View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com
ScienceDirect
Theoretical Computer Science 362 (2006) 33–53
Theoretical Computer Science
www.elsevier.com/locate/tcs

An algebraic view of the relation between largest common subtrees and smallest common supertrees

Francesc Rosselló[a,*,1], Gabriel Valiente[b,2]

[a] *Department of Mathematics and Computer Science, Research Institute of Health Science (IUNICS), University of the Balearic Islands, E-07122 Palma de Mallorca, Spain*

[b] *Algorithms, Bioinformatics, Complexity and Formal Methods Research Group, Department of Software, Technical University of Catalonia, E-08034 Barcelona, Spain*

Received 28 July 2004; accepted 4 May 2006

Communicated by Z. Esik

Abstract

The relationship between two important problems in tree pattern matching, the largest common subtree and the smallest common supertree problems, is established by means of simple constructions, which allow one to obtain a largest common subtree of two trees from a smallest common supertree of them, and vice versa. These constructions are the same for isomorphic, homeomorphic, topological, and minor embeddings, they take only time linear in the size of the trees, and they turn out to have a clear algebraic meaning.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Tree pattern matching; Subtree isomorphism; Subtree homeomorphism; Topological embedding; Minor containment; Largest common subtree; Smallest common supertree

1. Introduction

Subtree isomorphism and the related largest common subtree and smallest common supertree problems have practical applications in combinatorial pattern matching [14,19,28], pattern recognition [7,10,25], computational molecular biology [2,20,30], chemical structure search [3,4,11], and other areas of engineering and life sciences. In these areas, they are some of the most widely used techniques for comparing tree-structured data.

Largest common subtree is the problem of finding a largest tree that can be embedded in two given trees, while *smallest common supertree* is the dual problem of finding a smallest tree into which two given trees can be embedded. A tree S can be embedded in another tree T when there exists an injective mapping f from the nodes of S to the nodes of T that transforms arcs into paths in some specific way. The type of embedding depends on the properties

* Corresponding author. Tel.: +34 971 173202; fax: +34 971 173003.
E-mail addresses: cesc.rossello@uib.es (F. Rosselló), valiente@lsi.upc.edu (G. Valiente).
[1] Partially supported by the Spanish DGES and the EU program FEDER, project ALBIOM (BFM2003-00771).
[2] Partially supported by Spanish CICYT projects MAVERISH (TIC2001-2476-C03-01) and GRAMMARS (TIN2004-07925-C03-01), and by the Ministry of Education, Science, Sports and Culture of Japan through Grant-in-Aid for Scientific Research B-15300003 for visiting JAIST (Japan Advanced Institute of Science and Technology).

0304-3975/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2006.05.031

of the mapping f. In this paper we consider the following four types of tree embeddings, defined by suitable extra conditions on f:

Isomorphic embedding: if there is an arc from a to b in S, then there is an arc from $f(a)$ to $f(b)$ in T.

Homeomorphic embedding: if there is an arc from a to b in S, then there is a path from $f(a)$ to $f(b)$ in T with all intermediate nodes of total degree 2 and no intermediate node belonging to the image of f.

Topological embedding: if there is an arc from a to b in S, then there is a path from $f(a)$ to $f(b)$ in T with no intermediate node belonging to the image of f; and if there are arcs from a to two distinct nodes b and c in S, then the paths from $f(a)$ to $f(b)$ and to $f(c)$ in T have no common node other than $f(a)$.

Minor embedding: if there is an arc from a to b in S, then there is a path from $f(a)$ to $f(b)$ in T with no intermediate node belonging to the image of f.

The different *subtree embedding problems* of deciding whether a given tree can be embedded into another given tree, for the different types of embedding defined above, have been thoroughly studied in the literature. Their complexity is already settled: they are polynomial-time solvable for isomorphic, homeomorphic, and topological embeddings, and NP-complete for minor embeddings [8,16–18]. Efficient algorithms are known for subtree isomorphism [21,26], for subtree homeomorphism [5,27,28], for largest common subtree under isomorphic embeddings [26] and homeomorphic embeddings [19], and for both largest common subtree and smallest common supertree under isomorphic and topological embeddings [12]. The only (exponential) algorithm known for largest common subtree under minor embeddings is given in [22].

Particular cases of these embedding problems for trees have also been thoroughly studied in the literature. On ordered trees, they become polynomial-time solvable for isomorphic, homeomorphic, topological, and also minor embeddings. In this particular case, the largest common subtree problem under homeomorphic embeddings is known as the maximum agreement subtree problem [1,6,24], the largest common subtree problem under minor embeddings is known as the tree edit problem [9,23,31], and the smallest common supertree problem under minor embeddings is known as the tree alignment problem [13,15,29]. The smallest common supertree problem under minor embeddings was also studied in [18] for trees of bounded degree.

In this paper, we establish in a unified way the relationship between the largest common subtree and the smallest common supertree problems for isomorphic, homeomorphic, topological, and minor embeddings. A similar correspondence between largest common subgraphs and smallest common supergraphs under isomorphic embeddings was studied in [10]. More specifically, we give a simple and unique construction that allows one to obtain in all four cases a largest common subtree of two trees from any smallest common supertree of them, and vice versa, another simple and unique construction that allows one to obtain in all four cases a smallest common supertree of two trees from any largest common subtree of them. These constructions take only time linear in the size of the trees, and, moreover, they have a clear algebraic meaning: in all four types of embeddings, a largest common subtree of two trees is obtained as the *pullback* of their embeddings into a smallest common supertree, and a smallest common supertree of two trees is obtained as the *pushout* of the embeddings of a largest common subtree into them. This is, to the best of our knowledge, the first unified construction showing the relation between largest common subtrees and smallest common supertrees for isomorphic, homeomorphic, topological, and minor embeddings. These results answer the open problem of establishing the relationship between the largest common subtree and the smallest common supertree under any embedding relation, posed by the last author in his talk "Subgraph Isomorphism and Related Problems for Restricted Graph Classes" at Dagstuhl Seminar 04221, "Robust and Approximative Algorithms on Particular Graph Classes," May 23–28, 2004.

Roughly speaking, our constructions work as follows. Given two trees T_1 and T_2 and a largest common subtree T_μ explicitly embedded into them, a smallest common supertree of T_1 and T_2 is obtained by first making the disjoint sum of T_1 and T_2, then merging in this sum each two nodes of T_1 and T_2 that are related to the same node of T_μ, and finally removing all parallel arcs and all arcs subsumed by paths. Conversely, given two trees T_1 and T_2 embedded into a smallest common supertree T of them, a largest common subtree of T_1 and T_2 is obtained by removing all nodes in T not coming from both T_1 and T_2, and then replacing by arcs all paths between pairs of remaining nodes that do not contain other remaining nodes. Unfortunately, the justification for these simple constructions, as well as the proof of their algebraic meaning, is rather intricate, and at some points it differs substantially for the different notions of embedding.

Beyond their theoretical interest, these constructions provide an efficient solution of the smallest common supertree problem under homeomorphic embeddings, for which no algorithm was known until now. The solution extends the largest common homeomorphic subtree algorithm of [19], which in turn extended the subtree homeomorphism algorithm of [27,28]. Likewise, these constructions also provide a solution to the smallest common supertree problem under minor embeddings, for which no algorithm was known previously, either. The solution extends the unordered tree edit algorithm of [22].

2. Preliminaries

In this section we recall the categorical notions of pushouts and pullbacks, as they are needed in the following sections, and the notions of isomorphic, homeomorphic, topological, and minor embeddings of trees, together with some results about them that will be used in the rest of the paper.

2.1. Pushouts and pullbacks

A *category* is a structure consisting of: a class of *objects*; for every pair of objects A, B, a class $\mathrm{Mor}(A, B)$ of *morphisms*; and, for every objects A, B, C, a binary operation

$$\begin{aligned} \circ : \mathrm{Mor}(A, B) \times \mathrm{Mor}(B, C) &\to \mathrm{Mor}(A, C) \\ (f, g) &\mapsto g \circ f \end{aligned}$$

called *composition*, which satisfies the following two properties:

Associativity: for every $f \in \mathrm{Mor}(A, B)$, $g \in \mathrm{Mor}(B, C)$, and $h \in \mathrm{Mor}(C, D)$, $h \circ (g \circ f) = (h \circ g) \circ f \in \mathrm{Mor}(A, D)$.

Existence of identities: for every object A, there exists an *identity morphism* $\mathrm{Id}_A \in \mathrm{Mor}(A, A)$ such that $\mathrm{Id}_A \circ f = f$, for every $f \in \mathrm{Mor}(B, A)$, and $g \circ \mathrm{Id}_A = g$, for every $g \in \mathrm{Mor}(A, B)$.

It is usual to indicate that $f \in \mathrm{Mor}(A, B)$ by writing $f : A \to B$.

All categories considered in this paper have all trees as objects and different types of embeddings of trees as morphisms: see the next subsection.

A *pushout* in a category $\mathcal{C}$ of two morphisms $f_1 : A \to B_1$ and $f_2 : A \to B_2$ is an object P together with two morphisms $g_1 : B_1 \to P$ and $g_2 : B_2 \to P$ satisfying the following two conditions:

(i) $g_1 \circ f_1 = g_2 \circ f_2$.

(ii) (*Universal property*) If X is any object together with a pair of morphisms $g_1' : B_1 \to X$ and $g_2' : B_2 \to X$ such that $g_1' \circ f_1 = g_2' \circ f_2$, then there exists a unique morphism $h : P \to X$ such that $h \circ g_1 = g_1'$ and $h \circ g_2 = g_2'$.

A *pullback* in a category $\mathcal{C}$ of two morphisms $f_1 : A_1 \to B$ and $f_2 : A_2 \to B$ is an object Q together with two morphisms $g_1 : Q \to A_1$ and $g_2 : Q \to A_2$ satisfying the following two conditions:

(i) $f_1 \circ g_1 = f_2 \circ g_2$.

(ii) (*Universal property*) If X is any object together with a pair of morphisms $g_1' : X \to A_1$ and $g_2' : X \to A_2$ such that $f_1 \circ g_1' = f_2 \circ g_2'$, then there exists a unique morphism $h : X \to Q$ such that $g_1' = g_1 \circ h$ and $g_2' = g_2 \circ h$.

Two pushouts in $\mathcal{C}$ of the same pair of morphisms, as well as two pullbacks in $\mathcal{C}$ of the same pair of morphisms, are always isomorphic in $\mathcal{C}$.

2.2. Embeddings of trees

A *directed graph* is a structure $G = (V, E)$ consisting of a set V, whose elements are called *nodes*, and a set E of ordered pairs $(a, b) \in V \times V$ with $a \neq b$; the elements of E are called *arcs*. For every arc $(v, w) \in E$, v is its *source* node and w its *target* node. A graph is *finite* if its set of nodes is finite. The *in-degree* of a node v in a finite graph is the number of arcs that have v as target node and its *out-degree* is the number of arcs that have v as source node.

An *isomorphism* $f : G \to G'$ between graphs $G = (V, E)$ and $G' = (V', E')$ is a bijective mapping $f : V \to V'$ such that, for every $a, b \in V$, $(a, b) \in E$ if and only if $(f(a), f(b)) \in E'$.

A *path* in a directed graph $G = (V, E)$ is a sequence of nodes $(v_0, v_1, \ldots, v_k)$ such that $(v_0, v_1), (v_1, v_2), (v_2, v_3), \ldots, (v_{k-1}, v_k) \in E$; its *origin* is v_0, its *end* is v_k, and its *intermediate nodes* are $v_1, \ldots, v_{k-1}$. Such a path is *non-trivial* if $k \geqslant 1$. We shall represent a path *from a to b*, that is, a path with origin a and end b, by $a \rightsquigarrow b$.

A (*rooted*) *tree* is a directed finite graph $T = (V, E)$ with V either empty or containing a distinguished node $r \in V$, called the *root*, such that for every other node $v \in V$ there exists one, and only one, path $r \rightsquigarrow v$. Note that every node in a tree has in-degree 1, except the root that has in-degree 0. Henceforth, and unless otherwise stated, given a tree T we shall denote its set of nodes by $V(T)$ and its set of arcs by $E(T)$. The *size* of a tree T is its number $|E(T)|$ of arcs.

The *children* of a node v in a tree T are those nodes w such that $(v, w) \in E(T)$: in this case we also say that v is the *parent* of its children. The only node without parent is the root, and the nodes without children are the *leaves* of the tree.

A path $(v_0, v_1, \ldots, v_k)$ in a tree T is *elementary* if, for every $i = 1, \ldots, k-1$, v_{i+1} is the only child of v_i; in other words, if all its intermediate nodes have out-degree 1. In particular, an arc forms an elementary path.

Two non-trivial paths $(a, v_1, \ldots, v_k)$ and $(a, w_1, \ldots, w_\ell)$ in a tree T are said to *diverge* if their origin a is their only common node. Note that, by the uniqueness of paths in trees, this condition is equivalent to $v_1 \neq w_1$. The definition of trees also implies that, for every two nodes b, c of a tree that are not connected by a path, there exists one, and only one, node a such that there exist divergent paths $a \rightsquigarrow b$ and $a \rightsquigarrow c$: we shall call this node the *least common ancestor* of b and c. The adjective "least" refers to the obvious fact that if there exist paths from a node x to b and to c, then these paths consist of a path from x to the least common ancestor of b and c followed by the divergent paths from this node to b and c.

Definition 1. Let S and T be trees.

(i) S is a *minor* of T if there exists an injective mapping $f : V(S) \to V(T)$ satisfying the following condition: for every $a, b \in V(S)$, if $(a, b) \in E(S)$, then there exists a path $f(a) \rightsquigarrow f(b)$ in T with no intermediate node in $f(V(S))$. In this case, the mapping f is said to be a *minor embedding* $f : S \to T$.

(ii) S is a *topological subtree* of T if there exists a minor embedding $f : S \to T$ such that, for every $(a, b), (a, c) \in E(S)$ with $b \neq c$, the paths $f(a) \rightsquigarrow f(b)$ and $f(a) \rightsquigarrow f(c)$ in T diverge. In this case, f is called a *topological embedding* $f : S \to T$.

(iii) S is a *homeomorphic subtree* of T if there exists a minor embedding $f : S \to T$ satisfying the following extra condition: for every $(a, b) \in E(S)$, the path $f(a) \rightsquigarrow f(b)$ in T is elementary. In this case, f is said to be a *homeomorphic embedding* $f : S \to T$.

(iv) S is an *isomorphic subtree* of T if there exists an injective mapping $f : V(S) \to V(T)$ satisfying the following condition: if $(a, b) \in E(S)$, then $(f(a), f(b)) \in E(T)$. Such a mapping f is called an *isomorphic embedding* $f : S \to T$.

Lemma 2. *Every isomorphic embedding is a homeomorphic embedding, every homeomorphic embedding is a topological embedding, and every topological embedding is a minor embedding.*

Proof. It is obvious from the definitions that every isomorphic embedding is a homeomorphic embedding and that every topological embedding is a minor embedding. Now, let $f : S \to T$ be a homeomorphic embedding and let $(a, b), (a, c) \in E(S)$ be such that $b \neq c$. Then, the paths $f(a) \rightsquigarrow f(b)$ and $f(a) \rightsquigarrow f(c)$ are elementary and they do not contain any intermediate node in $f(V(S))$. This implies that neither $f(b)$ is intermediate in the path $f(a) \rightsquigarrow f(c)$, nor $f(c)$ is intermediate in the path $f(a) \rightsquigarrow f(b)$. Therefore, $f(b)$ and $f(c)$ are not connected by a path. But then the least common ancestor x of $f(b)$ and $f(c)$ must have out-degree at least 2, and thus it cannot be intermediate in the paths from $f(a)$ to these nodes. Since there exists a path $f(a) \rightsquigarrow x$, we conclude that $f(a) = x$, that is, the paths $f(a) \rightsquigarrow f(b)$ and $f(a) \rightsquigarrow f(c)$ diverge. This shows that f is a topological embedding. □

The implications in the last lemma are strict, as the following example shows.

Example 3. Let S and T be the trees described in Fig. 1, with roots r and 1, respectively.

(a) The mapping $f_0 : V(S) \to V(T)$ defined by $f_0(r) = 1$, $f_0(x) = 3$ and $f_0(y) = 4$ is not a minor embedding, because, although it transforms arcs in S into paths in T, the path $f_0(r) \rightsquigarrow f_0(y)$ contains the node $3 = f_0(x)$, which belongs to $f_0(V(S))$.

(b) The mapping $f_1 : V(S) \to V(T)$ defined by $f_1(r) = 1$, $f_1(x) = 5$ and $f_1(y) = 6$ is a minor embedding, because the arcs $(r, x), (r, y) \in E(S)$ become paths $f_1(r) \rightsquigarrow f_1(x)$ and $f_1(r) \rightsquigarrow f_1(y)$ in T with no intermediate node in $f_1(V(S))$. But it is not a topological embedding, because these paths do not diverge.

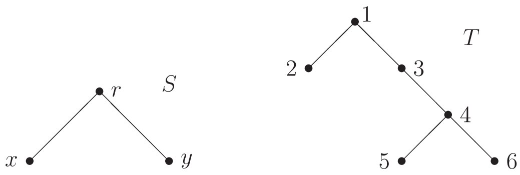

Fig. 1. The trees *S* and *T* in Example 3.

(c) The mapping $f_2 : V(S) \to V(T)$ defined by $f_2(r) = 1$, $f_2(x) = 2$ and $f_2(y) = 6$ is a topological embedding, because the arcs $(r, x), (r, y) \in E(S)$ become divergent paths $f_2(r) \rightsquigarrow f_2(x)$ and $f_2(r) \rightsquigarrow f_2(y)$ in T without intermediate nodes in $f_2(V(S))$. But it is not a homeomorphic embedding, because the path $f_2(r) \rightsquigarrow f_2(y)$ contains an intermediate node with more than one child.

(d) The mapping $f_3 : V(S) \to V(T)$ defined by $f_3(r) = 1$, $f_3(x) = 2$ and $f_3(y) = 4$ is a homeomorphic embedding, because the arcs $(r, x), (r, y) \in E(S)$ become elementary paths $f_3(r) \rightsquigarrow f_3(x)$ and $f_3(r) \rightsquigarrow f_3(y)$ in T with no intermediate node in $f_3(V(S))$. But it is not an isomorphic embedding, because the path $f_3(r) \rightsquigarrow f_3(y)$ is not an arc.

(e) The mappings $f_4 : V(S) \to V(T)$ defined by $f_4(r) = 1$, $f_4(x) = 2$ and $f_4(y) = 3$, and $f_5 : V(S) \to V(T)$ defined by $f_5(r) = 4$, $f_5(x) = 5$ and $f_5(y) = 6$ are isomorphic embeddings, because they transform every arc in S into an arc in T.

The following lemmas will be used several times in the next sections.

Lemma 4. *Let $f : S \to T$ be a minor embedding. For every $a, b \in V(S)$, there exists a path $a \rightsquigarrow b$ in S if and only if there exists a path $f(a) \rightsquigarrow f(b)$ in T. Moreover, if the path $f(a) \rightsquigarrow f(b)$ is elementary, then the path $a \rightsquigarrow b$ is also elementary, and if there is an arc from $f(a)$ to $f(b)$ in T, then there is an arc from a to b in S.*

Proof. Since the arcs in S become paths in T without intermediate nodes in $f(V(S))$, it is obvious that a path $a \rightsquigarrow b$ in S becomes, under f, a path $f(a) \rightsquigarrow f(b)$ in T whose intermediate nodes belonging to $f(V(S))$ are exactly the images under f of the intermediate nodes of the path $a \rightsquigarrow b$.

Assume now that there exists a path $f(a) \rightsquigarrow f(b)$ in T, and let r be the root of S. If $a = r$ or $a = b$, it is clear that there exists a path $a \rightsquigarrow b$ in S. If $a \neq r$ and $a \neq b$, then the images of the paths $r \rightsquigarrow a$ and $r \rightsquigarrow b$ in S are paths $f(r) \rightsquigarrow f(a)$ and $f(r) \rightsquigarrow f(b)$ in T. Now, the uniqueness of paths in T implies that the path $f(r) \rightsquigarrow f(b)$ splits into the path $f(r) \rightsquigarrow f(a)$ and the path $f(a) \rightsquigarrow f(b)$. Therefore, $f(a)$ is an intermediate node of the path $f(r) \rightsquigarrow f(b)$. As a consequence, since f is injective and any intermediate node of this path belonging to $f(V(S))$ must be the image under f of an intermediate node of the path $r \rightsquigarrow b$, the node a must be intermediate in the path $r \rightsquigarrow b$, which yields a path $a \rightsquigarrow b$ in S.

Moreover, if a node in S has more than one child, then its image under f has also more than one child. This implies that if the path $f(a) \rightsquigarrow f(b)$ is elementary, then the path $a \rightsquigarrow b$ is elementary, too. Finally, if there is an arc from $f(a)$ to $f(b)$, then the path $a \rightsquigarrow b$ cannot have any intermediate node: it must be an arc. □

By Lemma 2, the last lemma applies also to isomorphic, homeomorphic, and topological embeddings.

Lemma 5. *Let $f : S \to T$ be a topological embedding. For every $a, b \in V(S)$ not connected by a path, if x is their least common ancestor in S, then $f(x)$ is the least common ancestor of $f(a)$ and $f(b)$ in T.*

Proof. Since a and b are not connected by a path in S, by the last lemma we know that $f(a)$ and $f(b)$ are not connected by a path in T, either. Let now x be the least common ancestor of a and b in S, and let v and w be the children of x contained in the divergent paths $x \rightsquigarrow a$ and $x \rightsquigarrow b$, respectively. Then, since f is a topological embedding, there exist in T divergent paths $f(x) \rightsquigarrow f(v)$ and $f(x) \rightsquigarrow f(w)$, which are followed by paths $f(v) \rightsquigarrow f(a)$ and $f(w) \rightsquigarrow f(b)$, respectively. This means that $f(x)$ is the node in T from which there exist divergent paths to $f(a)$ and to $f(b)$, that is, the least common ancestor of these two nodes. □

By Lemma 2, the last lemma applies also to isomorphic and homeomorphic embeddings. But the thesis of this lemma need not hold if f is only a minor embedding: see, for instance Example 3(b), where r is the least common ancestor of x and y, but the least common ancestor of $f_1(x) = 5$ and $f_1(y) = 6$ is 4, and not $f_1(r) = 1$.

Lemma 6. *Every bijective minor embedding is an isomorphism of graphs.*

Proof. Let $f : S \to T$ be a minor embedding such that $f : V(S) \to V(T)$ is bijective, and let $a, b \in V(S)$. If $(a, b) \in E(S)$, then there exists a path $f(a) \rightsquigarrow f(b)$ in T without any intermediate node in $f(V(S))$. Since f is bijective, this means that this path has no intermediate node, and thus it is an arc. This proves that if $(a, b) \in E(S)$, then $(f(a), f(b)) \in E(T)$. The converse implication is given by Lemma 4. □

By Lemma 2, the last lemma implies that every bijective isomorphic, homeomorphic, or topological embedding is an isomorphism of graphs.

Definition 7. Let S and T be trees.

(i) A *largest common isomorphic subtree* (*homeomorphic subtree*, *topological subtree*, *minor*) of S and T is a tree that is an isomorphic subtree (respectively, homeomorphic subtree, topological subtree, minor) of both of them and has the largest size among all trees with this property.

(ii) A *smallest common isomorphic supertree* (*homeomorphic supertree*, *topological supertree*, *supertree under minor embeddings*) of S and T is a tree such that both S and T are isomorphic subtrees (respectively, homeomorphic subtrees, topological subtrees, minors) of it and has the least size among all trees with this property.

We shall denote by $\mathsf{Tree}_{\mathrm{iso}}$, $\mathsf{Tree}_{\mathrm{hom}}$, $\mathsf{Tree}_{\mathrm{top}}$, and $\mathsf{Tree}_{\mathrm{min}}$ the categories with objects all trees and with morphisms the isomorphic, homeomorphic, topological, and minor embeddings, respectively. Whenever we denote generically any one of these categories by Tree_*, we shall use the following notations. By a Tree_**-embedding* we shall mean a morphism in the corresponding category. By a *common* Tree_**-subtree* of two trees we shall mean a tree together with Tree_*-embeddings into these two trees. By a *largest common* Tree_**-subtree* of two trees we shall mean a largest size common Tree_*-tree. By a *common* Tree_**-supertree* of two trees we shall mean a tree together with Tree_*-embeddings of these two trees into it. By a *smallest common* Tree_**-supertree* of two trees we shall mean a least size common Tree_*-supertree. And by a Tree_**-path* we shall understand an arc if Tree_* stands for $\mathsf{Tree}_{\mathrm{iso}}$, an elementary path if Tree_* denotes $\mathsf{Tree}_{\mathrm{hom}}$, and an arbitrary path if Tree_* means $\mathsf{Tree}_{\mathrm{top}}$ or $\mathsf{Tree}_{\mathrm{min}}$. Note in particular that all trivial paths and all arcs are Tree_*-paths, for every category Tree_*.

The following corollary is a simple rewriting of the definitions.

Corollary 8. *Let* Tree_* *denote any category* $\mathsf{Tree}_{\mathrm{iso}}$, $\mathsf{Tree}_{\mathrm{hom}}$, *or* $\mathsf{Tree}_{\mathrm{min}}$. *For every trees S, T, a mapping $f : V(S) \to V(T)$ is a* Tree_**-embedding if and only if, for every $(a, b) \in E(S)$, there is a* Tree_**-path $f(a) \rightsquigarrow f(b)$ in T with no intermediate node belonging to $f(V(S))$.*

And the following corollary is a direct consequence of Lemma 4.

Corollary 9. *Let* Tree_* *be any category* $\mathsf{Tree}_{\mathrm{iso}}$, $\mathsf{Tree}_{\mathrm{hom}}$, $\mathsf{Tree}_{\mathrm{top}}$, *or* $\mathsf{Tree}_{\mathrm{min}}$, *and let $f : S \to T$ be a* Tree_**-embedding. For every $a, b \in V(S)$, if there exists a* Tree_**-path $f(a) \rightsquigarrow f(b)$ in T, then there exists a* Tree_**-path $a \rightsquigarrow b$ in S.*

Finally, we have the following result, which will be used later.

Lemma 10. *Let* Tree_* *be any category* $\mathsf{Tree}_{\mathrm{iso}}$, $\mathsf{Tree}_{\mathrm{hom}}$, $\mathsf{Tree}_{\mathrm{top}}$, *or* $\mathsf{Tree}_{\mathrm{min}}$, *let S, T, U be trees and $f : V(S) \to V(T)$ and $g : V(T) \to V(U)$ mappings between their sets of nodes. If $g \circ f : S \to U$ and $g : T \to U$ are* Tree_**-embeddings, then $f : S \to T$ is also a* Tree_**-embedding.*

Proof. Since $g \circ f$ is injective, it is clear that f is injective. Let now $a, b \in S$ be such that $(a, b) \in E(S)$. Since $g \circ f : S \to U$ is a Tree_*-embedding, there exists a Tree_*-path $g(f(a)) \rightsquigarrow g(f(b))$ in U without any intermediate node in $g(f(V(S)))$. Since $g : T \to U$ is a Tree$_*$-embedding, the existence of this path $g(f(a)) \rightsquigarrow g(f(b))$ in U implies, by Corollary 9, the existence of a Tree$_*$-path $f(a) \rightsquigarrow f(b)$ in T. This path cannot have any intermediate node in $f(V(S))$, because any such intermediate node would become, under g, an intermediate node belonging to $g(f(V(S)))$ in the path $g(f(a)) \rightsquigarrow g(f(b))$.

So, f is injective and if $(a, b) \in E(S)$, then there exists a Tree$_*$-path $f(a) \rightsquigarrow f(b)$ in T without intermediate nodes in $f(V(S))$. This already shows, by Corollary 8, that f is a Tree$_*$-embedding when Tree$_*$ stands for Tree$_{\text{iso}}$, Tree$_{\text{hom}}$, or Tree$_{\text{min}}$.

As far as Tree$_{\text{top}}$ goes, we have already proved that f transforms arcs into paths without intermediate nodes in $f(V(S))$, and thus it remains to prove that if $a, b, c \in V(S)$ are such that $(a, b), (a, c) \in E(S)$ and $b \neq c$, then the paths $f(a) \rightsquigarrow f(b)$ and $f(a) \rightsquigarrow f(c)$ in T diverge. But since $g \circ f$ is a topological embedding, the paths $g(f(a)) \rightsquigarrow g(f(b))$ and $g(f(a)) \rightsquigarrow g(f(c))$ in U are divergent, and this clearly implies that the paths $f(a) \rightsquigarrow f(b)$ and $f(a) \rightsquigarrow f(c)$ in T are divergent, too: any common intermediate node in these paths would become, under g, a common intermediate node in the paths $g(f(a)) \rightsquigarrow g(f(b))$ and $g(f(a)) \rightsquigarrow g(f(c))$. □

3. Common subtrees as pullbacks

In this section we study the construction of common subtrees as pullbacks of embeddings into common supertrees, for each one of the types of tree embeddings considered in this paper. We start with the most general type, minor embeddings.

Let $f_1 : T_1 \to T$ and $f_2 : T_2 \to T$ be henceforth two minor embeddings. Without any loss of generality, and unless otherwise stated, we shall assume that $V(T_1), V(T_2) \subseteq V(T)$ and that the minor embeddings f_1 and f_2 are given by these inclusions. For simplicity, we shall denote thus the image of a node $a \in V(T_i)$ under the corresponding f_i again by a.

Let T_p be the graph with set of nodes $V(T_p) = V(T_1) \cap V(T_2)$ and set of arcs defined in the following way: for every $a, b \in V(T_1) \cap V(T_2)$, $(a, b) \in E(T_p)$ if and only if there are paths $a \rightsquigarrow b$ in T_1 and in T_2 without intermediate nodes in $V(T_1) \cap V(T_2)$. We shall call this graph T_p the *intersection* of T_1 and T_2 obtained through f_1 and f_2.

This graph satisfies the following useful lemma.

Lemma 11. *For every $a, b \in V(T_1) \cap V(T_2)$:*
(i) *If there exists a path $a \rightsquigarrow b$ in T_p, then there exist paths $a \rightsquigarrow b$ in T_1 and in T_2.*
(ii) *If there exists a path $a \rightsquigarrow b$ in some T_i, $i = 1, 2$, then there exists also a path $a \rightsquigarrow b$ in T_p, and its intermediate nodes are exactly the intermediate nodes of the path $a \rightsquigarrow b$ in T_i that belong to $V(T_1) \cap V(T_2)$.*

Proof. Point (i) is a direct consequence of the fact that every arc in T_p corresponds to paths in T_1 and T_2.

As far as point (ii) goes, we shall prove that if there exists a path $a \rightsquigarrow b$ in T_1, then there exists also a path $a \rightsquigarrow b$ in T_p with intermediate nodes the intermediate nodes of the path in T_1 that belong to $V(T_1) \cap V(T_2)$, by induction on the number n of such intermediate nodes belonging to $V(T_1) \cap V(T_2)$.

If $n = 0$, then there exists a path $a \rightsquigarrow b$ in T_1 that does not contain any intermediate node in $V(T_1) \cap V(T_2)$. Since f_1 transforms arcs into paths with no intermediate node belonging to T_1, this implies that there exists a path $a \rightsquigarrow b$ in T that does not contain any node in $V(T_1) \cap V(T_2)$, either. Then, by Lemma 4, this path is induced by a path $a \rightsquigarrow b$ in T_2, and by the same reason this path does not contain any intermediate node in $V(T_1) \cap V(T_2)$. So, there are paths $a \rightsquigarrow b$ in T_1 and T_2 without intermediate nodes in $V(T_1) \cap V(T_2)$, and therefore, by definition, there exists an arc from a to b in T_p.

As the induction hypothesis, assume that the claim is true for paths in T_1 with n intermediate nodes in $V(T_1) \cap V(T_2)$, and assume now that the path $a \rightsquigarrow b$ has $n + 1$ such nodes. Let a_0 be the first intermediate node of this path belonging to $V(T_1) \cap V(T_2)$. Then, by the case $n = 0$, there is an arc in T_p from a to a_0, and by the induction hypothesis there is a path $a_0 \rightsquigarrow b$ in T_p whose only intermediate nodes are the intermediate nodes of the path $a_0 \rightsquigarrow b$ in T_1 that belong to $V(T_1) \cap V(T_2)$; by concatenating these paths in T_p we obtain the path $a \rightsquigarrow b$ we were looking for. □

The intersection of two minors need not be a tree, as the following simple example shows.

Example 12. Let T be a tree with nodes a_1, a_2, b, c and arcs (a_1, a_2), (a_2, b), (a_2, c), let T_1 be its minor with nodes a_1, b, c and arcs (a_1, b), (a_1, c), and let T_2 be its minor with nodes a_2, b, c and arcs (a_2, b), (a_2, c). In this case T_p is the graph with nodes b, c and no arc, and in particular it is not a tree.

Now we have the following result.

Proposition 13. *T_1 and T_2 have always a common minor, which is either T_p together with its inclusions in T_1 and T_2, or obtained by adding a root to T_p.*

Proof. If T_p is empty, then it is a tree and its inclusions into T_1 and T_2 are clearly minor embeddings. In this case, T_p is a common minor of T_1 and T_2.

So, assume in the sequel that T_p is non-empty. If it had no node without parents, then it would contain a circuit and this would imply, by Lemma 11(i), the existence of circuits in the trees T_1 and T_2, which is impossible. Therefore, T_p contains nodes without parent. Now we must consider two cases:

(1) T_p has only one node r_p without a parent. Then every other node a in T_p can be reached from r_p through a path, because this graph does not contain any circuit (as we have seen) and hence it must contain a path from a node of in-degree 0 to a. To check that this path is unique, we shall prove that no node in T_p has in-degree greater than 1.

Indeed, assume that there are nodes $a, b, c \in V(T_p)$, with $b \neq c$, and arcs from b and c to a. This means that there are paths in T_1 and in T_2 from b and c to a that do not contain any intermediate node in $V(T_1) \cap V(T_2)$. But since, say, T_1 is a tree, if there exist paths $b \rightsquigarrow a$ and $c \rightsquigarrow a$ in T_1, one of the nodes b or c must be intermediate in the path from the other one to a, which yields a contradiction.

This proves that, in this case, T_p is a tree. And by definition, for every $a, b \in V(T_p)$, if $(a, b) \in T_p$, then there are paths $a \rightsquigarrow b$ in T_1 and in T_2 without any intermediate node in $V(T_p)$. Therefore, the inclusions $\imath_i : V(T_p) \hookrightarrow V(T_i)$ induce minor embeddings $\imath_i : T_p \to T_i$, for $i = 1, 2$, and hence T_p is a common minor of T_1 and T_2.

(2) T_p contains more than one node without a parent, say $x_1, \ldots, x_k$. The same argument used in (1) shows in this case that every other node $a \in V(T_p)$ can be reached from one of these nodes x_i through a path in T_p, and that no node in T_p has in-degree greater than 1.

Let now $\widetilde{T}_p$ be the graph obtained by adding to T_p one node r and arcs (r, x_i), for $i = 1, \ldots, k$. Then, r is the only node without a parent in $\widetilde{T}_p$ and every node in it is reached from r through a unique path. Indeed, each x_i is reached from r through the new arc (r, x_i), and then every other node in $\widetilde{T}_p$ is reached from r by the path going from some x_i to it in T_p preceded by the arc from r to this x_i. And these paths are unique, because no node in $\widetilde{T}_p$ has in-degree greater than 1. Therefore, $\widetilde{T}_p$ is a tree with root r.

Now, note that there is no non-trivial path in either T_1 or T_2 from any node belonging to $V(T_1) \cap V(T_2)$ to any x_i: such a path, by Lemma 11, would induce a non-trivial path in T_p and therefore the node x_i would have a parent in T_p. This implies in particular that neither the root of T_1 nor the root of T_2 belong to $V(T_1) \cap V(T_2)$: since $k \geqslant 2$, there are non-trivial paths from each one of these roots to some x_i.

Consider then the injective mappings $\widetilde{\imath}_i : \widetilde{T}_p \to T_i$, $i = 1, 2$, defined by the inclusions on $V(T_p)$ and sending r to the root of the corresponding T_i. It is clear that they are minor embeddings: on the one hand, arguing as in (1) above, we obtain that the restriction of each $\widetilde{\imath}_i$ to T_p sends every arc to a path in T_i without any intermediate node coming from $\widetilde{T}_p$; on the other hand, $\widetilde{\imath}_i$ sends every arc (r, x_ℓ) to the path in T_i going from its root to x_ℓ, which, as we saw above, does not contain any intermediate node in $V(T_1) \cap V(T_2)$. Thus, $\widetilde{T}_p$ is a common minor of T_1 and T_2. □

If we restrict ourselves from minor embeddings to topological embeddings, then only the first case in the last proposition can happen.

Proposition 14. *If $f_1 : T_1 \to T$ and $f_2 : T_2 \to T$ are topological embeddings, then T_p is a tree and the inclusions $V(T_p) \hookrightarrow V(T_i)$ are topological embeddings $\imath_i : T_p \to T_i$, for $i = 1, 2$, and therefore T_p is a common topological subtree of T_1 and T_2.*

Proof. Let us prove first of all that if f_1 and f_2 are not only minor but topological embeddings, then T_p does not have more than one node without a parent. Indeed, assume that $a, b \in V(T_1) \cap V(T_2)$ have no parent in T_p. Then, neither

T_1 nor T_2 contains any non-trivial path from some node in $V(T_1) \cap V(T_2)$ to a or b, because, by Lemma 11, such a path would imply a non-trivial path in T_p finishing in a or b and then one of these nodes would have a parent in T_p. In particular, there is no path connecting a and b in either T_1 or T_2. For every $i = 1, 2$, let $x_i \in V(T_i)$ be the least common ancestor of a and b in T_i. By Lemma 5, each x_i is also the least common ancestor of a and b in T. But then $x_1 = x_2 \in V(T_1) \cap V(T_2)$ and therefore both a and b can be reached from a node in $V(T_1) \cap V(T_2)$ through paths in T_1 and in T_2, which yields a contradiction.

So, since every topological embedding is a minor embedding, from the proof of Proposition 13 we know that the fact that T_p has at most (and hence, exactly) one node without a parent implies that it is a tree. Let us prove now that $\imath_1$ is a topological embedding. By point (1) in the proof of Proposition 13, we already know that it is a minor embedding. So, it remains to prove that if there are arcs from a to b and to c in T_p, then the paths $a \rightsquigarrow b$ and $a \rightsquigarrow c$ in T_1 diverge.

To prove it, note that, since, by the definition of T_p, the paths $a \rightsquigarrow b$ and $a \rightsquigarrow c$ in T_1 and in T_2 have no intermediate node in $V(T_1) \cap V(T_2)$, neither b nor c appears in the path from a to the other one, and therefore there is no path connecting b and c. Thus, if, for every $i = 1, 2$, $x_i \in V(T_i)$ denotes the least common ancestor of b and c in T_i, then, arguing as before, we deduce that $x_1 = x_2$ and in particular that this node belongs to $V(T_1) \cap V(T_2)$.

Now, the existence of the paths $a \rightsquigarrow b$ and $a \rightsquigarrow c$ in T_1, implies that either $x_1 = a$ or there exists a non-trivial path in T_1 from a to x_1. But the paths $a \rightsquigarrow b$ and $a \rightsquigarrow c$ in T_1 do not contain any intermediate node belonging to $V(T_1) \cap V(T_2)$, and therefore it must happen that $a = x_1$ and the paths $a \rightsquigarrow b$ and $a \rightsquigarrow c$ in T_1 diverge, as we wanted to prove. □

We have similar results if f_1 and f_2 are not only topological, but homeomorphic or isomorphic embeddings.

Proposition 15. *If $f_1 : T_1 \to T$ and $f_2 : T_2 \to T$ are homeomorphic embeddings, then T_p is a tree and the inclusions $V(T_p) \hookrightarrow V(T_i)$ are homeomorphic embeddings $\imath_i : T_p \to T_i$, for $i = 1, 2$, and therefore T_p is a common homeomorphic subtree of T_1 and T_2.*

Proof. We already know from Proposition 14 that T_p is a tree and that the inclusions $\imath_1 : T_p \to T_1$ and $\imath_2 : T_p \to T_2$ are topological embeddings. It remains to prove that they are not only topological, but homeomorphic embeddings. We shall do it for $\imath_1 : T_p \to T_1$.

Let $a, b \in V(T_p)$ be such that $(a, b) \in E(T_p)$. Then, by definition, there exists a path $a \rightsquigarrow b$ in T_1 without any intermediate node in $V(T_1) \cap V(T_2)$. Assume that this path has an intermediate node x with more than one child. The path $a \rightsquigarrow b$ induces, under the homeomorphic embedding $f_1 : T_1 \to T$, a path $a \rightsquigarrow b$ in T that contains x, and this node has also more than one child in T. Now, by Lemma 4, there is also a path $a \rightsquigarrow b$ in T_2. Since every arc in T_2 becomes, under the homeomorphic embedding $f_2 : T_2 \to T$, an elementary path in T, the nodes in the path $a \rightsquigarrow b$ in T that do not belong to $V(T_2)$ have only one child. Therefore, $x \in V(T_2)$ and hence $x \in V(T_1) \cap V(T_2)$, which contradicts the fact that the path $a \rightsquigarrow b$ in T_1 does not contain any intermediate node in $V(T_1) \cap V(T_2)$. This proves that this path is elementary, as we wanted. □

Proposition 16. *If $f_1 : T_1 \to T$ and $f_2 : T_2 \to T$ are isomorphic embeddings, then T_p is a tree and the inclusions $V(T_p) \hookrightarrow V(T_i)$ are isomorphic embeddings $\imath_i : T_p \to T_i$, for $i = 1, 2$, and therefore, T_p is a common isomorphic subtree of T_1 and T_2.*

Proof. We already know from Proposition 15 that T_p is a tree and that $\imath_1 : T_p \to T_1$ and $\imath_2 : T_p \to T_2$ are homeomorphic embeddings, i.e., that if $(a, b) \in E(T_p)$, then there are elementary paths $a \rightsquigarrow b$ in T_1 and in T_2 without any intermediate node in $V(T_1) \cap V(T_2)$. We want to prove that each one of these paths consists of a single arc, i.e., that a is the parent of b in both trees.

Let c_1 be the parent of b in T_1 and c_2 the parent of b in T_2: they exist because there is a path $a \rightsquigarrow b$ in each tree. Then, since T_1 and T_2 are isomorphic subtrees of T, both c_1 and c_2 are parents of b in T, and therefore $c_1 = c_2 \in V(T_1) \cap V(T_2)$. So, the parents in T_1 and in T_2 of b are the same and they belong to $V(T_1) \cap V(T_2)$. Since the paths $a \rightsquigarrow b$ in T_1 and in T_2 do not contain any intermediate node in $V(T_1) \cap V(T_2)$ and they must contain c_1 and c_2, respectively, this implies that $a = c_1 = c_2$, as we claimed. □

We have finally the following result, which gives an algebraic content to the construction of intersections in $\mathsf{Tree}_{\mathrm{iso}}$, $\mathsf{Tree}_{\mathrm{hom}}$, and $\mathsf{Tree}_{\mathrm{top}}$.

Proposition 17. *Let* Tree$_*$ *denote any category* Tree$_{\mathrm{iso}}$, Tree$_{\mathrm{hom}}$, *or* Tree$_{\mathrm{top}}$. *For every pair of* Tree$_*$*-embeddings* $f_1 : T_1 \to T$ *and* $f_2 : T_2 \to T$,

$$(T_p, \imath_1 : T_p \to T_1, \imath_2 : T_p \to T_2)$$

is a pullback of f_1 *and* f_2 *in* Tree$_*$.

Proof. We know from the previous propositions that, in each case, T_p is a tree and $\imath_1 : T_p \to T_1$ and $\imath_2 : T_p \to T_2$ are Tree$_*$-embeddings, and it is clear that $f_1 \circ \imath_1 = f_2 \circ \imath_2$. Let us check now the universal property of pullbacks in Tree$_*$.

Let S be any tree and let $g_1 : S \to T_1$ and $g_2 : S \to T_2$ be two Tree$_*$-embeddings such that $f_1 \circ g_1 = f_2 \circ g_2$. Then, at the level of nodes, there exists a unique mapping $g : V(S) \to V(T_1) \cap V(T_2) = V(T_p)$ such that each g_i is equal to g followed by the corresponding inclusion $\imath_i : V(T_p) \hookrightarrow V(T_i)$. And since each $\imath_i : T_p \to T_i$ and each composition $g_i = \imath_i \circ g : S \to T_i$ are Tree$_*$-embeddings, Lemma 10 implies that g is a Tree$_*$-embedding from S to T_p. This is the unique Tree$_*$ embedding that, when composed with $\imath_1$ and $\imath_2$, yields g_1 and g_2, respectively. □

Therefore, the categories Tree$_{\mathrm{iso}}$, Tree$_{\mathrm{hom}}$, and Tree$_{\mathrm{top}}$ have all binary pullbacks. It is not the case with Tree$_{\mathrm{min}}$, as the following simple example shows.

Remark 18. The minor embeddings $f_1 : T_1 \to T$ and $f_2 : T_2 \to T$ corresponding to the minors described in Example 12 do not have a pullback in Tree$_{\mathrm{min}}$. Indeed, let P, together with $g_1 : P \to T_1$ and $g_2 : P \to T_2$, be a pullback of them in Tree$_{\mathrm{min}}$. Then, since $f_1 \circ g_1 = f_2 \circ g_2 : V(P) \to V(T)$, we have that $g_1(V(P)) \subseteq \{b, c\}$ and $g_2(V(P)) \subseteq \{b, c\}$ and hence, P being a tree and g_1 and g_2 being minor embeddings, there are only two possibilities for P:

- P is empty. In this case, if we consider a tree Q with one node q and no arc, and the minor embeddings $h_1 : Q \to T_1$ and $h_2 : Q \to T_2$ given by $h_1(q) = h_2(q) = c$, then $f_1 \circ h_1 = f_2 \circ h_2$ but there is no minor embedding $h : Q \to P$ (because P is empty), which contradicts the definition of pullback.
- P consists of only one node, say $\{x\}$, and no arc, and g_1 and g_2 send x to the same node, b or c, in T_1 and in T_2. But then if we consider the same tree Q as before and the minor embeddings $h_1 : Q \to T_1$ and $h_2 : Q \to T_2$ that send q to the node different from $g_1(x)$ and $g_2(x)$, there is again no minor embedding $h : Q \to P$ such that $h_1 = g_1 \circ h$ and $h_2 = g_2 \circ h$, which contradicts the definition of pullback.

Nevertheless, arguing as in the proof of Proposition 17 we obtain the following result.

Proposition 19. *If* $f_1 : T_1 \to T$ *and* $f_2 : T_2 \to T$ *are minor embeddings such that* T_p *is a tree, then* $(T_p, \imath_1 : T_p \to T_1, \imath_2 : T_p \to T_2)$ *is a pullback of* f_1 *and* f_2 *in* Tree$_{\mathrm{min}}$.

Proof. We know from the proof of Proposition 13 that if T_p is a tree, then $\imath_1 : T_p \to T_1$ and $\imath_2 : T_p \to T_2$ are minor embeddings, and it is clear that $f_1 \circ \imath_1 = f_2 \circ \imath_2$. Then, exactly the same argument used in Proposition 17 shows that, in this case, $(T_p, \imath_1 : T_p \to T_1, \imath_2 : T_p \to T_2)$ satisfies the universal property of pushouts in Tree$_{\mathrm{min}}$. □

4. Common supertrees as pushouts

In this section we study the construction of common supertrees as pushouts of embeddings of largest common subtrees, for each one of the types of tree embeddings considered in this paper. Let Tree$_*$ be henceforth any one of the categories of trees Tree$_{\mathrm{iso}}$, Tree$_{\mathrm{hom}}$, Tree$_{\mathrm{top}}$, or Tree$_{\mathrm{min}}$.

Let T_1 and T_2 be two trees. Let T_μ be a largest common Tree$_*$-subtree of them, and let $m_1 : T_\mu \to T_1$ and $m_2 : T_\mu \to T_2$ be any Tree$_*$-embeddings. Let $T_1 + T_2$ be the graph obtained as the disjoint sum of the trees T_1 and T_2: that is,

$$V(T_1 + T_2) = V(T_1) \sqcup V(T_2), \quad E(T_1 + T_2) = E(T_1) \sqcup E(T_2).$$

Let θ be the equivalence relation on $V(T_1) \sqcup V(T_2)$ defined, up to symmetry, by the following condition:

$(a, b) \in \theta$ if and only if $a = b$ or there exists some $c \in V(T_\mu)$ such that $a = m_1(c)$ and $b = m_2(c)$.

We shall denote the equivalence class modulo θ of an element $x \in V(T_1) \sqcup V(T_2)$ by $[x]$.

Let T_{po} be the quotient graph of $T_1 + T_2$ by this equivalence:

- its set of nodes $V(T_{\mathrm{po}})$ is the quotient set $(V(T_1) \sqcup V(T_2))/\theta$, with elements the equivalence classes of the nodes of T_1 or T_2;
- its arcs are those induced by the arcs in T_1 or T_2, in the sense that $([a], [b]) \in E(T_{\mathrm{po}})$ if and only if there exist $a' \in [a]$, $b' \in [b]$ and some $i = 1, 2$ such that $(a', b') \in E(T_i)$.

Note that every equivalence class $[a] \in V(T_{\mathrm{po}})$ is either a 2-elements set $\{m_1(x), m_2(x)\}$, with $x \in V(T_\mu)$, or a singleton $\{a\}$, with $a \in V(T_i) - m_i(V(T_\mu))$ for some $i = 1, 2$. Since every node in T_1 and T_2 has in-degree at most 1, every $[a] \in V(T_{\mathrm{po}})$ has in-degree at most 2, and if it is 2, then $[a]$ must be of the first kind.

Let $\ell_i : V(T_i) \to V(T_{\mathrm{po}})$, $i = 1, 2$, denote the inclusion $V(T_i) \hookrightarrow V(T_1) \sqcup V(T_2)$ followed by the quotient mapping $V(T_1) \sqcup V(T_2) \to (V(T_1) \sqcup V(T_2))/\theta$: that is, $\ell_i(x) = [x]$ for every $x \in V(T_i)$. Note that, by construction,

$$V(T_{\mathrm{po}}) = \ell_1(V(T_1)) \cup \ell_2(V(T_2))$$

and

$$\ell_1(V(T_1)) \cap \ell_2(V(T_2)) = \ell_1(m_1(V(T_\mu))) = \ell_2(m_2(V(T_\mu))).$$

It is straightforward to check that these mappings ℓ_i are injective, satisfy that $\ell_1 \circ m_1 = \ell_2 \circ m_2$, and they define *morphisms of graphs* $\ell_i : T_i \to T_{\mathrm{po}}$, $i = 1, 2$, in the sense that if $(a, b) \in E(T_i)$, then $(\ell_i(a), \ell_i(b)) \in E(T_{\mathrm{po}})$.

We shall call this graph T_{po}, together with these injective morphisms $\ell_i : T_i \to T_{\mathrm{po}}$, $i = 1, 2$, the *join* of T_1 and T_2 obtained through m_1 and m_2.

Lemma 20. *Let T_1 and T_2 be trees, let T_μ be a largest common* Tree$_*$*-subtree of T_1 and T_2, let $m_1 : T_\mu \to T_1$ and $m_2 : T_\mu \to T_2$ be any* Tree$_*$*-embeddings, and let T_{po} be the join of T_1 and T_2 obtained through m_1 and m_2.*

(i) *If r is the root of T_μ, then $m_1(r)$ is the root of T_1 or r_2 is the root of T_2.*

(ii) *For every $x, y \in V(T_\mu)$, if T_{po} contains a path from $[m_1(x)] = [m_2(x)]$ to $[m_1(y)] = [m_2(y)]$, then T_μ contains a path from x to y.*

(iii) *T_{po} contains no circuit.*

Proof. (i) Assume that both $m_1(r)$ and $m_2(r)$ have parents, say v_1 and v_2, respectively. Lemma 4 implies that $v_i \notin m_i(V(T_\mu))$, for each $i = 1, 2$: otherwise, there would be an arc in T_μ from the preimage of v_i to r. Then, we can enlarge T_μ by adding a new node r_0 and an arc (r_0, r) and we can extend m_1 and m_2 to this new tree by sending r_0 to v_1 and v_2, respectively. In this way we obtain a tree strictly larger than T_μ and Tree$_*$-embeddings of this new tree into T_1 and T_2, against the assumption that T_μ is a largest Tree$_*$-subtree of them.

(ii) We shall prove that if T_{po} contains a path $[m_1(x)] \rightsquigarrow [m_1(y)]$, then T_μ contains a path $x \rightsquigarrow y$, by induction on its number n of intermediate nodes in $\ell_1(m_1(V(T_\mu))) = \ell_2(m_2(V(T_\mu)))$.

If $n = 0$, that is, if no intermediate node in the path $[m_1(x)] \rightsquigarrow [m_1(y)]$ comes from a node of T_μ, then all intermediate nodes come only from one of the trees T_1 or T_2: assume, to fix ideas, that they come from T_1, and that this path is

$$([m_1(x)], [v_1], \ldots, [v_k], [m_1(y)]),$$

with $[v_1], \ldots, [v_k] \in \ell_1(V(T_1)) - \ell_1(m_1(V(T_\mu)))$. Since the nodes of $\ell_1(V(T_1)) - \ell_1(m_1(V(T_\mu)))$ are (as equivalence classes) singletons, and an arc in T_{po} involving one node of this set must be induced by an arc in $E(T_1)$, we conclude that there exists a path

$$(m_1(x), v_1, \ldots, v_k, m_1(y))$$

in T_1. Since m_1 is a Tree$_*$-embedding, and in particular a minor embedding, by Lemma 4 this implies that there exists a path $x \rightsquigarrow y$ in T_μ.

As the induction hypothesis, assume that the claim is true for paths in T_{po} with n intermediate nodes in $\ell_1(m_1(V(T_\mu))) = \ell_2(m_2(V(T_\mu)))$, and assume now that the path $[m_1(x)] \rightsquigarrow [m_1(y)]$ has $n + 1$ such nodes. Let $[m_1(a)]$ be the first intermediate node of this path belonging to $\ell_1(m_1(V(T_\mu)))$. Then, by the case $n = 0$, there is a path $x \rightsquigarrow a$ in T_μ, and by the induction hypothesis there is a path $a \rightsquigarrow y$; by concatenating them we obtain the path $x \rightsquigarrow y$ in T_μ we were looking for.

(iii) Assume that T_{po} contains a circuit. If at most one node in this circuit belongs to $\ell_1(m_1(V(T_\mu)))$, then, arguing as in the proof of (ii), we conclude that all arcs in this circuit are induced by arcs in the same tree T_1 or T_2, and they would form a circuit in this tree, which is impossible. Therefore, two different nodes in this circuit must belong to $\ell_1(m_1(V(T_\mu)))$. This implies that there exist $x, y \in V(T_\mu)$, $x \neq y$, such that T_{po} contains a path $[m_1(x)] \rightsquigarrow [m_1(y)]$ and a path $[m_1(y)] \rightsquigarrow [m_1(x)]$. By point (ii), this implies that T_μ contains a path $x \rightsquigarrow y$ and a path $y \rightsquigarrow x$, and hence a circuit, which is impossible. Therefore, T_{po} cannot contain any circuit. □

Proposition 21. *Let T_1 and T_2 be trees, let T_μ be a largest common* Tree$_*$*-subtree of T_1 and T_2, let $m_1 : T_\mu \to T_1$ and $m_2 : T_\mu \to T_2$ be any* Tree$_*$*-embeddings, and let T_{po} be the join of T_1 and T_2 obtained through m_1 and m_2.*

(i) *For every $v, w \in V(T_{po})$, if $(v, w) \in E(T_{po})$ and there is another path $v \rightsquigarrow w$ in T_{po}, then $v, w \in \ell_1(V(T_1)) \cap \ell_2(V(T_2))$, this path is unique, it is a* Tree$_*$*-path and it has no intermediate node in $\ell_1(V(T_1)) \cap \ell_2(V(T_2))$. In particular, if* Tree$_*$ *is* Tree$_{iso}$*, then this situation cannot happen.*

(ii) *For every $v, w \in V(T_{po})$, if there are two different paths from v to w in T_{po} without any common intermediate node, then one of them is the arc (v, w), and then* (ii) *applies. In particular, again, this situation cannot happen if* Tree$_*$ *is* Tree$_{iso}$.

Proof. (i) If $(v, w) \in E(T_{po})$, then there exist, say, $a, b \in V(T_1)$ such that $v = [a]$, $w = [b]$, and $(a, b) \in E(T_1)$. Assume now that there is another path from $[a]$ to $[b]$ in T_{po}. Since T_{po} does not contain circuits by point (iii) in the last lemma, this path cannot contain $[a]$ or $[b]$ as an intermediate node, and therefore its first intermediate node is different from $[b]$ and its last intermediate node is different from $[a]$. In particular, $[b]$ has in-degree 2 in T_{po}.

This implies that there exists some $y \in V(T_\mu)$ such that $b = m_1(y)$ and there exists some $c \in V(T_2)$ such that $(c, m_2(y)) \in E(T_2)$, and that there is a non-trivial path in T_{po} from $[a]$ to $[c]$. Since $a \in V(T_1)$ and $c \in V(T_2)$, this path must contain some node belonging to $\ell_1(V(T_1)) \cap \ell_2(V(T_2))$. If it is not $[a]$, then let $[m_1(x)]$ be the first intermediate node in the path $[a] \rightsquigarrow [c]$ coming from T_μ. Since, in this case, $a \in V(T_1) - m_1(V(T_\mu))$, all intermediate nodes in the path $[a] \rightsquigarrow [m_1(x)]$ come also from $V(T_1) - m_1(V(T_\mu))$, and therefore there exists a path $a \rightsquigarrow m_1(x)$ in T_1. But, on the other hand, since there is a path $[m_1(x)] \rightsquigarrow [m_1(y)]$ in T_{po} (consisting of the path $[m_1(x)] \rightsquigarrow [c]$ followed by the arc $([c], [m_1(y)])$), from Lemma 20(ii) we deduce that there exists a path $x \rightsquigarrow y$ in T_μ and hence a path $m_1(x) \rightsquigarrow m_1(y) = b$ in T_1. Summarizing, if $a \notin m_1(V(T_\mu))$, then T_1 contains both an arc from a to $m_1(y)$ and a non-trivial path from a to $m_1(y)$ (through $m_1(x)$), which is impossible.

So, $a = m_1(x)$ for some $x \in V(T_\mu)$. Since $(m_1(x), m_1(y)) \in E(T_1)$, Lemma 4 implies that $(x, y) \in E(T_\mu)$, and therefore there exists a Tree$_*$-path in T_2 from $m_2(x)$ to $m_2(y)$ without any intermediate node in $m_2(V(T_\mu))$: the uniqueness of paths in trees implies that this path contains c as its last intermediate node before $m_2(y)$. To begin with, this already shows that the situation considered in this point cannot happen if Tree$_*$ is Tree$_{iso}$: a Tree$_{iso}$-path is an arc, and therefore it does not contain any intermediate node.

Thus, we assume now that Tree$_*$ is Tree$_{min}$, Tree$_{hom}$ or Tree$_{top}$. The Tree$_*$-path $m_2(x) \rightsquigarrow m_2(y)$ in T_2 without any intermediate node in $m_2(V(T_\mu))$ and containing c as the last intermediate node induces a path from $[m_2(x)] = [a]$ to $[b]$ in T_{po} containing $[c]$ and with all its intermediate nodes in $\ell_2(V(T_2)) - \ell_2(m_2(V(T_\mu)))$. This path is a Tree$_*$-path. If Tree$_*$ stands for Tree$_{min}$ or Tree$_{top}$, it is obvious, because in these cases Tree$_*$-paths are simply paths. If Tree$_*$ is Tree$_{hom}$, then all intermediate nodes in the path $m_2(x) \rightsquigarrow m_2(y)$ in T_2 have only one child, and since they belong to $V(T_2) - m_2(V(T_\mu))$ and therefore they are not identified with any node from T_1, their equivalence classes in T_{po} have also out-degree 1, and hence the path $[m_2(x)] \rightsquigarrow [m_2(y)]$ in T_{po} it induces is also elementary.

This proves that $v, w \in \ell_1(V(T_1)) \cap \ell_2(V(T_2))$ and that, besides the arc (v, w), there exists a Tree$_*$-path $v \rightsquigarrow w$ without any intermediate node in $\ell_1(V(T_1)) \cap \ell_2(V(T_2))$, which contains $[c]$. Assume finally that there exists a “third” path from v to w other than the arc and this Tree$_*$-path. Since w has in-degree at most 2 and T_{po} contains no circuit, arguing as in the first paragraph of this proof we deduce that this path consists of a path from v to $[c]$ followed by the arc $([c], w)$. But $[c]$ has in-degree 1 in T_{po}, as well as all intermediate nodes in the path from v to $[c]$ induced by the path $m_2(x) \rightsquigarrow c$ in T_2. Therefore, this is the only path in T_{po} from v to $[c]$. This shows that there is only one path from v to w in T_{po} other than the arc (v, w), and it is the Tree$_*$-path without any intermediate node in $\ell_1(V(T_1)) \cap \ell_2(V(T_2))$ obtained above.

(ii) Assume that there exist two different paths from v to w without any intermediate node in common, and let v_1 and v_2 be the nodes that precede w in each one of these two paths; by assumption $v_1 \neq v_2$ and $(v_1, w), (v_2, w) \in E(T_{po})$. Then, w has in-degree 2 in T_{po}, and this implies that there exist $y \in V(T_\mu)$, $b \in V(T_1)$ and $c \in V(T_2)$ such that, say,

$v_1 = [b]$, $v_2 = [c]$, $w = [m_1(y)] = [m_2(y)]$, and $(b, m_1(y)) \in E(T_1)$, $(c, m_2(y)) \in E(T_2)$. By Lemma 21(i), y has a parent x in T_μ, and then there are Tree_*-paths $m_1(x) \rightsquigarrow m_1(y)$ in T_1 and $m_2(x) \rightsquigarrow m_2(y)$ in T_2. This yields, up to symmetry, three possibilities:

- If $m_1(x) = b$ and $m_2(x) = c$, then $[b] = [c]$, against the assumption $v_1 \neq v_2$. Besides, if $\mathsf{Tree}_* = \mathsf{Tree}_{\mathrm{iso}}$, then, since $\mathsf{Tree}_{\mathrm{iso}}$-paths are arcs, it must happen that $m_1(x) = b$ and $m_2(x) = c$. So, if $\mathsf{Tree}_* = \mathsf{Tree}_{\mathrm{iso}}$, the situation described in the point we are proving cannot happen. In the remaining two cases we understand that $\mathsf{Tree}_* \neq \mathsf{Tree}_{\mathrm{iso}}$.
- If $m_1(x) = b$ and $m_2(x) \neq c$, then in T_{po} we have on the one hand the arc $([b], w)$ and on the other hand a path $[b] \rightsquigarrow w$ induced by the path $m_2(x) \rightsquigarrow m_2(y)$ in T_2: since c is the parent of $m_2(y)$ in T_2, it is the last intermediate node in the path $m_2(x) \rightsquigarrow m_2(y)$, and therefore $[c]$ is the last intermediate node in the path $[b] \rightsquigarrow w$ induced by $m_2(x) \rightsquigarrow m_2(y)$. By (i), these are the only two paths from $[b]$ to w.

  Let us prove now that the path $v \rightsquigarrow w$ containing $[c]$ also contains $[b]$. Assume first that this path contains some node in $\ell_2(m_2(T_\mu))$ other than w, and let $[m_2(z)]$ be the last such node before w. This means that T_{po} contains a path $[m_2(z)] \rightsquigarrow [m_2(y)]$ and therefore, by Lemma 21(ii), there is a path $z \rightsquigarrow y$ in T_μ. But then this path must contain the parent x of y, which implies that the path $[m_2(z)] \rightsquigarrow [m_2(y)]$, and hence the path $v \rightsquigarrow w$ through $[c]$, contains in this case $[b] = [m_2(x)]$.

  Assume now that the path $v \rightsquigarrow w$ containing $[c]$ does not contain any node in $\ell_2(m_2(T_\mu))$ other than w. Since $c \in V(T_2)$, this would mean that this path is completely induced by a path in T_2, that is, $v = [a]$ for some $a \in V(T_2) - m_2(V(T_\mu))$ and there exists a path $(a, \ldots, c, m_2(y))$ in T_2 with no intermediate node in $m_2(V(T_\mu))$. In this case, since there is a path $m_2(x) \rightsquigarrow m_2(y)$ in T_2 and $m_2(x)$ is not contained in the path $a \rightsquigarrow m_2(y)$, there would exist a non-trivial path $m_2(x) \rightsquigarrow a$, which would induce a path from $[b] = [m_2(x)]$ to $v = [a]$ forming a circuit with the path $v \rightsquigarrow [b]$. So, this case cannot happen.

  So, the path $v \rightsquigarrow w$ containing $[c]$ also contains $[b]$. But, by assumption, the paths $v \rightsquigarrow w$ containing $[b]$ and $[c]$ have no common intermediate node. Therefore, it must happen that $v = [b]$, and hence one of the paths from v to w is an arc, as it is claimed in the statement.
- If $m_1(x) \neq b$ and $m_2(x) \neq c$, then there are Tree_*-paths $(m_1(x), \ldots, b, m_1(y))$ and $(m_2(x), \ldots, c, m_2(y))$ in T_1 and T_2, respectively, without intermediate nodes coming from $V(T_\mu)$.

  In this case, we can enlarge T_μ by adding a new node x_0 and replacing the arc (x, y) by two arcs (x, x_0) and (x_0, y), and we can extend m_1 and m_2 to this new node by sending it, respectively, to b and c. It is clear that this new tree is strictly larger than T_μ. Moreover, the extensions of m_1 and m_2 are Tree_*-embeddings: the new arc (x, x_0) is transformed under them into the Tree_*-paths—without intermediate nodes coming from $V(T_\mu)$—that go from $m_1(x)$ to b and from $m_2(x)$ to c, respectively; the new arc (x_0, y) is transformed under them into the arcs $(b, m_1(y))$ and $(c, m_2(y))$, respectively; and it is clear that if m_1 and m_2 were topological embeddings, then their extensions are still so, because the new node x_0 has only one child. Thus, in this way we obtain a new common Tree_*-subtree of T_1 of T_2 that is strictly larger than T_μ, which yields a contradiction.

Summarizing, if Tree_* is $\mathsf{Tree}_{\mathrm{iso}}$, then there cannot exist two different paths $v \rightsquigarrow w$, and if Tree_* is $\mathsf{Tree}_{\mathrm{hom}}$, $\mathsf{Tree}_{\mathrm{top}}$, or $\mathsf{Tree}_{\mathrm{iso}}$, there can exist two different paths $v \rightsquigarrow w$ without common intermediate nodes, but then the only case that does not yield a contradiction is when one of these paths is an arc. □

Let now T_σ be the graph obtained from T_{po} by removing every arc that is subsumed by a path: that is, we remove from T_{po} each arc (v, w) for which there is another path $v \rightsquigarrow w$ in T_{po}. Note in particular that $V(T_\sigma) = V(T_{\mathrm{po}})$. We shall call this graph the Tree_*-*sum* of T_1 and T_2 obtained through m_1 and m_2.

As a direct consequence of Lemma 21(i), we have that if $\mathsf{Tree}_* = \mathsf{Tree}_{\mathrm{iso}}$, then $T_\sigma = T_{\mathrm{po}}$, because if $(v, w) \in E(T_{\mathrm{po}})$, there does not exist any other path $v \rightsquigarrow w$, and therefore no arc is removed from T_{po} in the construction of T_σ. In the other three categories, still by Lemma 21(i) and its proof, if the arc $([a], [b])$ induced by an arc, say, $(a, b) \in E(T_1)$ is removed because of the existence of a second path $[a] \rightsquigarrow [b]$, then $a, b \in m_1(T_\mu)$, this second path is a Tree_*-path, and all its intermediate nodes are equivalence classes of nodes in $V(T_2) - m_2(V(T_\mu))$. In particular, since the arcs (v, w) removed in the construction of T_σ are such that $v, w \in \ell_1(m_1(V(T_\mu))) = \ell_2(m_2(V(T_\mu)))$ and the Tree_*-paths that make these arcs to be removed have no intermediate node in this set, these paths are not modified in the construction of T_σ, and the arcs can be removed in any order.

Proposition 22. *For every two trees T_1 and T_2, any* Tree$_*$*-sum of T_1 and T_2 is a common* Tree$_*$*-supertree of them.*

Proof. Let T_1 and T_2 be two trees, let T_μ be a largest common Tree$_*$-subtree of them and let $m_1 : T_\mu \to T_1$ and $m_2 : T_\mu \to T_2$ be any Tree$_*$-embeddings. Let T_σ be the Tree$_*$-sum of T_1 and T_2 obtained through m_1 and m_2, and let $\ell_i : V(T_i) \to V(T_\sigma) = (V(T_1) \sqcup V(T_2))/\theta$, $i = 1, 2$, stand for the corresponding restrictions of the quotient mappings.

Every arc removed from the join T_{po} of T_1 and T_2 in the construction of T_σ is subsumed by a path in T_{po}. This implies that, for every $x, y \in V(T_{\mathrm{po}})$, there is a path $x \rightsquigarrow y$ in T_{po} if and only if there is a path $x \rightsquigarrow y$ in T_σ. In particular, since the only nodes in T_{po} than can possibly have no parent are the images of the roots of T_1 and T_2, the same also happens in T_σ.

Now, by Lemma 20(i), if r is the root of T_μ then $m_1(r)$ is the root r_1 of T_1 or $m_2(r)$ is the root r_2 of T_2. If $m_1(r) = r_1$ and $m_2(r) = r_2$, then $[r_1] = [r_2]$ is the only node in T_σ without parent, and every node v in T_{po} (as well as in T_σ, as we said) can be reached from this node through a path: if $v = [a_1]$, with $a_1 \in V(T_1)$, through the image of the path $r_1 \rightsquigarrow a_1$ in T_1, and if $v = [a_2]$, with $a_2 \in V(T_2)$, through the image of the path $r_2 \rightsquigarrow a_2$ in T_2. If, on the contrary, say, $m_1(r) = r_1$ but $m_2(r) \neq r_2$, then $[r_2]$ is the only node in T_σ with no parent and every node in T_σ can be reached from this node through a path: every node of the form $[a_2]$, with $a_2 \in V(T_2)$, through the image of the path $r_2 \rightsquigarrow a_2$ in T_2, and every node of the form $[a_1]$, with $a_1 \in V(T_1)$, through the path obtained by concatenating the image of the path $r_2 \rightsquigarrow m_2(r)$ in T_2 and the image of the path $r_1 \rightsquigarrow a_1$ in T_1.

Thus, T_σ has one, and only one, node without parent, and every other node in T_σ can be reached from it through a path. Moreover, every node in T_σ has in-degree at most 1. Indeed, if a node w has in-degree 2 in T_{po}, say $(v_1, w), (v_2, w) \in E(T_{\mathrm{po}})$, then there will exist some node v and paths $v \rightsquigarrow v_1$ and $v \rightsquigarrow v_2$ with no common intermediate node. But then, by Lemma 21(ii), v will be one of the nodes v_1 or v_2, say $v = v_1$, and then the arc $(v_1, w) \in E(T_{\mathrm{po}})$ is subsumed by the path $v_1 \rightsquigarrow w$ through v_2, and hence it is removed in the construction of T_σ, leaving only the arc (v_2, w). So, every node in T_σ has in-degree at most 1, and it can be reached through a path from the only node without a parent. This proves that T_σ is a tree.

Now we have to prove that $\ell_1 : T_1 \to T_\sigma$ and $\ell_2 : T_2 \to T_\sigma$ are Tree$_*$-embeddings. We shall prove that ℓ_1 is a Tree$_*$-embedding. Recall that the mapping $\ell_1 : V(T_1) \to V(T_{\mathrm{po}}) = V(T_\sigma)$ is injective, and note that, by Lemma 21, if $(a, b) \in E(T_1)$, then there is a Tree$_*$-path in T_σ from $\ell_1(a) = [a]$ to $\ell_1(b) = [b]$ that does not contain any intermediate node in $\ell_1(V(T_1)) \cap \ell_2(V(T_2))$: either the arc $([a], [b])$ induced by the arc in T_1 or the Tree$_*$-path $[a] \rightsquigarrow [b]$ that made this arc to be removed. This shows that ℓ_1 is a Tree$_*$-embedding when Tree$_*$ is Tree$_{\mathrm{iso}}$, Tree$_{\mathrm{hom}}$ or Tree$_{\mathrm{min}}$.

In the case of Tree$_{\mathrm{top}}$, it remains to prove that if $(a, b), (a, c) \in E(T_1)$, then the paths $[a] \rightsquigarrow [b]$ and $[a] \rightsquigarrow [c]$ are divergent. Up to symmetry, there are three possibilities to discuss:

- If the paths $[a] \rightsquigarrow [b]$ and $[a] \rightsquigarrow [c]$ are both arcs, then the injectivity of ℓ_1 implies that they are different and therefore they define divergent paths.
- If the path $[a] \rightsquigarrow [b]$ is an arc and the path $[a] \rightsquigarrow [c]$ has intermediate nodes, and if they did not diverge, $[b]$ would be the first intermediate node of the path $[a] \rightsquigarrow [c]$. But this is impossible, because, since the arc $([a], [c]) \in E(T_{\mathrm{po}})$ has been removed in the construction of T_σ, all intermediate nodes of the path $[a] \rightsquigarrow [c]$ are equivalence classes of nodes in $V(T_2) - m_2(V(T_\mu))$.
- If both paths $[a] \rightsquigarrow [b]$ and $[a] \rightsquigarrow [c]$ have intermediate nodes, then both arcs $([a], [b]), ([a], [c]) \in E(T_{\mathrm{po}})$ have been removed in the construction of T_σ, and therefore there are $x, y, z \in V(T_\mu)$ such that $(x, y), (x, z) \in E(T_\mu)$, $m_1(x) = a$, $m_1(y) = b$, $m_1(z) = c$, and the intermediate nodes of the paths $[a] \rightsquigarrow [b]$ and $[a] \rightsquigarrow [c]$ are the equivalence classes of the intermediate nodes of the paths $m_2(x) \rightsquigarrow m_2(y)$ and $m_2(x) \rightsquigarrow m_2(z)$ in T_2. Now, since m_2 is a topological embedding, these paths $m_2(x) \rightsquigarrow m_2(y)$ and $m_2(x) \rightsquigarrow m_2(z)$ have no common intermediate node. Since ℓ_2 is injective, no image of an intermediate node of the path $m_2(x) \rightsquigarrow m_2(y)$ is equal to the image of an intermediate node of the path $m_2(x) \rightsquigarrow m_2(z)$, and thus the paths $[a] \rightsquigarrow [b]$ and $[a] \rightsquigarrow [c]$ are divergent.

Therefore, if Tree$_*$ = Tree$_{\mathrm{top}}$, ℓ_1 is a topological embedding. □

Theorem 24 below extends the last proposition in the algebraic direction, by showing that Tree$_*$-sums are not only common Tree$_*$-subtrees, but pushouts. In its proof we shall use several times the following technical fact, which we establish first as a lemma.

Lemma 23. *Let* Tree_* *be* $\mathsf{Tree}_{\mathrm{hom}}$, $\mathsf{Tree}_{\mathrm{top}}$ *or* $\mathsf{Tree}_{\min}$. *Let* T_1 *and* T_2 *be trees, let* T_μ *be a largest common* Tree_**-subtree of* T_1 *and* T_2, *let* $m_1 : T_\mu \to T_1$ *and* $m_2 : T_\mu \to T_2$ *be any* Tree_**-embeddings, and let* $f_1 : T_1 \to T$ *and* $f_2 : T_2 \to T$ *be any* Tree_**-embeddings such that* $f_1 \circ m_1 = f_2 \circ m_2$.

There do not exist $x \in V(T_\mu)$, $p \in V(T_1) - m_1(V(T_\mu))$, *and* $q \in V(T_2) - m_2(V(T_\mu))$ *such that* $(m_1(x), p) \in E(T_1)$, $(m_2(x), q) \in E(T_2)$, *and* $f_1(p)$ *and* $f_2(q)$ *are connected by a path.*

Proof. Assume that there exist $x \in V(T_\mu)$, $p \in V(T_1) - m_1(V(T_\mu))$, and $q \in V(T_2) - m_2(V(T_\mu))$ such that $(m_1(x), p) \in E(T_1)$, $(m_2(x), q) \in E(T_2)$, and there is, say, a path $f_2(q) \rightsquigarrow f_1(p)$, in such a way that $f_2(q)$ is an intermediate node in the path $f_1(m_1(x)) \rightsquigarrow f_1(p)$. We shall look for a contradiction.

Under these assumptions, we can enlarge T_μ by adding a new node y, a new arc (x, y), and replacing by a new arc (y, z) every arc (x, z) such that the path $m_1(x) \rightsquigarrow m_1(z)$ in T_1 contains p. It is clear that the graph $\widehat{T_\mu}$ obtained in this way is a tree, strictly larger than T_μ. We can extend m_1 and m_2 to $\widehat{T_\mu}$ by defining $m_1(y) = p$ and $m_2(y) = q$. If we prove that the mappings $m_1 : V(\widehat{T_\mu}) \to V(T_1)$ and $m_2 : V(\widehat{T_\mu}) \to V(T_2)$ defined in this way are Tree_*-embeddings $m_1 : \widehat{T_\mu} \to T_1$ and $m_2 : \widehat{T_\mu} \to T_2$, this will contradict the assumption that T_μ is a largest common Tree_*-subtree of T_1 and T_2.

Now, on the one hand, the arc (x, y) is transformed under m_1 and m_2 into the arcs $(m_1(x), p)$ and $(m_1(x), q)$, respectively. Assume now that $\widehat{T_\mu}$ contains a new arc (y, z). This means that T_μ contained (x, z) and that p is the first intermediate node of the Tree_*-path $m_1(x) \rightsquigarrow m_1(z)$, which does not have any intermediate node in $m_1(V(T_\mu))$. This implies that there exists in T_1 a Tree_*-path without intermediate nodes in $m_1(V(\widehat{T_\mu}))$ from $p = m_1(y)$ to $m_1(z)$. As far as m_2 goes, note that the arc (x, z) in T_μ induces under $f_1 \circ m_1$ a Tree_*-path from $f_1(m_1(x)) = f_2(m_2(x))$ to $f_1(m_1(z)) = f_2(m_2(z))$ that contains $f_1(p)$. This path also contains $f_2(q)$, because this node is contained in the path from $f_1(m_1(x)) = f_2(m_2(x))$ to $f_1(p)$. So, there exists a Tree_*-path $f_2(q) \rightsquigarrow f_2(m_2(z))$, which entails the existence of a Tree_*-path $q \rightsquigarrow m_2(z)$ in T_2. And since this path is actually a piece of the path $m_2(x) \rightsquigarrow m_2(z)$, it has no intermediate node in $m_2(V(T_\mu))$.

This shows that $m_1 : \widehat{T_\mu} \to T_1$ and $m_2 : \widehat{T_\mu} \to T_2$ transform arcs into Tree_*-paths without any intermediate node coming from $\widehat{T_\mu}$, and hence that they are Tree_*-morphisms when Tree_* is $\mathsf{Tree}_{\mathrm{hom}}$ or $\mathsf{Tree}_{\min}$. When $\mathsf{Tree}_* = \mathsf{Tree}_{\mathrm{top}}$, it remains to check that m_1 and m_2 transform pairs of arcs with the same source node into divergent paths. To do it, note first that in this case x has at most one child z such that the path $m_1(x) \rightsquigarrow m_1(z)$ in T_1 contains p, because the paths in T_1 from $m_1(x)$ to the images under m_1 of the children of x diverge. Therefore, the new node y has out-degree at most 1 in $\widehat{T_\mu}$. So, to prove that m_1 and m_2 are topological embeddings, it is enough to check that if y_1 is any child of x in $\widehat{T_\mu}$ other than y, the paths $m_i(x) \rightsquigarrow m_i(y)$ and $m_i(x) \rightsquigarrow m_i(y_1)$ in each T_i diverge. For $i = 1$ it is obvious, because the path $m_1(x) \rightsquigarrow m_1(y)$ is simply the arc $(m_1(x), p)$ and, by assumption, p is not contained in the path $m_1(x) \rightsquigarrow m_1(y_1)$. As far as the case $i = 2$ goes, the path $m_2(x) \rightsquigarrow m_2(y)$ is simply the arc $(m_2(x), q)$, and thus it is enough to check that q is not contained in the path $m_2(x) \rightsquigarrow m_2(y_1)$. But the paths from $f_1(m_1(x)) = f_2(m_2(x))$ to $f_1(p)$ and to $f_1(m_1(y_1)) = f_2(m_2(y_1))$ diverge because f_1 is a topological embedding, and therefore, since $f_2(q)$ is contained in the fist one, it cannot be contained in the second one, which implies that q cannot be contained in the path $m_2(x) \rightsquigarrow m_2(y_1)$. This finishes the proof that, when $\mathsf{Tree}_* = \mathsf{Tree}_{\mathrm{top}}$, m_1 and m_2 are topological embeddings. □

Theorem 24. *Let* T_1 *and* T_2 *be trees, let* T_μ *be a largest common* Tree_**-subtree of* T_1 *and* T_2, *and let* $m_1 : T_\mu \to T_1$ *and* $m_2 : T_\mu \to T_2$ *be any* Tree_**-embeddings.*

Then, the Tree_**-sum* T_σ *of* T_1 *and* T_2 *obtained through* m_1 *and* m_2, *together with the* Tree_**-embeddings* $\ell_1 : T_1 \to T_\sigma$ *and* $\ell_2 : T_2 \to T_\sigma$, *is a pushout in* Tree_* *of* m_1 *and* m_2.

Proof. It is clear that $\ell_1 \circ m_1 = \ell_2 \circ m_2$. Therefore, it remains to prove that T_σ, together with the Tree_*-embeddings $\ell_1 : T_1 \to T_\sigma$ and $\ell_2 : T_2 \to T_\sigma$, satisfies the universal property of pushouts in Tree_*.

So, let $f_1 : T_1 \to T$ and $f_2 : T_2 \to T$ be any Tree_*-embeddings such that $f_1 \circ m_1 = f_2 \circ m_2$. It is well-known that there exists one, and only one, mapping $f : (V(T_1) \sqcup V(T_2))/\theta \to V(T)$ such that $f \circ \ell_1 = f_1$ and $f \circ \ell_2 = f_2$: namely, the one defined by $f([a]) = f_1(a)$ if $a \in V(T_1)$ and $f([a]) = f_2(a)$ if $a \in V(T_2)$. We must prove that this mapping f is a Tree_*-embedding.

Let us prove first that it is injective. Assume that there exist $v, w \in V(T)$, $v \neq w$, such that $f(v) = f(w)$. Since f_1 and f_2 are injective, it is clear that they cannot be classes of nodes of the same tree T_i. Thus, there exist $a \in V(T_1) - m_1(V(T_\mu))$ and $b \in V(T_2) - m_2(V(T_\mu))$ such that $v = [a]$ and $w = [b]$ and $f_1(a) = f_2(b)$.

By Lemma 20(i), the image under some m_i of the root of T_μ is the root of the corresponding T_i. This implies that there exists a path from the image of a node in T_μ to one of these nodes a or b in the corresponding tree. Moreover, if there exists, say, some $x \in V(T_\mu)$ such that there is a path $m_1(x) \rightsquigarrow a$ in T_1, then there is a path from $f_1(m_1(x)) = f_2(m_2(x))$ to $f_1(a) = f_2(b)$ in T, and hence a path $m_2(x) \rightsquigarrow b$ in T_2. By symmetry, if there exists some $x \in V(T_\mu)$ such that there is a path $m_2(x) \rightsquigarrow b$ in T_2, then there is a path $m_1(x) \rightsquigarrow a$ in T_1.

This shows that there exists a node $x_0 \in V(T_\mu)$ such that there exist paths $m_1(x_0) \rightsquigarrow a$ in T_1 and $m_2(x_0) \rightsquigarrow b$ in T_2 without any intermediate node in $m_1(V(T_\mu))$ or $m_2(V(T_\mu))$, respectively. These paths induce, through f_1 and f_2, the same path from $f_1(m_1(x_0)) = f_2(m_2(x_0))$ to $f_1(a) = f_2(b)$ in T (because of the uniqueness of paths in trees). Let now e be the child of $m_1(x_0)$ contained in the path $m_1(x_0) \rightsquigarrow a$ in T_1, and d the child of $m_2(x_0)$ in the path $m_2(x_0) \rightsquigarrow b$ in T_2. Then $f_1(e)$ and $f_2(d)$ are contained in the path from $f_1(m_1(x_0)) = f_2(m_2(x_0))$ to $f_1(a) = f_2(b)$ in T, and hence, they are connected by a path.

When Tree_* is $\mathsf{Tree}_{\mathrm{hom}}$, $\mathsf{Tree}_{\mathrm{top}}$ or $\mathsf{Tree}_{\mathrm{min}}$, Lemma 23 says that this situation is impossible, and therefore f must be injective. In the case when $\mathsf{Tree}_* = \mathsf{Tree}_{\mathrm{iso}}$, since f_1 and f_2 transform arcs into arcs, it must happen that $f_1(e) = f_2(d)$. This allows us to enlarge T_μ, by adding a new node y_0 and a new arc (x_0, y_0): it is clear that the graph $\widehat{T_\mu}$ obtained in this way is a tree. We extend m_1 and m_2 to $\widehat{T_\mu}$ by defining $m_1(y_0) = e$ and $m_2(y_0) = d$. The mappings $m_1 : V(\widehat{T_\mu}) \to V(T_1)$ and $m_2 : V(\widehat{T_\mu}) \to V(T_2)$ defined in this way are isomorphic embeddings $m_1 : \widehat{T_\mu} \to T_1$ and $m_2 : \widehat{T_\mu} \to T_2$. Indeed, they are injective because their restrictions to T_μ are injective and, by assumption, $e \notin m_1(V(\widehat{T_\mu}))$ and $d \notin m_2(V(\widehat{T_\mu}))$, and they transform arcs into arcs because their restrictions to T_μ do so and $(m_i(x_0), m_i(y_0)) \in E(T_i)$ for each $i = 1, 2$. In this way we obtain a common isomorphic subtree of T_1 and T_2 that is strictly larger than T_μ, which yields a contradiction. Therefore, f is also injective in this case.

So, $f : V(T_\sigma) \to V(T)$ is always injective. Now, assume $(v, w) \in T_\sigma$. Then, for some $i = 1, 2$, there exist $a, b \in V(T_i)$ such that $v = [a]$, $w = [b]$, and $(a, b) \in E(T_i)$: to fix ideas, assume that $i = 1$. This implies that there is a Tree_*-path from $f(v) = f_1(a)$ to $f(w) = f_1(b)$ in T. If $\mathsf{Tree}_* = \mathsf{Tree}_{\mathrm{iso}}$, this already proves that f is an isomorphic embedding.

Thus, henceforth, we shall assume that $\mathsf{Tree}_* \neq \mathsf{Tree}_{\mathrm{iso}}$. In this case, we must check that no intermediate node of this Tree_*-path $f(v) \rightsquigarrow f(w)$ belongs to $f(V(T_\sigma)) = f_1(V(T_1)) \cup f_2(V(T_2))$. Now, f_1 being a Tree_*-embedding, we already know that no intermediate node of this path belongs to $f_1(V(T_1))$, and therefore we only have to check that no intermediate node belongs to $f_2(V(T_2))$, either. Before proceeding, note that we have already proved that f sends arcs to Tree_*-paths, and hence that this mapping transforms paths in T_σ into paths in T.

Assume that there is some $c \in V(T_2)$ such that $f_2(c)$ is an intermediate node of the path $f_1(a) \rightsquigarrow f_1(b)$ in T. This prevents the existence of paths $[c] \rightsquigarrow [a]$ or $[b] \rightsquigarrow [c]$ in T_σ: the image of such a path under f would be a path in T that would build up a circuit with the path from $f_1(a) = f([a])$ to $f_2(c) = f([c])$ or from $f_2(c) = f([c])$ to $f_1(b) = f([b])$, respectively, that we already know to exist. Moreover, $c \notin m_2(V(T_\mu))$, because if $c \in m_2(V(T_\mu))$, then $f_2(c) \in f_2(m_2(V(T_\mu))) = f_1(m_1(V(T_\mu))) \subseteq f_1(V(T_1))$.

After excluding these possibilities, we still must discuss several cases:

- $a = m_1(x)$ and $b = m_1(y)$ for some $x, y \in V(T_\mu)$. In this case, by Lemma 4, the existence of an arc from $\ell_2(m_2(x)) = \ell_1(m_1(x)) = [a]$ to $\ell_2(m_2(y)) = \ell_1(m_1(y)) = [b]$ implies the existence of an arc from $m_2(x)$ to $m_2(y)$ in T_2. Since f_2 is a Tree_*-embedding, the path from $f_2(m_2(x)) = f_1(a)$ to $f_2(m_2(y)) = f_1(b)$ does not contain any intermediate node in $f_2(V(T_2))$, which contradicts the existence of c.
- $a = m_1(x)$ for some $x \in V(T_\mu)$, but $b \notin m_1(V(T_\mu))$. In this case, since f_2 is a Tree_*-embedding, the existence of a Tree_*-path $f_2(m_2(x)) = f_1(a) \rightsquigarrow f_2(c)$ in T implies, by Corollary 9, the existence of a Tree_*-path $m_2(x) \rightsquigarrow c$ in T_2. And this path cannot have any intermediate node in $m_2(V(T_\mu))$: any intermediate node in this set would become, under f_2, an intermediate node in $f_2(m_2(V(T_\mu))) = f_1(m_1(V(T_\mu))) \subseteq f_1(V(T_1))$ of the path $f_2(m_2(x)) \rightsquigarrow f_2(c)$. Let d be the child of $m_2(x)$ contained in this path $m_2(x) \rightsquigarrow c$. The path $f_2(m_2(x)) = f_1(a) \rightsquigarrow f_2(c)$ contains $f_2(d)$, and therefore $f_2(d)$ is an intermediate node of the path $f_1(a) \rightsquigarrow f_1(b)$. But then this situation is impossible by Lemma 23.
- $a \notin m_1(V(T_\mu))$. Since, by Lemma 20(i), the image under m_1 or m_2 of the root of T_μ is the root of T_1 or T_2, respectively, we know that there exists some $x \in V(T_\mu)$ such that there is a path $m_1(x) \rightsquigarrow a$ in T_1 or a path $m_2(x) \rightsquigarrow c$ in T_2. It turns

out that the existence of such a path $m_1(x) \rightsquigarrow a$ in T_1 or $m_2(x) \rightsquigarrow c$ in T_2 implies the existence of paths $m_1(x) \rightsquigarrow a$ *and* $m_2(x) \rightsquigarrow c$ in T_1 and T_2, respectively. Indeed, if there exists a path $m_1(x) \rightsquigarrow a$, then there is a path $f_1(m_1(x)) \rightsquigarrow f_1(a)$ in T, which, composed with the path $f_1(a) \rightsquigarrow f_2(c)$, yields a path $f_2(m_2(x)) = f_1(m_1(x)) \rightsquigarrow f_2(c)$, and this path, on its turn, implies a path $m_2(x) \rightsquigarrow c$ in T_2. Conversely, if there exists a path $m_2(x) \rightsquigarrow c$, then there is a path from $f_2(m_2(x))$ to $f_2(c)$ in T. Since there is also a path $f_1(a) \rightsquigarrow f_2(c)$ and $f_2(m_2(x)) = f_1(m_1(x))$ cannot be an intermediate node of the path $f_1(a) \rightsquigarrow f_2(c)$ (because this path does not contain any intermediate node in $f_1(V(T_1))$), it must happen that $f_1(a)$ is intermediate in the path $f_2(m_2(x)) \rightsquigarrow f_2(c)$, that is, that there is a path $f_1(m_1(x)) = f_2(m_2(x)) \rightsquigarrow f_1(a)$ which, finally, implies a path $m_1(x) \rightsquigarrow a$ in T_1.

So, we can take $x \in V(T_\mu)$ such that, on the one hand, there exist paths $m_1(x) \rightsquigarrow a$ and $m_2(x) \rightsquigarrow c$ in T_1 and T_2 and, on the other hand, there do not exist paths $m_1(y) \rightsquigarrow a$ in T_1 or $m_2(y) \rightsquigarrow c$ in T_2 for any child y of it. Let then e be the child of $m_1(x)$ contained in the path $m_1(x) \rightsquigarrow a$ in T_1, and d the child of $m_2(x)$ contained in the path $m_2(x) \rightsquigarrow c$ in T_2. The uniqueness of paths in T implies that the path $f_2(m_2(x)) \rightsquigarrow f_2(c)$, which contains $f_2(d)$, is the concatenation of the path $f_1(m_1(x)) \rightsquigarrow f_1(a)$, which contains $f_1(e)$, and the path $f_1(a) \rightsquigarrow f_2(c)$. Therefore, $f_1(e)$ and $f_2(d)$ are connected by a path. By Lemma 23, this situation cannot happen.

Therefore, f transforms arcs into Tree_*-paths without intermediate nodes in $f(V(T_\sigma))$, and thus it is a Tree_*-embedding when Tree_* is $\mathsf{Tree}_{\mathrm{hom}}$ or $\mathsf{Tree}_{\mathrm{min}}$. This proves the universal property of pushouts, and with it the statement, for these categories. It remains to prove it in $\mathsf{Tree}_{\mathrm{top}}$.

So far, we know that, if we are in $\mathsf{Tree}_{\mathrm{top}}$, then f transforms arcs into paths without intermediate nodes in $f(V(T_\sigma))$. Now we must prove that it transforms arcs with the same source node into divergent paths. So, assume there are arcs (v, w) and (v, u) in T_σ with $w \neq u$.

If these arcs are induced by arcs in the same tree, i.e., if there exist $(a, b), (a, c) \in V(T_i)$, for some $i = 1, 2$, such that $v = [a]$, $w = [b]$ and $u = [c]$, then, since f_i is a topological embedding, the paths from $f(v) = f_i(a)$ to $f(w) = f_i(b)$ and to $f(u) = f_i(c)$ are divergent. Now consider the case when each one of these arcs is induced by an arc in a different tree. In this case, there exist $x \in V(T_\mu)$, $b \in V(T_1)$ and $c \in V(T_2)$ such that, say, $v = [m_1(x)] = [m_2(x)]$, $w = [b]$ and $u = [c]$, and there are arcs $(m_1(x), b) \in E(T_1)$ and $(m_2(x), c) \in E(T_2)$.

If there exists $y \in V(T_\mu)$ such that $m_1(y) = b$, then, by Lemma 4, $(x, y) \in E(T_\mu)$ and hence there exists a path $m_1(x) \rightsquigarrow m_2(y)$ in T_2. But since there is an arc from $[m_2(x)] = [m_1(x)]$ to $[m_2(y)] = [b]$ in T_σ, the path $m_2(x) \rightsquigarrow m_2(y)$ in T_2 must also be an arc (otherwise, it would induce a path in T_{po} that would have made the arc (v, w) to be removed in the construction of T_σ). Therefore, the arc (v, w) is induced by the arc $(m_2(x), m_2(y))$ in T_2, and thus both arcs (v, w) and (v, u) are induced by arcs in T_2 and the paths $f(v) \rightsquigarrow f(w)$ and $f(v) \rightsquigarrow f(u)$ are divergent, as we have just seen. In a similar way, if there exists $y \in V(T_\mu)$ such that $m_2(y) = c$, then both arcs (v, w) and (v, u) are induced by arcs in T_1 and the paths $f(v) \rightsquigarrow f(w)$ and $f(v) \rightsquigarrow f(u)$ are divergent.

Consider finally the case when neither b nor c have a preimage in T_μ. There are two possibilities to discuss:

- If there exists an arc $(x, z) \in V(T_\mu)$ such that b is the first intermediate node of the path $m_1(x) \rightsquigarrow m_1(z)$, then $w = [b]$ is the first intermediate node of the path $[m_1(x)] \rightsquigarrow [m_1(z)]$. In particular, $u = [c]$ does not appear in this last path, which implies that the arc $(m_2(x), c)$ and the path $m_2(x) \rightsquigarrow m_2(z)$ are divergent. Since f_2 is a topological embedding, the paths in T from $f_2(m_2(x))$ to $f_2(c)$ and from $f_2(m_2(x)) = f_1(m_1(x))$ to $f_2(m_2(z)) = f_1(m_1(z))$ are also divergent. Since $f_1(b)$ is contained in this last path, we finally deduce that the paths from $f(v) = f_2(m_2(x)) = f_1(m_1(x))$ to $f(u) = f_1(b)$ and to $f(w) = f_2(c)$ are divergent.

  The case when there exists an arc $(x, z) \in V(T_\mu)$ such that c is the first intermediate node of the path $m_2(x) \rightsquigarrow m_2(z)$ is solved in a similar way.
- If there is no arc (x, z) in T_μ such that b or c are intermediate nodes of the paths $m_1(x) \rightsquigarrow m_1(z)$ or $m_2(x) \rightsquigarrow m_2(z)$, respectively, then we can enlarge T_μ by adding to it a new node y_0 and an arc (x, y_0), and we can extend m_1 and m_2 to this new tree by defining $m_1(y_0) = b$ and $m_2(y_0) = c$, and it is straightforward to prove that in this way we obtain a topological subtree of T_1 and T_2 strictly larger than T_μ, which contradicts the assumption that T_μ is a largest common topological subtree of T_1 and T_2. So, this possibility cannot happen.

This finishes the proof for $\mathsf{Tree}_{\mathrm{top}}$. □

Remark 25. To frame the last result, it is interesting to note that no category Tree_* considered in this paper has all binary pushouts, essentially because the category of sets with injective mappings as morphisms does not have all binary pushouts, either. As a matter of fact, the simplest counter-example does not involve arcs at all. Let S be the empty tree and, for every $i = 1, 2$, let T_i be the tree consisting of a single node $\{a_i\}$ and no arc, and let $m_i : V(S) \to V(T_i)$ be the empty mapping. It is clear that each m_i is a Tree_*-embedding, for every category Tree_*. Now, assume that $m_1 : S \to T_1$ and $m_2 : S \to T_2$ have a pushout $(P, g_1 : T_1 \to P, g_2 : T_2 \to P)$ in Tree_*.

Consider the tree P' consisting of two nodes a_1, a_2 and no arc and the mappings $g'_1 : V(T_1) \to V(P')$ and $g'_2 : V(T_2) \to V(P')$ defined by $g'_1(a_1) = a_1$ and $g'_2(a_2) = a_2$. These mappings are Tree_*-embeddings, for every category Tree_*. Since $g'_1 \circ m_1 = g'_2 \circ m_2$, by the universal property of pushouts there exists a Tree_*-embedding $g' : P \to P'$ such that $g' \circ g_1 = g'_1$ and $g' \circ g_2 = g'_2$: in particular, $g'(g_1(a_1)) = a_1 \neq a_2 = g'(g_2(a_2))$, and therefore $g_1(a_1) \neq g_2(a_2)$.

Consider now the tree P'' consisting of a single node a and no arc and the mappings $g''_1 : V(T_1) \to V(P'')$ and $g''_2 : V(T_2) \to V(P'')$ defined by $g''_1(a_1) = g''_2(a_2) = a$. Again, these mappings are Tree_*-embeddings, for every category Tree_*, and they satisfy that $g''_1 \circ m_1 = g''_2 \circ m_2$. Then, by the universal property of pushouts, there exists a Tree_*-embedding $g'' : P \to P''$ such that $g'' \circ g_1 = g''_1$ and $g'' \circ g_2 = g''_2$. But then $g''(g_1(a_1)) = g''_1(a_1) = a = g''_2(a_2) = g''(g_2(a_2))$, and hence g'' is not injective. Therefore, it cannot be a Tree_*-embedding, which yields a contradiction.

This shows that m_1 and m_2 do not have a pushout in any category Tree_*. Of course, in this case S is not a least common Tree_*-subtree of T_1 and T_2.

5. Largest common subtrees and smallest common supertrees

Let Tree_* still denote any category $\mathsf{Tree}_{\mathrm{iso}}$, $\mathsf{Tree}_{\mathrm{hom}}$, $\mathsf{Tree}_{\mathrm{top}}$ or $\mathsf{Tree}_{\mathrm{min}}$. In this section, we show that the constructions presented in the last two sections can be used to obtain largest common Tree_*-subtrees and smallest common Tree_*-supertrees of pairs of trees. The key will be the following result.

Lemma 26. *Let T_1 and T_2 be two trees, and let T_μ be a largest common Tree_*-subtree of them. For every common Tree_*-supertree T of T_1 and T_2, we have that $|V(T)| \geqslant |V(T_1)| + |V(T_2)| - |V(T_\mu)|$.*

Proof. Propositions 13–16 show that, for every two Tree_*-embeddings $f_1 : T_1 \to T$ and $f_2 : T_2 \to T$, there exists a common Tree_*-subtree T_0 of T_1 and T_2 with set of nodes containing $f_1(V(T_1)) \cap f_2(V(T_2))$: after a relabeling of the nodes (so that f_1 and f_2 are given by inclusions of the sets of nodes), it will be the intersection T_p of T_1 and T_2 in $\mathsf{Tree}_{\mathrm{iso}}$, $\mathsf{Tree}_{\mathrm{hom}}$, and $\mathsf{Tree}_{\mathrm{top}}$, and its one-node extension $\widetilde{T}_p$ in $\mathsf{Tree}_{\mathrm{min}}$. Then,

$$|f_1(V(T_1)) \cap f_2(V(T_2))| \leqslant |V(T_0)| \leqslant |V(T_\mu)|$$

and hence,

$$\begin{aligned} |V(T)| &\geqslant |f_1(V(T_1)) \cup f_2(V(T_2))| \\ &= |f_1(V(T_1))| + |f_2(V(T_2))| - |f_1(V(T_1)) \cap f_2(V(T_2))| \\ &\geqslant |V(T_1)| + |V(T_2)| - |V(T_0)| \\ &\geqslant |V(T_1)| + |V(T_2)| - |V(T_\mu)|, \end{aligned}$$

as we claimed. □

Theorem 27. *For every pair of trees T_1 and T_2, any Tree_*-sum of T_1 and T_2 is a smallest common Tree_*-supertree of them.*

Proof. By Proposition 22, any Tree_*-sum T_σ of T_1 and T_2 is a common Tree_*-supertree of them, and by construction

$$|V(T_\sigma)| = |V(T_1)| + |V(T_2)| - |V(T_\mu)|,$$

for some largest common Tree_*-subtree T_μ of them. Thus, T_σ achieves the lower bound established in Lemma 26 for common Tree_*-supertrees of T_1 and T_2, which implies that it is a smallest common Tree_*-supertree of them. □

Theorem 28. *For every two trees T_1 and T_2, any intersection of T_1 and T_2 obtained through Tree_*-embeddings into a smallest common Tree_*-supertree of them is a largest common Tree_*-subtree of T_1 and T_2.*

Proof. Let T_1 and T_2 be two trees, let T'_σ be a smallest common Tree$_*$-supertree of T_1 and T_2, let $p_1 : T_1 \to T'_\sigma$ and $p_2 : T_2 \to T'_\sigma$ be any Tree$_*$-embeddings, and let T'_p be any common Tree$_*$-subtree of T_1 and T_2 obtained by expanding the intersection T_p of T_1 and T_2 obtained through p_1 and p_2, which exists by Propositions 13–16.

Now, by Theorem 27 we have that, for any largest common Tree$_*$-subtree T_μ of T_1 and T_2,

$$|V(T'_\sigma)| = |V(T_1)| + |V(T_2)| - |V(T_\mu)|$$

and we know that

$$|p_1(V(T_1)) \cap p_2(V(T_2))| \leqslant |V(T'_p)| \leqslant |V(T_\mu)|.$$

Then

$$\begin{aligned} |V(T_1)| + |V(T_2)| - |V(T_\mu)| &= |V(T'_\sigma)| \geqslant |p_1(V(T_1)) \cup p_2(V(T_2))| \\ &= |p_1(V(T_1))| + |p_2(V(T_2))| - |p_1(V(T_1)) \cap p_2(V(T_2))| \\ &\geqslant |V(T_1)| + |V(T_2)| - |V(T'_p)| \\ &\geqslant |V(T_1)| + |V(T_2)| - |V(T_\mu)|. \end{aligned}$$

This implies that $|V(T'_p)| = |V(T_\mu)| = |p_1(V(T_1)) \cap p_2(V(T_2))|$. From these equalities we deduce, on the one hand, that T'_p is also a largest common Tree$_*$-subtree of T_1 and T_2, and on the other hand, that $V(T'_p) = p_1(V(T_1)) \cap p_2(V(T_2))$, i.e., that $T'_p = T_p$, as we claimed. □

Thus, for every pair of trees T_1 and T_2, the pushout in Tree$_*$ of any Tree$_*$-embeddings from a largest common Tree$_*$-subtree of them yields a smallest common Tree$_*$-supertree of them, and the pullback in Tree$_*$ of any Tree$_*$-embeddings into a smallest common Tree$_*$-supertree of them yields a largest common Tree$_*$-subtree of them. Moreover, all smallest common Tree$_*$-supertrees and all largest common Tree$_*$-subtrees are obtained in this way up to isomorphisms, as the following corollaries show.

Corollary 29. *Every smallest common* Tree$_*$*-supertree of a pair of trees* T_1 *and* T_2 *is, up to an isomorphism, the* Tree$_*$*-sum of* T_1 *and* T_2 *obtained through the embeddings of a largest common* Tree$_*$*-subtree into them.*

Proof. Let T_1 and T_2 be two trees, let T'_σ be a smallest common Tree$_*$-supertree of T_1 and T_2 and let $p_1 : T_1 \to T'_\sigma$ and $p_2 : T_2 \to T'_\sigma$ be any Tree$_*$-embeddings. By Theorem 28, the intersection T_p of T_1 and T_2 obtained through p_1 and p_2, together with the corresponding inclusions $\imath_1 : T_p \to T_1$ and $\imath_2 : T_p \to T_2$, is a largest common Tree$_*$-subtree of T_1 and T_2. Let now T_σ, together with $m_1 : T_1 \to T_\sigma$ and $m_2 : T_2 \to T_\sigma$, be the sum of T_1 and T_2 obtained through $\imath_1$ and $\imath_2$. By Theorem 27, T_σ is a smallest common Tree$_*$-supertree of T_1 and T_2, and by Theorem 24, $(T_\sigma, m_1 : T_1 \to T_\sigma, m_2 : T_2 \to T_\sigma)$ is a pushout of $\imath_1 : T_p \to T_1$ and $\imath_2 : T_p \to T_2$ in Tree$_*$. Since $p_1 \circ \imath_1 = p_2 \circ \imath_2$, by the universal property of pushouts there exists a Tree$_*$-embedding $p : T_\sigma \to T'_\sigma$ such that $p \circ m_1 = p_1$ and $p \circ m_2 = p_2$. Now, T_σ and T'_σ have the same size, because they are both smallest common Tree$_*$-supertrees of T_1 and T_2. Therefore, $p : T_\sigma \to T'_\sigma$ is bijective, and thus an isomorphism by Lemma 6. □

A similar argument, which we leave to the reader, proves also the following result.

Corollary 30. *Every largest common* Tree$_*$*-supertree of a pair of trees* T_1 *and* T_2 *is, up to an isomorphism, the intersection of* T_1 *and* T_2 *obtained through their embeddings into a smallest common* Tree$_*$*-supertree.*

Corollary 31. *The problems of finding a largest common* Tree$_*$*-subtree and a smallest common* Tree$_*$*-supertree of two trees, in each case together with a pair of witness* Tree$_*$*-embeddings, are reducible to each other in time linear in the size of the trees.*

Proof. Given two trees T_1 and T_2, if we know a largest common Tree$_*$-subtree T_μ of them, together with a pair of witness Tree$_*$-embeddings $m_1 : T_\mu \to T_1$ and $m_2 : T_\mu \to T_2$, then the construction of the pushout

$$(T_\sigma, \ell_1 : T_1 \to T_\sigma, \ell_2 : T_2 \to T_\sigma)$$

of m_1 and m_2 described in Theorem 24 gives a smallest common Tree$_*$-supertree of T_1 and T_2, and this construction can be obtained in time linear in the size of T_1 and T_2, as follows.

First, make copies T_1' and T_2' of T_1 and T_2, with $\ell_1 : T_1 \to T_1'$ and $\ell_2 : T_2 \to T_2'$ identity mappings. Second, sum up T_1' and T_2' into a graph T_σ. Third, for each $a \in V(T_\mu)$, merge nodes $\ell_1(m_1(a))$ and $\ell_2(m_2(a))$, and remove all parallel arcs.

Next, remove from T_σ all arcs subsumed by paths, as follows. For each node $y \in V(T_\sigma)$ of in-degree 2, let $x, x' \in V(T_\sigma)$ be the source nodes of the two arcs coming into y. Now, perform a simultaneous traversal of the paths of arcs coming into x and x', until reaching node x' along the first path or x along the second path. The simultaneous traversal of incoming paths may stop along either path, but continue along the other one, because a node of in-degree 0 or in-degree 2 is reached. Finally, remove from T_σ either arc (x', y), if node x' was reached along the first path, or arc (x, y), if node x was reached along the second path.

Conversely, if we know a smallest common Tree$_*$-supertree T of T_1 and T_2, together with a pair of witness Tree$_*$-embeddings $f_1 : T_1 \to T$ and $f_2 : T_2 \to T$, then, by Theorem 28, the pullback

$$(T_p, \imath_1 : T_p \to T_1, \imath_2 : T_p \to T_2)$$

of f_1 and f_2 described in Section 3 yields a largest common Tree$_*$-subtree of T_1 and T_2, and this construction can also be obtained in time linear in the size of T_1 and T_2, as follows.

First, make a copy T_p of T, with $g : T \to T_p$ the identity mapping. Second, for each $a \in V(T_1)$, mark $g(f_1(a))$ in T_p. Third, for each $a \in V(T_2)$, if $g(f_2(a))$ is already marked in T_p, double-mark it. Next, for each node of T_p which is not double-marked, add a new arc from its parent (if any) to each of its children (if any) in T_p, and remove the node not double-marked. Finally, set mappings $\imath_i : T_p \to T_i$ for $i = 1, 2$, as follows: for each $a \in V(T_i)$, if $g(f_i(a))$ is defined, set $\imath_i(g(f_i(a))) = a$. □

6. Conclusion

Subtree isomorphism and the related problems of largest common subtree and smallest common supertree belong to the most widely used techniques for comparing tree-structured data, with practical applications in combinatorial pattern matching, pattern recognition, chemical structure search, computational molecular biology, and other areas of engineering and life sciences. Four different embedding relations are of interest in these application areas: isomorphic, homeomorphic, topological, and minor embeddings.

The complexity of the largest common subtree problem and the smallest common supertree problem under these embedding relations is already settled: they are polynomial-time solvable for isomorphic, homeomorphic, and topological embeddings, and they are NP-complete for minor embeddings. Moreover, efficient algorithms are known for largest common subtree under isomorphic, homeomorphic, and topological embeddings, and for smallest common supertree under isomorphic and topological embeddings, and an exponential algorithm is known for largest common subtree under minor embeddings.

In this paper, we have established the relationship between the largest common subtree and the smallest common supertree of two trees by means of simple constructions, which allow one to obtain the largest common subtree from the smallest common supertree, and vice versa. We have given these constructions for isomorphic, homeomorphic, topological, and minor embeddings, and have shown their implementation in time linear in the size of the trees. In doing so, we have filled the gap by providing a simple extension of previous largest common subtree algorithms for solving the smallest common supertree problem, in particular under homeomorphic and minor embeddings for which no algorithm has been known previously.

Beside the practical interest of these extensions to previous algorithms, we have provided a unified algebraic construction showing the relation between largest common subtrees and smallest common supertrees for the four different embedding problems studied in the literature: isomorphic, homeomorphic, topological, and minor embeddings. The unified construction shows that smallest common supertrees are pushouts and largest common subtrees are pullbacks.

Acknowledgements

The authors acknowledge with thanks the anonymous referees, whose comments, suggestions and criticism have led to a substantial improvement of this paper.

References

[1] A. Amir, D. Keselman, Maximum agreement subtree in a set of evolutionary trees: metrics and efficient algorithms, SIAM J. Comput. 26 (6) (1997) 1656–1669.
[2] K.F. Aoki, A. Yamaguchi, Y. Okuno, T. Akutsu, N. Ueda, M. Kanehisa, H. Mamitsuka, Efficient tree-matching methods for accurate carbohydrate database queries, in: Proc. 14th Internat. Conf. Genome Informatics, Universal Academic Press, New York, 2003, pp. 134–143.
[3] P.J. Artymiuk, R.V. Spriggs, P. Willett, Graph theoretic methods for the analysis of structural relationships in biological macromolecules, J. Amer. Soc. Inform. Sci. Technol. 56 (5) (2005) 518–528.
[4] J.M. Barnard, Substructure searching methods: old and new, J. Chem. Inform. Comput. Sci. 33 (1993) 532–538.
[5] M.-J. Chung, $O(n^{2.5})$ time algorithms for the subgraph homeomorphism problem on trees, J. Algorithms 8 (1) (1987) 106–112.
[6] R. Cole, M. Farach-Colton, R. Hariharan, T.M. Przytycka, M. Thorup, An $O(n \log n)$ algorithm for the maximum agreement subtree problem for binary trees, SIAM J. Comput. 30 (5) (2000) 1385–1404.
[7] D. Conte, P. Foggia, C. Sansone, M. Vento, Thirty years of graph matching in pattern recognition, Internat. J. Pattern Recognition Artificial Intell. 18 (3) (2004) 265–298.
[8] A. Dessmark, A. Lingas, A. Proskurowski, Faster algorithms for subgraph isomorphism of k-connected partial k-trees, Algorithmica 27 (1) (2000) 337–347.
[9] S. Dulucq, L. Tichit, RNA secondary structure comparison: exact analysis of the Zhang-Shasha tree edit algorithm, Theoret. Comput. Sci. 306 (1–3) (2003) 471–484.
[10] M.-L. Fernández, G. Valiente, A graph distance measure combining maximum common subgraph and minimum common supergraph, Pattern Recognition Lett. 22 (6–7) (2001) 753–758.
[11] V.J. Gillet, P. Willett, J. Bradshaw, Similarity searching using reduced graphs, J. Chem. Inform. Comput. Sci. 43 (2) (2003) 338–345.
[12] A. Gupta, N. Nishimura, Finding largest subtrees and smallest supertrees, Algorithmica 21 (2) (1998) 183–210.
[13] J. Jansson, A. Lingas, A fast algorithm for optimal alignment between similar ordered trees, Fund. Inform. 56 (1–2) (2003) 105–120.
[14] T. Jiang, G. Lin, B. Ma, K. Zhang, The longest common subsequence problem for arc-annotated sequences, J. Discrete Algorithms 2 (2) (2004) 257–270.
[15] T. Jiang, L. Wang, K. Zhang, Alignment of trees—an alternative to tree edit, Theoret. Comput. Sci. 143 (1) (1995) 137–148.
[16] P. Kilpeläinen, H. Mannila, Ordered and unordered tree inclusion, SIAM J. Comput. 24 (2) (1995) 340–356.
[17] J. Matoušek, R. Thomas, On the complexity of finding isomorphisms and other morphisms for partial k-trees, Discrete Math. 108 (1–3) (1992) 343–364.
[18] N. Nishimura, P. Ragde, D.M. Thilikos, Finding smallest supertrees under minor containment, Internat. J. Foundations Comput. Sci. 11 (3) (2000) 445–465.
[19] R.Y. Pinter, O. Rokhlenko, D. Tsur, M. Ziv-Ukelson, Approximate labelled subtree homeomorphism, in: Proc. 15th Annu. Symp. Combinatorial Pattern Matching, Lecture Notes in Computer Science, Vol. 3109, Springer, Berlin, 2004, pp. 55–69.
[20] R.Y. Pinter, O. Rokhlenko, E. Yeger-Lotem, M. Ziv-Ukelson, Alignment of metabolic pathways, Bioinformatics 21 (16) (2005) 3401–3408.
[21] R. Shamir, D. Tsur, Faster subtree isomorphism, J. Algorithms 33 (2) (1999) 267–280.
[22] D. Shasha, J.T.-L. Wang, K. Zhang, F.Y. Shih, Exact and approximate algorithms for unordered tree matching, IEEE Trans. Systems, Man Cybernat. 24 (4) (1994) 668–678.
[23] D. Shasha, K. Zhang, Fast algorithms for the unit cost editing distance between trees, J. Algorithms 11 (4) (1990) 581–621.
[24] M.A. Steel, T. Warnow, Kaikoura tree theorems: computing the maximum agreement subtree, Inform. Process. Lett. 48 (2) (1993) 77–82.
[25] A. Torsello, D.H. Rowe, M. Pelillo, Polynomial-time metrics for attributed trees, IEEE Trans. Pattern Anal. Mach. Intell. 27 (7) (2005) 1087–1099.
[26] G. Valiente, Algorithms on Trees and Graphs, Springer, Berlin, 2002.
[27] G. Valiente, Constrained tree inclusion, in: Proc. 14th Annu. Symp. Combinatorial Pattern Matching, Lecture Notes in Computer Science, Vol. 2676, Springer, Berlin, 2003, pp. 361–371.
[28] G. Valiente, Constrained tree inclusion, J. Discrete Algorithms 3 (2–4) (2005) 431–447.
[29] L. Wang, J. Zhao, Parametric alignment of ordered trees, Bioinformatics 19 (17) (2003) 2237–2245.
[30] K. Zhang, RNA structure comparison and alignment, in: J.T.-L. Wang, M.J. Zaki, H. Toivonen, D. Shasha (Eds.), Data Mining in Bioinformatics, Springer, Berlin, 2005, pp. 59–81.
[31] K. Zhang, D. Shasha, Simple fast algorithms for the editing distance between trees and related problems, SIAM J. Comput. 18 (6) (1989) 1245–1262.