edge.

4. (many-one child) A set of subtrees rooted at children of a is mapped to a

subtree rooted at a child v of u, where (a, v) is an edge in a previous expansion. In this case (u, v) is a weak edge.

We formalize the possible associations of children by a tuple for each possible pair of subsets A of children of a and X of children of u and each possible mapping among vertices.

Definition 3 *Given two sets A, X we define $\Pi(A, X)$ as the set containing all tuples*

$$(\{A^{\mathrm{e}}, A^{\mathrm{t}}, A^{\mathrm{o}}, A^{\mathrm{m}}\}, \{X^{\mathrm{e}}, X^{\mathrm{t}}, X^{\mathrm{o}}, X^{\mathrm{m}}\}, \tau, \alpha, \chi)$$

that satisfy the following properties:

1. *$\{A^{\mathrm{e}}, A^{\mathrm{t}}, A^{\mathrm{o}}, A^{\mathrm{m}}\}$ is a partition of A (some parts may be empty);*
2. *$\{X^{\mathrm{e}}, X^{\mathrm{t}}, X^{\mathrm{o}}, X^{\mathrm{m}}\}$ is a partition of X (some parts may be empty);*
3. *$\tau : A^{\mathrm{t}} \to X^{\mathrm{t}}$ is a bijection;*
4. *$\alpha : X^{\mathrm{m}} \to A^{\mathrm{o}}$ is a surjection; and*
5. *$\chi : A^{\mathrm{m}} \to X^{\mathrm{o}}$ is a surjection.*

It is not difficult to show that in the trees Q rooted at a and R rooted at u, $f_{\mathcal{E}}$ preserves the parent-child orientation of the strong edges. Suppose instead that (a_1, a_2) and (u_1, u_2) are $\mathcal{E}$-counterparts where a_1 is in the path between a and a_2 in Q but, in R, u_2 is in the path connecting u and u_1. Applying Lemma 7 for paths $P_Q(a, a_2)$ and $P_R(u, u_2)$, we conclude $a_1 \in P_Q(a, a_2)$ is adjacent, in $\mathcal{E}$, to a vertex in $P_R(u, u_2)$. But a_1 is also adjacent to u_1 (in $\mathcal{E}$), and hence by property 1 of the definition of an expansion a_1 must be adjacent to u_2, a contradiction as (a_1, a_2) and (u_1, u_2) are strong edges. Using this observation we will always assume from now on that if (a_1, a_2) is strong and (u_1, u_2) is its $\mathcal{E}$-counterpart, then a_1 (u_1) is the endpoint closer to a (u) in Q (R). In general, whenever we mention an edge, the first endpoint of the pair will be the one that is closer to the root of the tree to which it belongs.

We call two edges of a rooted tree *comparable* if one of them is in the path connecting the other with the root. If we have three mutually incomparable edges such that exactly two of them have a vertex different from the root as a common ancestor, we call the two edges the *close pair* of the triple. The definitions of Q_a and R_u below are found in the first paragraph of Section 2.

Lemma 9 *For any expansion $\mathcal{E}$ of two trees $Q = Q_a$ and $R = R_u$ such that $(a, u) \in \mathcal{E}$, if e_1, e_2 are strong edges of Q and e_1' and e_2' are their $\mathcal{E}$-counterparts in R, then e_1 is comparable with e_2 if and only if e_1' is comparable with e_2'.*

Proof. We prove the lemma by contradiction. Without loss of generality, $e_1 = (a_1, a_2)$ is in the path connecting $e_2 = (a_3, a_4)$ and a, and $e_1' = (u_1, u_2)$ and $e_2' = (u_3, u_4)$ are not comparable. The incomparability of e_1' and e_2' means that u_2 is not in $P_R(u_1, u_4)$. By applying Lemma 7 for paths $P_Q(a_1, a_4)$ and $P_R(u_1, u_4)$, $a_2 \in P_Q(a_1, a_4)$ will be adjacent, in $\mathcal{E}$, to some vertex in $P_R(u_1, u_4)$. As a_2 is also adjacent to u_2 in $\mathcal{E}$, property 1 of the definition of an expansion requires that a_2 be adjacent to u_1. Since e_1 and e'_1 are strong edges we have obtained a contradiction. The proof of the other direction is symmetric. □

Lemma 10 *For any expansion $\mathcal{E}$ of two trees $Q = Q_a$ and $R = R_u$ such that $(a,u) \in \mathcal{E}$, if e_1, e_2 and e_3 are strong mutually incomparable edges of Q and e'_1, e'_2 and e'_3 are their $\mathcal{E}$-counterparts in R, then e_1, e_2 is the close pair of e_1, e_2, e_3 if and only if e'_1, e'_2 is the close pair of e'_1, e'_2, e'_3.*

Proof. In a proof by contradiction, we let $e_i = (a_i, b_i), i = 1,2,3,\ e'_i = (u_i, v_i), i = 1,2,3$, and suppose that e_1 and e_2 form a close pair and e'_2 and e'_3 form a close pair. Let b be the common ancestor of e_1 and e_2 and v be the common ancestor of e'_2 and e'_3. As a consequence of Lemma 7 applied to $P_Q(b_1, b_2)$ and $P_R(v_1, v_2)$, b must be adjacent in $\mathcal{E}$ to some vertex in $P_R(v_1, v_2)$. Similarly, we can prove that b must be adjacent in $\mathcal{E}$ to some vertex in $P_R(v_1, v_3)$ and to some vertex in $P_R(v_2, v_3)$. It is not hard to see that these three facts and property 1 of the definition of $\mathcal{E}$ prove that b and v must be adjacent.

Using the same technique, by applying Lemma 7 to $P_Q(b_2, b_3)$ and $P_R(v_2, v_3)$, we conclude that $a \in P_Q(b_2, b_3)$ will be adjacent, in $\mathcal{E}$, to some vertex in $P_R(v_2, v_3)$. As a is also adjacent to u, by property 1 of the definition of an expansion, a must be adjacent to v. By Lemma 7 for $P_Q(b_1, b_2)$ and $P_R(v_1, v_2)$, by symmetry we can show that b is connected to u in $\mathcal{E}$. We have shown that $\{(a,u),(a,v),(b,u),(b,v)\} \subseteq E(\mathcal{E})$ which violates property 3 of the definition of an expansion. The proof of the other direction is symmetric. □

Given a child b of a in Q_a we denote as $\tilde{Q}_b$ the graph Q_b augmented with the edge (a,b), and given a child v of u in R_u we denote as $\tilde{R}_v$ the graph R_v augmented with the edge (u,v).

Lemma 11 *For any expansion $\mathcal{E}$ of two trees $Q = Q_a$ and $R = R_u$ such that $(a,u) \in \mathcal{E}$, there exists a tuple $(\{A^{\mathrm{e}}, A^{\mathrm{t}}, A^{\mathrm{o}}, A^{\mathrm{m}}\},\ \{X^{\mathrm{e}}, X^{\mathrm{t}}, X^{\mathrm{o}}, X^{\mathrm{m}}\},\ \tau, \alpha, \chi)$ in $\Pi(\mathsf{children}(a), \mathsf{children}(u))$, such that the following hold:*

1. *there are no strong edges in $Q_{A^{\mathrm{e}}}$ or $R_{X^{\mathrm{e}}}$;*
2. *all edges from a to vertices in A^t and from u to vertices in X^t are strong;*
3. *all edges from a to vertices in A^o and from u to vertices in X^o are weak;*
4. *for all $b \in A^t$, $v \in X^m$, and $c \in A^m$, $f_{\mathcal{E}}$ maps (a,b) to $(u, \tau(b))$, the strong edges in Q_b to the strong edges in $R_{\tau(b)}$, the strong edges in R_v to the strong edges in $Q_{\alpha(v)}$, and the strong edges in Q_c to the strong edges in $R_{\chi(c)}$.*

Proof. We let E_a be the set of edges induced in Q by a and its children and let E_u be the set of edges induced in R by u and its children. To construct the desired partition, we first define sets A^{e}, A^{t}, X^{e}, and X^{t} as follows: A^{e} is the maximum subset of the children of a in Q with the property that for each b in A^e, Q_b contains no strong edges; A^{t} consists of the the children b of a such that (a,b) is the $\mathcal{E}$-counterpart of an edge (u,v) in E_u; X^e and X^t are defined analogously. We form the bijection τ by setting $\tau(b) = v$ for (a,b) and (u,v) $\mathcal{E}$-counterparts, for

$b \in A^t$. We have now satisfied conditions 1 and 2, and it is straightforward to see that, for all $b \in A^{\mathrm{t}}$, $f_{\mathcal{E}}$ maps (a, b) to $(u, \tau(b))$.

We now claim that for any $b \in A^{\mathrm{t}}$, $f_{\mathcal{E}}$ maps the strong edges of Q_b to the strong edges of $R_{\tau(b)}$. Suppose instead that edges $(a, b) \in E_a$, $(c, d) \in E(Q_b)$, $(u, v) = f_{\mathcal{E}}((a, b))$, and $(w, x) = f_{\mathcal{E}}((c, d)) \notin E(R_{\tau(b)})$ were a counterexample. Then, since (a, u), $(d, x) \in E(\mathcal{E})$, we can apply Lemma 7 to $P_Q(a, d)$ and $P_R(u, x)$ in order to conclude that the neighborhood of b in $\mathcal{E}$ contains a vertex in $P_R(u, x)$. As (w, x) is not an edge of $R_{\tau(b)}$, $v = \tau(b)$ is not a vertex of this path. Since v is adjacent to b in $\mathcal{E}$, by property 1 of the definition of an expansion, u must be a neighbor of b in $\mathcal{E}$. This results in a contradiction, as (a, b) and (u, v) are strong edges of Q and R respectively.

We now define A^{o} to include any child b of a for which $\tilde{Q}_b$ contains strong edges whose $\mathcal{E}$-counterparts are in more than one of the trees $\tilde{R}_w$ for children w of u. Notice that A^{o} and A^{t} are disjoint, as for any $b \in A^{\mathrm{t}}$, the counterparts of the strong edges of $\tilde{Q}_b$ are all in one tree $\tilde{R}_w$, namely $\tilde{R}_{\tau(b)}$. Furthermore, we define X^{m} so that, for any $b \in A^{\mathrm{o}}$, X^{m} contains the children $v_1, \ldots, v_r$ of u such that $\tilde{R}_{v_1}, \ldots, \tilde{R}_{v_r}$ are the trees that contain $\mathcal{E}$-counterparts of strong edges in $\tilde{Q}_b$. The surjection α maps any vertex v_i in X^{m} to the corresponding vertex b of A^{o}.

We claim that for any $b \in A^{\mathrm{o}}$ the edge (a, b) is weak. Suppose instead that (a, b) were strong; since $b \notin A^t$, its $\mathcal{E}$-counterpart (x, w) must be in $\tilde{Q}_v$ for some child v of u. Since $b \in A^o$, the $\mathcal{E}$-counterparts of the strong edges in $\tilde{Q}_b$ are in more than one tree in $\mathcal{R}_u$, and thus some tree $\tilde{R}_{v'}$ different from $\tilde{R}_v$ contains at least one $\mathcal{E}$-counterpart (y, z) of a strong edge (c, d) in $\tilde{Q}_b$. Clearly (a, b) and (c, d) are comparable, contradicting Lemma 9 as (x, w) and (y, z) are incomparable. Therefore, all the edges connecting a with vertices in A^{o} are weak. As a consequence, X^{m} and X^{t} are disjoint as for any $v_i \in X^{\mathrm{m}}$ the $\mathcal{E}$-counterparts of the strong edges of $\tilde{R}_{v_i}$ belong to trees $\tilde{Q}_b$ for children b of a such that (a, b) is weak.

To show that the $\mathcal{E}$-counterparts of the strong edges in the $\tilde{R}_{v_i}$'s are all in $\tilde{Q}_b$, for $X^{\mathrm{m}} = \{v_1, \ldots, v_r\}$, we suppose instead that some tree (without loss of generality $\tilde{R}_{v_1}$) containing a strong edge e'_1 with its $\mathcal{E}$-counterpart e_1 in $\tilde{Q}_{b'}$ for some child $b' \neq b$ of a. By definition, $\tilde{R}_{v_1}$ contains a strong edge e'_2 different from e'_1 that is the counterpart of a strong edge in $\tilde{Q}_b$. In addition, also by definition, $\tilde{Q}_b$ contains at least one strong edge e_3 different from e_2 whose $\mathcal{E}$-counterpart e'_3 is in a tree (without loss of generality $\tilde{R}_{v_2}$) different from $\tilde{R}_{v_1}$. By Lemma 9, e_2 and e_3 (e'_1 and e'_2) are incomparable as their $\mathcal{E}$-counterparts e'_2 and e'_3 (e_1 and e_2) are incomparable. Moreover, the close pair of the first triple is e_2 and e_3 and the close pair of the second triple is e'_2 and e'_1, violating Lemma 10. We can conclude that $\mathcal{E}$-counterparts of the strong edges in the $\tilde{R}_{v_i}$'s are all in $\tilde{Q}_b$. This completes the proofs of conditions 3 and 4 as far as sets X^{m} and A^{o} are concerned.

Working symmetrically, we can include in X^{o} all the children v of u such that the strong edges of $\tilde{R}_v$ have $\mathcal{E}$-counterparts in more than one tree in $\tilde{Q}_b$ for children b of a. As before, X^{o} and X^{t} are disjoint. Moreover, X^{o} and X^{m} are also disjoint as, according to the discussion above, for any $v_i \in X^{\mathrm{m}}$ the strong edges of $\tilde{R}_{v_i}$ are all in a single $\tilde{Q}_b$. Applying the same arguments as before, we can define the set

A^{m} and surjection $\chi : A^{\mathrm{m}} \to X^{\mathrm{o}}$ and verify that conditions 3 and 4 are satisfied for $X^{\mathrm{o}}, A^{\mathrm{m}}$, and χ.

The construction of the desired tuple is not yet complete. If b is a child of a that has not yet been classified as a member of A^{e}, A^{t}, A^{o}, or A^{m}, then the $\mathcal{E}$-counterparts of the strong edges of $\tilde{Q}_b$ are *all* in exactly one tree $\tilde{R}_v$ but (a, b) and (u, v) are not both strong edges. We can make a similar claim for unclassified children v of u. Therefore, there is a bijection σ between the unclassified children of a and the unclassified children of u that allows us to classify each one of them arbitrarily in A^{o} and X^{m} respectively or in A^{m} and X^{o} respectively. For each such arbitrary choice α or χ is augmented by σ on the new pair of elements. By repeating the same arguments one can prove that, after this enhancement, the sets defined still satisfy properties 3 and 4 while α and χ remain surjections. In conclusion, the tuple $(\{A^{\mathrm{e}}, A^{\mathrm{t}}, A^{\mathrm{o}}, A^{\mathrm{m}}\}, \{X^{\mathrm{e}}, X^{\mathrm{t}}, X^{\mathrm{o}}, X^{\mathrm{m}}\}, \tau, \alpha, \chi)$ satisfies properties 1-5 and the lemma holds. □

In order to define the recurrence for our dynamic programming algorithm, we need to be able to decompose a minimum size expansion of two trees into minimum size expansions of pairs of subtrees. The following two lemmas consider decompositions induced by removal of a strong edge or a weak edge, respectively. For notational convenience, we will use short forms for various subgraphs of Q_a and R_u. For b a child of a in Q_a, we define $Q_{\neg b}$ to be $Q_a \setminus Q_b$; similarly, for v a child of u in R_u, we define $R_{\neg v}$ to be $R_u \setminus R_v$. For any subset X of children of u in R_u, we define $R_{\neg X}$ to be $R[(V(R) \setminus V(R_X)) \cup \{u\}]$.

Lemma 12 *If $\mathcal{E}$ is a minimum size expansion of $Q = Q_a$ and $R = R_u$, (a, b) is a strong edge of Q, and $(u, v) = f_\epsilon((a, b))$ is its $\mathcal{E}$-counterpart, then $\mathcal{E}$ is the union of a minimum size expansion of $Q_{\neg b}$ and $R_{\neg v}$ containing (a, u) and a minimum size expansion of Q_b and R_v containing (b, v).*

Proof. We first claim that $\mathcal{E}_1 = \mathcal{E}[V(Q_b) \cup V(R_v)]$ is an expansion of Q_b and R_v containing (b, v). In order to prove this, it is enough to show that all the neighbors in $\mathcal{E}$ of all the vertices in $V(Q_b)$ $(V(R_v))$ are in $V(R_v)$ $(V(Q_b))$. Suppose to the contrary that $\mathcal{E}$ contains an edge (c, w) such that $c \in V(Q_b)$ and $w \notin V(R_v)$. If we now apply Lemma 7 for $P_Q(a, c)$ and $P_R(u, w)$, $b \in P_Q(a, c)$ must be adjacent to some vertex not in R_v. As b is adjacent to v in $\mathcal{E}$, by property 1 of the definition of an E, b must be adjacent to u, a contradiction as (a, b) and (u, v) are strong edges. By symmetry we can prove that $\mathcal{E}_2 = \mathcal{E}[V(Q_{\neg b}) \cup V(R_{\neg v})]$ is an expansion of $Q_{\neg b}$ and $R_{\neg v}$ containing (a, u).

It now remains to prove that $\mathcal{E}_1$ and $\mathcal{E}_2$ are both minimum size expansions. Suppose instead that there is an expansion $\mathcal{E}'$ of Q_b and R_v, that has size smaller than the one of $\mathcal{E}_1$. Then, by Lemma 6, $\mathcal{E}' \cup \mathcal{E}_2$ is an expansion of Q and R with size smaller than $\mathcal{E}$, contradicting the minimality of $\mathcal{E}$. The proof for $\mathcal{E}_2$ is similar. □

Lemma 13 *If $\mathcal{E}$ is a minimum size expansion of $Q = Q_a$ and $R = R_u$, (a, b) is a weak edge of Q, and (a, u) and (b, u) are edges of $\mathcal{E}$, then*

1. *there exists a subset X of the children of u such that $\mathcal{E}$ is the union of a*

minimum size expansion of Q_b and R_X containing (b, u) and a minimum size expansion of $Q_{\neg b}$ and $R_{\neg X}$ containing (a, u), and

2. *if Q_b contains only weak edges, then for any vertex $c \in V(Q_b)$, $N_{\mathcal{E}}(c) = \{u\}$.*

Proof. We let X be the set containing any child v of u for which R_v contains a neighbor, in $\mathcal{E}$, of a vertex in Q_b. We then let $\mathcal{E}_1 = \mathcal{E}[V(Q_b) \cup V(R_X)]$ and $\mathcal{E}_2 = \mathcal{E}[V(Q_{\neg b}) \cup V(R_{\neg X})]$. To show that $\mathcal{E}_1$ is an expansion of Q_b and R_X and that $\mathcal{E}_2$ is an expansion of $Q_{\neg b}$ and $R_{\neg X}$ by inheriting the properties of an expansion from $\mathcal{E}$, it will suffice to show that in $\mathcal{E}$ all neighbors of vertices in $V(Q_b)$ ($V(R_X)$,$V(Q_{\neg b})$, and $V(R_{\neg X})$, respectively) are in $V(R_X)$ ($V(Q_b)$,$V(R_{\neg X})$, and $V(Q_{\neg b})$, respectively). The first of the four statements follows from the definition of X.

To prove the third claim by contradiction, suppose instead that a vertex $c \in Q_{\neg b}$ is adjacent in $\mathcal{E}$ to a vertex w outside of $R_{\neg X}$. Clearly, $X \neq \emptyset$ and w is in one of the connected components of $R_X - \{u\}$. Let v be the vertex of X such that $w \in R_v$. In addition, by the definition of X, Q_b contains at least one vertex d adjacent, in $\mathcal{E}$, to a vertex x in R_v. We now apply Lemma 7 to paths $P_Q(d, c)$ and $P_R(w, x)$ to conclude that $a \in P_Q(d, c)$ is adjacent, in $\mathcal{E}$, to some vertex in R_v. Since a is also adjacent to u in $\mathcal{E}$, by property 1 of the definition of an expansion, a must be adjacent to v in $\mathcal{E}$. Similarly, we can show that b is adjacent to v. Therefore, the neighborhoods of a and b have two vertices, i.e. u and v, in common. This contradicts property 3 of the definition of an expansion and hence the claim holds. The remaining claims can be proved in a similar manner.

To prove the first statement in the lemma, it now remains to show that $\mathcal{E}_1$ and $\mathcal{E}_2$ are both minimum size expansions. Suppose to the contrary that there is an expansion $\mathcal{E}'$ of one of the pairs Q_b and R_X or $Q_{\neg b}$ and $R_{\neg X}$ that has size smaller than the one established above, say Q_b and R_X have an expansion $\mathcal{E}'$ smaller than $\mathcal{E}_1$. Then, by Lemma 5, $\mathcal{E}' \cup \mathcal{E}_2$ is an expansion of Q and R with size smaller than $\mathcal{E}$, contradicting the minimality of $\mathcal{E}$.

To prove the second statement in the lemma, it suffices to show that if Q_b contains only weak edges, then $X = \emptyset$. Suppose instead that $|X| \geq 1$. Then $|E(R_X)| \geq 1$. Since $\mathcal{E}_2$ is a minimum size expansion of Q_b and R_X, and Q_b contains only weak edges, R_X contains only weak edges. But then $|E(T_{\mathcal{E}_2})| = |E(Q_b)| + |E(R_X)|$, contradicting Lemma 8. □

The following lemma uses the structural information of Lemma 11, followed by repeated applications of Lemmas 12 and 13.

Lemma 14 *For any minimum size expansion $\mathcal{E}$ of two trees $Q = Q_a$ and $R = R_u$ such that $(a, u) \in \mathcal{E}$, there is a tuple $(\{A^{\mathrm{e}}, A^{\mathrm{t}}, A^{\mathrm{o}}, A^{\mathrm{m}}\}, \{X^{\mathrm{e}}, X^{\mathrm{t}}, X^{\mathrm{o}}, X^{\mathrm{m}}\}, \tau, \alpha, \chi)$ in $\Pi(\mathsf{children}(a), \mathsf{children}(u))$ such that $(\mathcal{E}_e \cup \mathcal{E}_t \cup \mathcal{E}_a \cup \mathcal{E}_u)$ is a partition of $\mathcal{E}$ where*

1. *$\mathcal{E}_e$ relates epsilon children; $\mathcal{E}_e = \{(a, z) \mid z \in V(R_{X^{\mathrm{e}}})\} \cup \{(c, u) \mid c \in V(Q_{A^{\mathrm{e}}})\}$.*

2. *$\mathcal{E}_t$ relates terminal children; $\mathcal{E}_t = \bigcup_{b \in A^{\mathrm{t}}} \mathcal{E}_{t,b}$ where, for any vertex $b \in A^{\mathrm{t}}$, $\mathcal{E}_{t,b}$ is a minimum expansion of Q_b and $R_{\tau(b)}$ that contains $(b, \tau(b))$.*

3. *$\mathcal{E}_a$ relates one-many children; $\mathcal{E}_a = \bigcup_{b \in A^{\mathrm{o}}} \mathcal{E}_{a,b}$ where, for any vertex $b \in A^{\mathrm{o}}$, $\mathcal{E}_{a,b}$ is a minimum expansion of Q_b and $R_{\alpha^{-1}(b)}$ that contains (b, u).*

4. $\mathcal{E}_u$ *relates many-one children;* $\mathcal{E}_u = \bigcup_{v \in X^{\mathrm{o}}} \mathcal{E}_{u,v}$ *where, for any vertex* $v \in X^{\mathrm{o}}$*,* $\mathcal{E}_{u,v}$ *is a minimum expansion of* $Q_{\chi^{-1}(v)}$ *and* R_v *that contains* (a, v)*.*

Proof. We will prove the lemma by decomposing $\mathcal{E}$ in groups of subexpansions of the four types described in Lemma 11. This decomposition will proceed step by step by applying Lemmas 12 and 13, inductively, as appropriate, depending on the type of subexpansion it is possible to extract.

Let $(\{A^{\mathrm{e}}, A^{\mathrm{t}}, A^{\mathrm{o}}, A^{\mathrm{m}}\}, \{X^{\mathrm{e}}, X^{\mathrm{t}}, X^{\mathrm{o}}, X^{\mathrm{m}}\}, \tau, \alpha, \chi) \in \Pi(\mathsf{children}(a), \mathsf{children}(u))$ be as determined by Lemma 11. We will extract the decomposition of $\mathcal{E}$ using induction on $j = |A^{\mathrm{e}}| + |X^{\mathrm{e}}| + |A^{\mathrm{t}}| + |A^{\mathrm{o}}| + |X^{\mathrm{o}}|$. If $j = 0$, the result is trivial. We assume that it holds if $0 \leq j < k$ and we will prove that it also holds when $j = k$. Let $b \in A^{\mathrm{e}} \cup X^{\mathrm{e}} \cup A^{\mathrm{t}} \cup A^{\mathrm{o}} \cup X^{\mathrm{o}}$. We may assume that b is a vertex in $A^{\mathrm{e}} \cup A^{\mathrm{t}} \cup A^{\mathrm{o}}$, as the case where b is a vertex in $X^{\mathrm{e}} \cup X^{\mathrm{o}}$ is symmetric. We set $\mathcal{E}_{\neg b} = \mathcal{E}[(V(Q \setminus Q_b) \cup V(R \setminus R_{\sigma(b)})]$ where, if $b \in A^{\mathrm{e}}$, (resp. $b \in A^{\mathrm{t}}$, $b \in A^{\mathrm{o}}$), then $\sigma(b) = \emptyset$, (resp. $\sigma(b) = \tau(b)$, $\sigma(b) = \alpha^{-1}(b)$).

We claim that $\mathcal{E}_{\neg b}$ is a minimum size expansion of $Q_{\neg b}$ and $R_{\neg\sigma(b)}$ and that $\mathcal{E}_b = \mathcal{E}[V(Q_b) \cup V(R_{\sigma(b)})]$ is a minimum size expansion of Q_b and $R_{\sigma(b)}$. When $b \in A^{\mathrm{t}}$, the claim is a consequence of Lemma 12 and when $b \in A^{\mathrm{e}} \cup A^{\mathrm{o}}$, the claim is a consequence of Lemma 13.

We now apply the induction hypothesis on $\mathcal{E}_{\neg b}$ and derive the tuple $(\{A^{\mathrm{e}}_{\neg b}, A^{\mathrm{t}}_{\neg b}, A^{\mathrm{o}}_{\neg b}, A^{\mathrm{m}}_{\neg b}\}, \{X^{\mathrm{e}}_{\neg b}, X^{\mathrm{t}}_{\neg b}, X^{\mathrm{o}}_{\neg b}, X^{\mathrm{m}}_{\neg b}\}, \tau_{\neg b}, \alpha_{\neg b}, \chi_{\neg b}) \in P(\mathsf{children}(a) \setminus \{b\}, \mathsf{children}(u) \setminus \{\sigma(b)\})$ and the corresponding partition $(\mathcal{E}_{e,\neg b}, \mathcal{E}_{t,\neg b}, \mathcal{E}_{a,\neg b}, \mathcal{E}_{u,\neg b})$ of $\mathcal{E}_{\neg b}$ satisfying conditions 1–4. If $b \in A^{\mathrm{t}}$, and v is as defined in Lemma 12, for each member m of the tuple, $m = m_{\neg b}$ with the following exceptions: $A^{\mathrm{t}} = A^{\mathrm{t}}_{\neg b} \cup \{b\}$, $X^{\mathrm{t}} = X^{\mathrm{t}}_{\neg b} \cup \{v\}$, and $\tau = \tau_{\neg b} \cup \{(b, v)\}$. Suppose now that $b \in A^{\mathrm{e}} \cup A^{\mathrm{o}}$ and X is as defined in Lemma 13. In this case, it is easy to see that if $X = \emptyset$, then $A^{\mathrm{e}} = A^{\mathrm{e}}_{\neg b} \cup \{b\}$, and for each other member m of the tuple $m = m_{\neg b}$. Finally, if $X \neq \emptyset$, $A^{\mathrm{o}} = A^{\mathrm{o}}_{\neg b} \cup \{b\}$, $X^{\mathrm{m}} = X^{\mathrm{m}}_{\neg b} \cup X$, and $\alpha = \alpha_{\neg b} \cup \{(w, b) \mid w \in X\}$, with all other members of the tuple unchanged. We construct the partition $(\mathcal{E}_e \cup \mathcal{E}_t \cup \mathcal{E}_a \cup \mathcal{E}_u)$ of $\mathcal{E}$ as follows.

If $b \in A^{\mathrm{e}}$, then, by Lemma 13, $\mathcal{E}_b = \{(c, u) \mid c \in V(Q_b)\}$. We set $\mathcal{E}_e = \mathcal{E}_b \cup \mathcal{E}_{e,\neg b}$, $\mathcal{E}_t = \mathcal{E}_{t,\neg b}$, $\mathcal{E}_a = \mathcal{E}_{a,\neg b}$, and $\mathcal{E}_u = \mathcal{E}_{u,\neg b}$.

If $b \in A^{\mathrm{o}}$, then, by Lemma 13, $\mathcal{E}_b$ is a minimum expansion of Q_b and $R_{\alpha^{-1}(b)} = R_{\sigma(b)}$. We set $\mathcal{E}_e = \mathcal{E}_{e,\neg b}$, $\mathcal{E}_t = \mathcal{E}_{t,\neg b}$, $\mathcal{E}_a = \mathcal{E}_b \cup \mathcal{E}_{a,\neg b}$, and $\mathcal{E}_u = \mathcal{E}_{u,\neg b}$.

If $b \in A^{\mathrm{t}}$, then, by Lemma 12, $\mathcal{E}_b$ is a minimum expansion of Q_b and $R_{\tau(b)} = R_{\sigma(b)}$. We set $\mathcal{E}_e = \mathcal{E}_{e,\neg b}$, $\mathcal{E}_t = \mathcal{E}_b \cup \mathcal{E}_{t,\neg b}$, $\mathcal{E}_a = \mathcal{E}_{a,\neg b}$, and $\mathcal{E}_u = \mathcal{E}_{u,\neg b}$.

It now remains to verify that, in all cases, the tuple

$$(\{A^{\mathrm{e}}, A^{\mathrm{t}}, A^{\mathrm{o}}, A^{\mathrm{m}}\}, \{X^{\mathrm{e}}, X^{\mathrm{t}}, X^{\mathrm{o}}, X^{\mathrm{m}}\}, \tau, \alpha, \chi)$$

along with the partition $(\mathcal{E}_e \cup \mathcal{E}_t \cup \mathcal{E}_a \cup \mathcal{E}_u)$ satisfy conditions 1–4. This check is straightforward except for conditions 2 and 3 where we have to prove that the new expansions $\mathcal{E}_t$ and $\mathcal{E}_a$ are minimum expansions. This follows from Lemmas 5 and 6 as, by their construction, $\mathcal{E}_t$ and $\mathcal{E}_a$ are unions of minimum expansions. □

Procedure Expansion(Q, R, a, u)
Input: Two trees Q, R and two vertices $a \in V(Q), u \in V(R)$.
Output: $\min\{|\mathcal{E}| \ : \ \mathcal{E}$ is an expansion of Q and R and $(a, u) \in \mathcal{E}\}$.

1: Root Q and R at a and u respectively.
2: Topologically sort $V(Q)$, giving $L_Q := \{a_1, \ldots, a_{|V(Q)|}\}$ where $a = a_{|V(Q)|}$.
3: Topologically sort $V(R)$, giving $L_R := \{u_1, \ldots, u_{|V(R)|}\}$ where $u = u_{|V(R)|}$.
4: **for** $i := 1 \ldots |V(Q)|$ **do**
5:   **for** $j := 1 \ldots |V(R)|$ **do**
6:     **if** a_i and u_j are leaves **then** $I(a_i, u_j, \emptyset, \emptyset) := 1$
7:     **else**
8:       **for all** $X \subseteq \mathsf{children}(u_j)$ and $A \subseteq \mathsf{children}(a_i)$ **do**
9:         $x := |V(Q)| + |V(R)|$
10:         **for all** $(\{A^{\mathrm{e}}, A^{\mathrm{t}}, A^{\mathrm{o}}, A^{\mathrm{m}}\}, \{X^{\mathrm{e}}, X^{\mathrm{t}}, X^{\mathrm{o}}, X^{\mathrm{m}}\}, \tau, \alpha, \chi) \in \Pi(A, X)$ **do**
11:           $x := \min\{\, x,\ |V(Q_{A^{\mathrm{e}}})| + |V(R_{X^{\mathrm{e}}})| - 1 +$ (i)
$\sum_{b \in A^{\mathrm{t}}} I(b, f_{\mathtt{t}}(b), \mathsf{children}(b), \mathsf{children}(\tau(b))) +$ (ii)
$\sum_{b \in A^{\mathrm{o}}} I(b, u_i, \mathsf{children}(b), \alpha^{-1}(b)) +$ (iii)
$\sum_{v \in X^{\mathrm{o}}} I(a_i, v, \chi^{-1}(v), \mathsf{children}(v)) \,\}$ (iv)
12:         $I(a_i, u_i, A, X) := x$
13: **return** $I(a, u, \mathsf{children}(a), \mathsf{children}(u))$
14: **end**

Theorem 2 *For any trees Q and R rooted at a and u respectively,* Expansion (Q, R, a, u) *returns the minimum number of edges in any expansion $\mathcal{E}$ containing (a, u).*

Proof. We prove that for Q and R rooted at a and u respectively, for any $c \in V(Q)$, $z \in V(R)$, and any $A \subseteq \mathsf{children}(c)$ and $X \subseteq \mathsf{children}(z)$, the quantity $I(c, z, A, X)$ computed by the algorithm is the minimum number of edges over all expansions $\mathcal{E}$ of Q_A and R_X, where $(c, z) \in \mathcal{E}$. The proof is by induction on the order of computation.

Consider the computation of $I(c, z, A, X)$. As L_Q and L_R are topological sorts of $V(Q)$ and $V(R)$ respectively, we can conclude that $I(d, y, A_d, X_y)$ has already been computed in the following three cases, which cover the expressions on the right-hand side of step 11.

1. $d \in \mathsf{children}(c)$, $y \in \mathsf{children}(z)$, $A_d \subseteq \mathsf{children}(d)$, and $X_y \subseteq \mathsf{children}(y)$.

2. $d \in \mathsf{children}(c)$, $y = z$, $A_d \subseteq \mathsf{children}(d)$, and $X_y \subseteq \mathsf{children}(z)$.

3. $d = c$, $y \in \mathsf{children}(z)$, $A_d \subseteq \mathsf{children}(c)$, and $X_y \subseteq \mathsf{children}(y)$.

If we assume by the inductive hypothesis that the values $I(d,\, y,\, A_d,\, X_y)$ are correct, then, by Lemma 14, there is a choice of $(\{A^{\mathrm{e}}, A^{\mathrm{t}}, A^{\mathrm{o}}, A^{\mathrm{m}}\},\, \{X^{\mathrm{e}}, X^{\mathrm{t}},\, X^{\mathrm{o}}, X^{\mathrm{m}}\},$ $\tau, \alpha, \chi)$ that results, at step 11, in x taking on the minimum number of edges in an expansion $\mathcal{E}$ of Q_A and R_X containing (c, z), as required. □

Theorem 3 *For any pair of trees Q and R of bounded degree,* $\mathsf{sctmj}(Q, R)$ *can be computed in $O(n^3)$ time where $n = \max\{|V(Q)|, |V(R)|\}$.*

Proof. The `if`-statement at step 6 is invoked $O(n^2)$ times, and because the maximum degrees of Q and R are bounded by a constant, the loops at steps 8 and 10 result in a constant number of iterations of step 11. This quantity is multiplied by $O(n)$, the number of rootings to check. □

5. Extensions of the algorithm

In this section we describe how our algorithm can be generalized to the problem of determining the edit distance (under certain conditions) of a pair of a edge-labeled, unrooted, unordered trees. The set of operations used is edge contraction, edge relabeling, and edge insertion. In this last operation, a vertex v is chosen, replaced by a pair of vertices v_1 and v_2 such that $N(v_1)$ and $N(v_2)$ partition $N(v)$, and a labeled edge (v_1, v_2) is inserted. The final condition we impose on the edit sequence is that all insertions must be completed before any other operations.

Let Q and R be edge-labeled trees, i.e. for some given alphabet Σ, there exist two functions $q : E(Q) \to \Sigma$ and $r : E(R) \to \Sigma$. We denote as g, h the cost functions where for any $\sigma \in \Sigma$, $g(\sigma)$ and $h(\sigma)$ represent the cost of the contraction and the insertion respectively of an edge labeled with σ. Finally, for $\sigma, \rho \in \Sigma$, we denote as $l(\sigma, \rho)$ the cost of changing the label of an edge from σ to ρ. We can now define as $\mathsf{dist}(Q, R)$ the smallest possible total cost of a sequence of operations which transforms Q to R, subject to the constraint that all insertions occur first.

Given such a sequence, we can reorder it (without altering its cost) so that the relabelings precede the contractions and follow the expansions. Let T_1 be the tree after all expansions, and T_2 the tree after all relabelings. Clearly, if labels are removed, T_1 is isomorphic to T_2, and both are majors of both Q and R. Thus, for every edit sequence, there is a natural common supertree.

Conversely, let T be a common major of Q and R corresponding to some extension $\mathcal{E}$ of Q and R. It is easy to see that Q can be transformed to R after the following sequence of operations: first insert in Q all the edges in $E(T) - E(Q)$, then relabel all the strong edges of T to the labelings they should have in R, and, finally, contract all the edges in $E(T) - E(R)$. Notice that, if $S(T)$ contains the strong edges of T, the total cost of this sequence of operations is

$$\sum_{e \in E(T)-E(Q)} h(q(e)) + \sum_{e \in E(T)-E(R)} g(r(e)) + \sum_{e \in S(T)} l(q(e), r(e))$$

which, in turn, is equal to

$$\sum_{e \in E(T)} h(q(e)) + \sum_{e \in E(T)} g(r(e)) + \sum_{e \in S(T)} l(q(e), r(e)) - C(R, Q) =$$

$$\sum_{e \in E(T)} (h(q(e)) + g(r(e))) + \sum_{e \in S(T)} l(q(e), r(e)) - C(R, Q)$$

where $C(R,Q) = \sum_{e\in E(Q)} h(q(e)) + \sum_{e\in E(R)} g(r(e))$. Therefore, in order to compute $\mathsf{dist}(Q,R)$ we have to find an expansion $\mathcal{E}$ with major T where the quantity

$$Q(T) = \sum_{e\in E(T)} (h(q(e)) + g(r(e))) + \sum_{e\in S(T)} l(q(e), r(e))$$

is minimized. Following the methodology of the previous sections we set up a general version of $I(c,z,A,X)$, representing the minimum value of $Q(T)$ over the T's corresponding to all expansions $\mathcal{E}$ of Q_A and Q_X where $(c,z) \in \mathcal{E}$. The only modification required for Procedure $\mathsf{Expansion}(Q,R,a,u)$ concerns the way x is computed in line 11, which should change to the following:

$$
\begin{array}{lll}
11:\ x := \min\{\, x, & \sum_{e\in E(Q_{A^{\mathsf{e}}})} h(q(e)) + \sum_{e\in E(R_{X^{\mathsf{e}}})} g(r(e)) + & \text{(i)}\\
& \sum_{b\in A^{\mathsf{t}}}(\ I(b, f_{\mathsf{t}}(b), \mathsf{children}(b), \mathsf{children}(\tau(b))) + & \\
& \qquad l(q((a_i,b)), r((u_i, f_{\mathsf{t}}(b))))\ ) + & \text{(ii)}\\
& \sum_{b\in A^{\mathsf{o}}}(\ I(b, u_i, \mathsf{children}(b), \alpha^{-1}(b)) + h(q((a_i,b)))\ ) + & \text{(iii)}\\
& \sum_{v\in X^{\mathsf{o}}}(\ I(a_i, v, \chi^{-1}(v), \mathsf{children}(v)) + g(r((u_j,v)))\ )\ \} & \text{(iv)}
\end{array}
$$

For completeness, in line 9, x should now be initialized as $C(R,Q)$.

Clearly, the above modifications do not require more time asymptotically, and we have the following:

Theorem 4 *The edit distance (under operations edge contraction, edge relabeling, and edge insertion, where all insertions come first) of any pair of edge-labeled trees Q and R of bounded degree, can be computed in $O(n^3)$ time where $n = \max\{|V(Q)|, |V(R)|\}$.*

6. Conclusions and further work

We have demonstrated an $O(n^3)$ algorithm for finding the smallest common tree major of two trees Q and R, where both Q and R are unrooted and undirected, and have degree bounded by a fixed constant. The degree restriction can be relaxed to maximum degree $O(\log n/\log\log n)$ while keeping the running time of the algorithm polynomial, since the multiplicative factor is $d^{O(d)}$ for trees of maximum degree d (this factor arises from the number of tuples examined at line 10 of the algorithm). Our algorithm can be generalized to the problem of determining the edit distance (under the operations of edge contraction, edge relabeling, and edge insertion, where all insertions come first) of a pair of a edge-labeled, unrooted, unordered trees, by incorporating labels into the definition of the expansion. All of our algorithms can be implemented in NC using the technique of Brent restructuring to parallelize dynamic programming on trees [9]. Our work is also related to work on intertwines [15]: the value $\mathsf{sctmj}(Q,R)$ is the minimum size of an acyclic intertwine of Q and R.

Although the NP-completeness of minor containment for general trees suggests the intractability of finding the largest common subgraph under minors, there is hope for solving other related problems. The problem of determining whether or not G is a minor of H is solvable in polynomial time for G and H both bounded-degree partial k-trees [16] or for G and H both k-connected k-paths [11]; solving the largest common supergraph problem for each of these graph classes would be an obvious extension to our work. Another obvious extension would be to solve the largest common tree major problem for three or more input trees.

7. Acknowledgements

We would like to thank an anonymous referee for comments which improved the presentation of this paper.

References

1. A. Amir and D. Keselman. Maximum agreement subtree in a set of evolutionary trees: metrics and efficient algorithms. *SIAM Journal on Computing*, 26(6):1656–1669, December 1997.
2. J. A. Bondy and U.S.R. Murty. *Graph Theory with Applications*. North-Holland, 1976.
3. M. J. Chung. O($n^{2.5}$) time algorithms for the subgraph homeomorphism problem on trees. *Journal of Algorithms*, 8:106–112, 1987.
4. R. Cole and R. Hariharan. An $O(n \log n)$ algorithm for the maximum agreement subtree problem for binary trees. In *Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 323–332, 1996.
5. M. Dubiner, Z. Galil, and E. Magen. Faster tree pattern matching. In *Proceedings of the 31st Annual Symposium on Foundations of Computer Science*, pages 145–150, 1990.
6. P. Duchet. Tree minors. Presentation at *AMS-IMS-SIAM Joint Summer Research Conference on Graph Minors*, 1991 (personal communication, A. Gupta).
7. M. Farach, T. Przytycka, and M. Thorup. On the agreement of many trees. *Information Processing Letters*, 55(6):297–301, 1995.
8. M. Farach and M. Thorup. Fast comparison of evolutionary trees. In *Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 481–488, 1994.
9. A. Gupta and N. Nishimura. The parallel complexity of tree embedding problems. *Journal of Algorithms*, 18(1):176–200, 1995.
10. A. Gupta and N. Nishimura. Finding largest subtrees and smallest supertrees. *Algorithmica*, 21:183–210, 1998.
11. A. Gupta, N. Nishimura, A. Proskurowski, and P. Ragde. Embeddings of k-connected graphs of pathwidth k. In *Proceedings of the 7th Scandinavian Workshop on Algorithm Theory*, 2000.
12. T. Jiang, L. Wang, and K. Zhang. Alignment of trees – an alternative to tree edit. In *Combinatorial Pattern Matching*, pages 75–86, 1994.
13. P. Kilpeläinen and H. Mannila. Ordered and unordered tree inclusion. *SIAM Journal on Computing*, 24(2):340–356, 1995.
14. S. R. Kosaraju. Efficient tree pattern matching. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, pages 178–183, 1989.
15. J. Lagergren. The size of an intertwine. In *Proceedings of the 23rd International Colloquium on Automata, Languages, and Programming*, volume 820 of *Lecture Notes in Computer Science*, pages 520–531, 1994.
16. J. Matoušek and R. Thomas. On the complexity of finding iso- and other morphisms

for partial k-trees. *Discrete Mathematics*, 108:343–364, 1992.

17. N. Robertson and P. Seymour. Graph minors II. Algorithmic aspects of tree-width. *Journal of Algorithms*, 7:309–322, 1986.
18. K. Siddiqi, A. Shokoufandeh, S. Dickinson, and S. Zucker. Shock graphs and shape matching. *International Journal of Computer Vision*, to appear.
19. T. Warnow. Tree compatibility and inferring evolutionary history. In *Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 382–391, 1993.