

An edit distance between quotiented trees

Pascal Ferraro * Christophe Godin*

Submitted for publication to Algorithmica on August 2000,
Revised on december 2001, Accepted on August 2002

Abstract. In this paper we propose a dynamic programming algorithm to compare two quotiented trees using a constrained edit distance. A quotiented tree is a tree defined with an additional equivalent relation on vertices and such that the quotient graph is also a tree. The core of the method relies on an adaptation of an algorithm recently proposed by Zhang for comparing unordered rooted trees. This method is currently being used in plant architecture modelling to quantify different types of variability between plants represented by quotiented trees.

Key Words. quotiented tree, edit distance matching, dynamic programming.

1 Introduction

In the early seventies, Wagner and Fisher proposed an algorithm which computes the distance between two strings of characters as the minimum cost sequence of elementary operations needed to transform one of the strings into the other [1]. Given two strings of characters A and B of respective lengths N_A and N_B , a set of elementary operators on strings, called edit operations, and a cost associated with each edit operation, Wagner and Fisher defined a distance between two strings as the cost of the sequence of edit operations that transforms A into B with minimum cost. The Wagner and Fisher distance makes use of the dynamic programming principle to achieve an algorithm with linear complexity, *i.e.* in $O(N_A.N_B)$. Selkow [2], then Tai and Lu [3, 4] generalized this approach, based on edit operations, to define and compute metrics on labelled ordered trees. These algorithms have been used over recent decades in computer science and in various applied fields, such as evolutionary biology [5], chemistry [6] and molecular biology [7].

Zhang [8] extended these dynamic programming-based algorithms to define a distance for unordered labelled trees. In unordered trees, no ordering is considered for the set of sons of any vertex. This algorithm has recently been applied in plant modelling applications to compute a distance between individual plants whose topology is represented by unordered trees [9]. However, to take account of the multiscale nature of plant structures [10], plants are currently represented by quotiented trees [10]. A quotiented tree is a tree with an equivalence relation defined on the set of vertices, and such that the resulting quotient graph is also a tree. A

*Plant Modelling Program - CIRAD TA 40/PS2 - Boulevard de la Lironde - 34398 Montpellier Cedex 5, France - email ferraro|godin@cirad.fr

quotiented tree can thus be considered as an autosimilar structure represented by trees on two different scales.

In this paper, we define a distance between quotiented trees based on the computation of an optimal sequence of edit operations that preserves equivalence relations on tree vertices. In section 2, basic definitions concerning trees and quotiented trees are introduced. In section 3, we first recall how sequences of edit operations can be modelled using mappings between tree vertices [3]. Zhang's algorithm is then presented and reformulated in terms of recursive relations between sets of mappings in order to prepare its extension to quotiented trees, carried out in section 4. The properties of mappings between quotiented trees, i.e. preserving equivalence relations and called valid edit distance mappings, are then studied and leads to new recursive equations. Similarly to Zhang's algorithm, we show that these equations contain terms that can be computed as particular minimum cost maximum flow problems. Finally, a dynamic programming algorithm that computes a structural distance between two quotiented trees in polynomial time is depicted.

2 Definitions and notations

A *finite directed graph* (or simply a *graph*) is a pair (V, E) where V denotes a finite set of vertices and $E \subseteq V \times V$ denotes a finite set of edges. The number of vertices of a graph G is denoted by $|G|$. If $e = (x, y)$ is an edge in E , x and y are incident with e . Vertex x is called a *father* of y and y is a *son* of x . The set of sons of a vertex is denoted by $son(v)$ which is of size n_v and $\deg(G) = \max_{v \in V} \{n_v\}$. For every k in $\{1, \dots, n_v\}$, v_k denotes a son of v . A *path* (resp. a *chain*) from x_1 to x_n is a sequence of vertices (x_1, x_2, \dots, x_n) such that for any two consecutive vertices $\{x_i, x_{i+1}\}$ of the sequence, (x_i, x_{i+1}) is an edge (resp. either (x_i, x_{i+1}) or (x_{i+1}, x_i) is an edge). Vertex v is an ancestor of vertex w - and reciprocally w is a descendant of v - if a path exists from v to w . The set of descendants of v is denoted by $V[v]$ and contains v itself. A *cycle* is a non-empty path from one vertex to itself. A graph with no cycle is called a *directed acyclic graph*. A sub-graph of a graph $G = (V, E)$ is a graph $G' = (V', E')$ such that $V' \subseteq V$ and $E' \subseteq E$. This is denoted by $G' \subseteq G$. Two vertices of a graph are *connected* if a chain exists between them. A graph is connected if any pair of vertices are connected. The *connected components* of a graph are the maximum (for graph inclusion) connected subgraphs of this graph.

The ancestor relationship on a directed acyclic graph is a partial ordering relation on the set of vertices denoted by \leq . A tree is a connected graph such that there exists a unique vertex, called the root, which has no father, and any vertex different from the root is the son of exactly one vertex. A tree T rooted in v is denoted by $T[v]$. A tree contains no cycle. In a tree, the set of common ancestors of any two vertices x and y obviously contains at least the root vertex and is a totally ordered set (with respect to the ancestor relationship). The maximum element of this set is called the *greatest common ancestor* and is denoted by $x \wedge y$. If S is any set of vertices of a tree, $\bigwedge_{x \in S}$ denotes the greatest common ancestor of all the vertices in S . The graph $\theta = (\emptyset, \emptyset)$ is called the empty tree. An unordered tree is a tree for which no ordering distinction is made among the sons of any vertex. A *sub-tree* is a connected sub-graph of a tree. If x is any vertex of tree $T[v]$, $T[x] = (V[x], E[x])$ denotes the maximum sub-tree of $T[v]$ rooted in x . A *forest* is a graph whose connected components are trees. If x is any vertex of

tree $T[v]$, $F[x]$ denotes the *forest rooted in x* , i.e. obtained from $T[x]$ by removing the root x and all the edges incident with x . A forest rooted in a vertex is thus defined as a set of tree and a given vertex x , therefore there is never equality between a tree and a forest, even if x has only one son. Suppose that x is the only son of y , then by definition $T[x] \neq F[y]$ and $T[x] \subset F[y]$. In the following the term forest will be used to designate a forest. The set of all sub-trees and forests rooted in a vertex of $T[v]$ is denoted by $\mathcal{S}(v) = \{S[x] \mid x \in V[v] \text{ and } S[x] = T[x] \text{ or } S[x] = F[x]\}$. Suppose that x is the only son of y , then by definition $T[x] \neq F[y]$ and $T[x] \subset F[y]$.

A labeled graph is a graph (V, E) together with a mapping α which associates a label from a finite (or infinite) set of labels $\Sigma = \{a, b, c, \dots\}$, with each vertex in V . We assume in the sequel that a distance d is defined on Σ . d enables us to define a distance between any two vertices x and y of labelled graphs: $d(x, y) = d(\alpha(x), \alpha(y))$.

A quotiented graph H is a 3-uple (G, W, π) where $G = (V, E)$ is a directed graph called the *support* of H , W is a set of vertices and π is a surjective mapping from V to W . For any vertex x in V , the vertex $\pi(x)$ is called the *complex* of x and reciprocally x is a *component* of $\pi(x)$. $\pi^{-1}(z)$ denotes the set of components of a vertex z of W and if x is a vertex of V , $\Pi(x)$ denotes the set $\pi^{-1}(\pi(x))$ of components of $\pi(x)$. The size of $\Pi(x)$ is denoted by $|\Pi(x)|$ and $\deg_{\pi}(H) = \max_{x \in V} \{|\Pi(x)|\}$. The function π induces a partition Π_H on V : $\Pi_H = \{\pi^{-1}(z) \mid z \in W\}$. The *quotient graph* $\mathcal{Q}(H)$ associated with H is the graph (W, E_{π}) such that:

$$\forall (x, y) \in E, \quad (\pi(x), \pi(y)) \in E_{\pi} \Leftrightarrow \pi(x) \neq \pi(y)$$

Quotiented graphs whose support and quotient graphs are trees are called *quotiented trees*. Let $H = (G, W, \pi)$ be a quotiented graph with support graph $G = (V, E)$ which is either a tree or a forest. Let $x \in V$, then $H[x]$ denotes the quotiented graph $(G[x], W[\pi(x)], \pi_{/x})$ where $G[x]$ is the sub-tree or a forest of G rooted in x , $W[\pi(x)]$ is the set of vertices of the sub-tree of $\mathcal{Q}(H)$ rooted in $\pi(x)$ and $\pi_{/x}$ is the restriction of π to $V[x]$. If $G[x]$ is a tree, $H[x]$ is a quotiented tree.

3 Distance between unordered tree graphs

3.1 EDMs

The tree-to-tree correction problem [11] consists in determining the distance between two trees measured by the minimum cost of the sequence of edit operations needed to transform one tree into the other. Based on definitions established by Wagner and Fisher [1], Tai [3] and Selkow [2], Zhang [12, 8] uses three edit operations: substitution, deletion and insertion.

- Substituting a vertex x means changing the label of x ;
- Deleting a vertex x means making the sons of x the sons of the father of x and removing x ;
- Inserting a vertex x means that x becomes a son of a vertex y and a subset of sons of y become the set of sons of x (insertion is the complement of deletion).

In order to characterize the effect of a sequence of edit operations on a tree, Tai [3] introduced a structure called *edit distance mapping (EDM)*. An EDM from a tree $T_1[v]$ to a tree $T_2[w]$ is a *partial mapping* from $V_1[v]$ to $V_2[w]$, based on the notion of *trace* between sequences [1]. Intuitively, an EDM is a description of how a sequence of edit operations transforms $T_1[v]$ into $T_2[w]$, ignoring the order in which the edit operations are applied. The relation between edit operations and EDMs is made explicit in [3], [9].

Definition 1 Let $T_1[v] = (V_1[v], E_1[v])$ and $T_2[w] = (V_2[w], E_2[w])$ be two trees, then an EDM M from $T_1[v]$ to $T_2[w]$ is a set of ordered pairs of vertices (z, t) of $V_1[v] \times V_2[w]$.

We recall that $T_1[v]$ and $T_2[w]$ are trees respectively rooted in v and w . The same definition is used to define an EDM from a forest $F_1[v]$ to a forest $F_2[w]$. The set of EDMs from $T_1[v]$ to $T_2[w]$ is denoted by $\mathcal{EDM}(v, w)$.

Let M be an EDM of $\mathcal{EDM}(v, w)$, by convention in any pair (z, t) of M , z is called an *image* of t by M and reciprocally t is called an *image* of z by M . Similarly, if z does not appear in a pair of M , we say that z *has no image*.

Let x be a vertex of $T_1[v]$ and let y be a vertex of $T_2[w]$, $M_{x/y}$ (resp. $M_{y/x}$) denotes the set of vertices of $T_1[x]$ (resp. $T_2[y]$) which have an image by M in $T_2[y]$ and $T_1[x]$:

$$\begin{aligned} M_{x/y} &= \{z_1 \in V_1[x] \mid \exists z_2 \in V_2[y]; (z_1, z_2) \in M\} \\ M_{y/x} &= \{z_2 \in V_2[y] \mid \exists z_1 \in V_1[x]; (z_1, z_2) \in M\} \\ \overline{M_{x/y}} &= V_1[x] / M_{x/y} \\ \overline{M_{y/x}} &= V_2[y] / M_{y/x} \end{aligned}$$

$M_{x/w}$ and $M_{y/v}$ will be denoted by M_x and M_y when no confusion is possible.

Let \hat{x} be the greatest common ancestor of the vertices of $T_2[w]$ which have an image in $T_1[x]$:

$$\hat{x} = \bigwedge_{y \in M_{w/x}} \{y\}$$

Similarly, for any y in $T_2[w]$, \hat{y} defines a vertex in $T_1[v]$.

Note that when $M_{x/w}$ is empty, then \hat{x} is not defined. The function “ $\hat{}$ ”

Furthermore, remark that if \hat{x} exists, \hat{x} is in $T_2[w]$ while x is in $T_1[v]$ and that \hat{x} is not necessarily an image of a vertex in $T_1[x]$.

This function can be used to associate a mapping \widehat{M}_{12} from $\mathcal{S}_1(v)$ to $\mathcal{S}_2(w)$ with any M in $\mathcal{EDM}(v, w)$.

Definition 2 Let M be an EDM from $T_1[v]$ to $T_2[w]$. \widehat{M}_{12} is a mapping from $\mathcal{S}_1(v)$ to $\mathcal{S}_2(w)$ such that:

$$\begin{aligned} \widehat{M}_{12} : \mathcal{S}_1(v) &\rightarrow \mathcal{S}_2(w) \\ S_1[x] &\mapsto \widehat{M}_{12}(S_1[x]) = \begin{cases} \theta & \text{if } M_{x/w} = \emptyset \\ T_2[\hat{x}] & \text{if } \hat{x} \in M_w \\ F_2[\hat{x}] & \text{otherwise} \end{cases} \end{aligned}$$

Note that in this definition when \widehat{x} is not defined, then the image of $S_1[x]$ is the empty tree.

Symetrically, a mapping \widehat{M}_{21} can be defined from $\mathcal{S}_2(w)$ to $\mathcal{S}_1(v)$. When no confusion is possible, \widehat{M}_{12} and \widehat{M}_{21} are simply denoted by \widehat{M} . Figure 1 illustrates the image by \widehat{M} of a tree and a forest. \widehat{M} gives a high-level interpretation of EDMs: whereas M expresses a relationship between vertices, \widehat{M} expresses a corresponding relationship between trees (or forests) rooted in these vertices. \widehat{M} has the following important property:

Proposition 1 \widehat{M} is an increasing mapping, i.e. for any x and y in $V_1[v]$, and for any $S_1[x]$ and $S_1[y]$ in $\mathcal{S}_1(v)$:

$$S_1[x] \subseteq S_1[y] \Rightarrow \widehat{M}(S_1[x]) \subseteq \widehat{M}(S_1[y])$$

The reciprocal of proposition 1, which is not true in general (fig. 1), is true with additional assumptions (see section 3.4). \widehat{M} enables us to work at the graph level, i.e. to express the algorithm properties in terms of relations between trees (or forests), while the original formulation of Zhang's algorithm was performed at the vertex level. This formulation will be used in the sequel to extend the original comparison algorithm to quotiented trees.

3.2 Cost of EDMs

According to the definition of the elementary cost between vertices, a cost is assigned to each EDM M from $T_1[v]$ to $T_2[w]$:

$$\gamma(M) = \sum_{(x,y) \in M} d(x,y)$$

A first dissimilarity measure can be defined as the minimum cost of an EDM from $T_1[v]$ to $T_2[w]$. However, this dissimilarity does not take account of vertices which have no image, and two trees of different sizes could thus be considered as similar. To account for these vertices, the cost of sets \overline{M}_v and \overline{M}_w can be added to the cost of an EDM. Let us define a symbol λ not in Σ and extend the distance d so that d is a distance over $\Sigma \cup \{\lambda\}$. The cost of inserting or deleting a vertex x is denoted $d(x, \lambda)$ and is defined as $d(\alpha(x), \lambda)$.

The cost $\Gamma_{v,w}(M)$ of an EDM M from $T_1[v]$ to $T_2[w]$ is thus defined as:

$$\Gamma_{v,w}(M) = \gamma(M) + \gamma(\overline{M}_v) + \gamma(\overline{M}_w)$$

A dissimilarity measure between a tree $T_1[v]$ and a tree $T_2[w]$ can thus be defined as the minimum cost of an EDM from $T_1[v]$ to $T_2[w]$:

$$D(T_1[v], T_2[w]) = \min_{M \in \mathcal{EDM}(v,w)} \{\Gamma_{v,w}(M)\}$$

When d is a distance, D is also a distance [13].

3.3 Valid EDMs

In the following, we consider an analogous dissimilarity measure, restricted to EDMs preserving structural properties of the mapped trees. These are called valid EDMs:

Definition 3 (valid EDM) *Let $T_1[v] = (V_1[v], E_1[v])$ and $T_2[w] = (V_2[w], E_2[w])$ be two trees, a valid EDM M from $T_1[v]$ to $T_2[w]$ is a set of ordered pairs of vertices $(x, y) \in V_1[v] \times V_2[w]$ satisfying the constraints:*

$$\forall (x_1, x_2), (y_1, y_2) \in M : \quad (1)$$

$$x_1 = y_1 \Leftrightarrow x_2 = y_2 \quad (1)$$

$$x_1 \leq y_1 \Leftrightarrow x_2 \leq y_2 \quad (2)$$

Sets of valid EDMs are denoted by $\mathcal{T}(v, w)$. A valid EDM between two rooted forests $F_1[v]$ and $F_2[w]$ is defined in a similar manner. The set of EDMs from $F_1[v]$ to $F_2[w]$ is denoted by $\mathcal{F}(v, w)$.

A new dissimilarity measure between two tree graphs $T_1[v]$ and $T_2[w]$ can be defined as an optimization problem:

Problem 1 *Find $\Gamma_{v,w}(M)$ minimum, such that M is a valid EDM from $T_1[v]$ to $T_2[w]$ satisfying constraints (1) and (2):*

$$D(T_1[v], T_2[w]) = \min_{M \in \mathcal{T}(v,w)} \{\Gamma_{v,w}(M)\}$$

Zhang [14] and Kilpelläinen [15] showed that for two trees, this definition of valid mapping leads to an NP-complete problem. To alleviate this difficulty, an algorithm which solves this problem in a polynomial time has been proposed by Tai [3] for ordered trees by introducing a new constraint which preserves the order. The corresponding dissimilarity measure was shown to be a distance [3]. In case of unordered trees, Zhang proposed considering a new constraint in the definition of a valid EDM [12, 8] based on an initial idea proposed by Tanaka and Tanaka [16] for ordered trees: two separate sub-trees of one tree should be mapped onto two separate sub-trees of the other tree. Zhang extended this idea from ordered to unordered trees and changed the definition of a valid EDM as follows:

Definition 4 *Let $T_1[v] = (V_1[v], E_1[v])$ and $T_2[w] = (V_2[w], E_2[w])$ be two trees, a valid EDM M from $T_1[v]$ to $T_2[w]$ is a set of ordered pairs of vertices $(x, y) \in V_1[v] \times V_2[w]$ satisfying constraints (1), (2) and:*

$$\forall (x_1, x_2), (y_1, y_2), (z_1, z_2) \in M : x_1 \wedge y_1 < z_1 \Leftrightarrow x_2 \wedge y_2 < z_2 \quad (3)$$

The dissimilarity measure between two unordered trees $T_1[v]$ and $T_2[w]$ is defined as:

$$D(T_1[v], T_2[w]) = \min_{M \in \mathcal{T}(v,w)} \{\Gamma_{v,w}(M)\}$$

Zhang showed that the dissimilarity measure $D(T_1[v], T_2[w])$ is actually a distance [8] and proposed an algorithm with bounded complexity to solve the new optimization problem 2.

Problem 2 *Find $\Gamma_{v,w}(M)$ minimum, such that M is a valid EDM from $T_1[v]$ to $T_2[w]$ satisfying constraints (1), (2) and (3).*

3.4 Properties of valid EDMs

In this section we consider a valid EDM (according to definition 4) M from $S_1[v]$ to $S_2[w]$, where $S_1[v]$ and $S_2[w]$ are both either a tree or a forest. We show several properties of valid EDMs that enable us to derive the basic algorithm for comparing unordered trees [12] using this new formulation.

3.4.1 Properties of \widehat{M}

For valid EDMs, property 1 of \widehat{M} can be extended as follows:

Proposition 2 *For any valid EDM M :*

$$\begin{aligned} \forall (S_1[x], S_1[y]) \in \mathcal{S}_1(v) \times \mathcal{S}_1(v) : \\ S_1[x] \subseteq S_1[y] \Leftrightarrow \widehat{M}(S_1[x]) \subseteq \widehat{M}(S_1[y]) \end{aligned}$$

Note that the three constraints used to define valid EDMs (definition 4) are necessary for the equivalence to hold.

Proposition 3 *M is a valid EDM from $S_1[v]$ to $S_2[w]$ if and only if \widehat{M} satisfies one and only one of these five assertions:*

1. $\widehat{M}(S_1[v]) = \theta$ and $\widehat{M}(S_2[w]) = \theta$;
2. $\widehat{M}(S_1[v]) \neq \theta$ and $\widehat{M}(S_2[w]) \neq \theta$ then:
 - (a) $\widehat{M}(S_1[v]) \subset S_2[w]$ and $\widehat{M}(S_2[w]) = S_1[v]$;
 - (b) $\widehat{M}(S_1[v]) = S_2[w]$ and $\widehat{M}(S_2[w]) \subset S_1[v]$;
 - (c) $\widehat{M}(S_1[v]) = S_2[w]$ and $\widehat{M}(S_2[w]) = S_1[v]$;
 - (d) $\widehat{M}(S_1[v]) \subset S_2[w]$ and $\widehat{M}(S_2[w]) \subset S_1[v]$.

This proposition can be used to solve optimization problem 2 recursively [8]. This is achieved by applying the dynamic programming principle (*e.g.* [17]), to the computation of the optimal valid EDM. To express the recursive nature of the optimality principle, the set $\mathcal{M}(v, w)$ can be split into subsets (figure 2) as follows:

figure 2

1. $\widehat{M}(S_1[v]) = \theta$ and $\widehat{M}(S_2[w]) = \theta$;
 - $\mathcal{M}(v, w)_{\theta, \theta} = \left\{ M \in \mathcal{M}(v, w) \mid \widehat{M}(S_1[v]) = \theta \text{ and } \widehat{M}(S_2[w]) = \theta \right\}$
2. $\widehat{M}(S_1[v]) \neq \theta$ and $\widehat{M}(S_2[w]) \neq \theta$ then
 - $\mathcal{M}(v, w)_{C, =} = \left\{ M \in \mathcal{M}(v, w) \mid \widehat{M}(S_1[v]) \subset S_2[w] \text{ and } \widehat{M}(S_2[w]) = S_1[v] \right\}$

- $\mathcal{M}(v, w)_{=, \subset} = \left\{ M \in \mathcal{M}(v, w) \mid \widehat{M}(S_1[v]) = S_2[w] \text{ and } \widehat{M}(S_2[w]) \subset S_1[v] \right\}$
- $\mathcal{M}(v, w)_{=, =} = \left\{ M \in \mathcal{M}(v, w) \mid \widehat{M}(S_1[v]) = S_2[w] \text{ and } \widehat{M}(S_2[w]) = S_1[v] \right\}$
- $\mathcal{M}(v, w)_{\subset, \subset} = \left\{ M \in \mathcal{M}(v, w) \mid \widehat{M}(S_1[v]) \subset S_2[w] \text{ and } \widehat{M}(S_2[w]) \subset S_1[v] \right\}$

These subsets form a partition of $\mathcal{M}(v, w)$. In these definitions, $\mathcal{M}(v, w)$ represents either $\mathcal{T}(v, w)$ or $\mathcal{F}(v, w)$. For example, $\mathcal{T}(v, w)_{=, =}$ is the set of valid EDMs M from $T_1[v]$ to $T_2[w]$ such that $\widehat{M}(T_1[v]) = T_2[w]$ and $\widehat{M}(T_2[w]) = T_1[v]$; $\mathcal{F}(v, w)_{=, \subset}$ is the set of valid EDMs M from $F_1[v]$ to $F_2[w]$ such that $\widehat{M}(F_1[v]) = F_2[w]$ and $\widehat{M}(F_2[w]) \subset F_1[v]$.

Proposition 4 *For any valid EDM from $T_1[v]$ to $T_2[w]$, M is in $\mathcal{T}(v, w)_{=, =}$ if and only if:*

$$(v, w) \in M$$

For any valid EDM M from $T_1[v]$ to $T_2[w]$ in $\mathcal{T}(v, w)$, if (v, w) is in M , $M \setminus \{(v, w)\}$ is denoted by M^* . Furthermore, if M is in $\mathcal{T}(v, w)_{=, =}$, $\widehat{M}^*(T_1[v]) \subset T_2[w]$ and $\widehat{M}^*(T_2[w]) \subset T_1[v]$ and then according to the previous proposition, $M^* \in \mathcal{F}(v, w)$.

Proposition 5 *For any valid EDM from $T_1[v]$ to $T_2[w]$, M^* is in $\mathcal{T}(v, w)_{\subset, \subset}$ if and only if M^* is a valid EDM from $F_1[v]$ to $F_2[w]$:*

$$\mathcal{T}(v, w)_{\subset, \subset} = \mathcal{F}(v, w)$$

EDMs in $\mathcal{R}(v, w) = \mathcal{F}(v, w)_{=, =} \cup \mathcal{F}(v, w)_{\subset, \subset}$ have an additional remarkable property.

Proposition 6 *For any M in $\mathcal{R}(v, w)$ and for any v_i son of v , such that $\widehat{M}(T_1[v_i]) \neq \theta$, there exists a unique w_j , son of w such that $\widehat{M}(T_1[v_i]) \subseteq T_2[w_j]$. And for any son v_k of v , if $\widehat{M}(T_1[v_k]) \neq \theta$ and $\widehat{M}(T_1[v_k]) \subseteq T_2[w_j]$ then $v_k = v_i$.*

For any EDM from $\mathcal{R}(v, w)$, the image of any tree of $F_1[v]$ is either the empty tree or included in a tree of $F_2[w]$. This means that vertices from a tree $T_1[v_i]$ can only be mapped onto vertices of one tree $T_2[w_j]$ and reciprocally. $\mathcal{R}(v, w)$ thus defines a mapping between trees of $F_1[v]$ and trees of $F_2[w]$, called restricted EDM [8].

A matching of a graph is any subset of its edges such that no two members of the subset are adjacent [18]. We define a bipartite graph $G(v, w) = (V, E)$, where V represents $son[v] \cup son[w]$ and E is $son[v] \times son[w]$. The set of all possible matching on this graph is denoted by $\mathcal{K}(v, w)$.

3.4.2 Recursive expression of EDM sets

Proposition 3 can be directly expressed in terms of valid EDMs and reveals the different cases used by Zhang to establish the recurrent relations between EDM sets.

Proposition 7 *Let M be a valid EDM:*

1. if M is in $\mathcal{T}(v, w)$ then M satisfies one and only one of the follows assertions:

- (a) $\exists w_k \in \text{son}[w]$ such that $M \in \mathcal{T}(v, w_k)$;
- (b) $\exists v_k \in \text{son}[v]$ such that $M \in \mathcal{T}(v_k, w)$;
- (c) $(v, w) \in M$ and $M^* \in \mathcal{F}(v, w)$;
- (d) $M \in \mathcal{F}(v, w)$;
- (e) $M = \emptyset$.

2. if M is in $\mathcal{F}(v, w)$, then M satisfies one and only one of the follows assertions:

- (a) $\exists w_k \in \text{son}[w]$ such that $M \in \mathcal{F}(v, w_k)$;
- (b) $\exists v_k \in \text{son}[v]$ such that $M \in \mathcal{F}(v_k, w)$;
- (c) $M \in \mathcal{R}(v, w)$;
- (d) $M = \emptyset$.

Cases 1.e and 2.d do not appear in the original formulation since they represent limit cases. However, these limit cases will be exploited in the extension of the algorithm discussed in the next section.

The equivalence of properties 3 and 7 shows that the original formulation of Zhang's algorithm can be expressed in terms of the properties of \widehat{M} . The new formulation introduced in proposition 3 is more compact than the original formulation since it does need to make a distinction between forests and trees (as in property 7).

3.5 Recursive expression of the distance between unordered trees

The above proposition 7 of valid EDMs can be used to compute recursively the cost of a valid EDM with minimum cost.

Theorem 1 [12, 8] $D(T_1[v], T_2[w])$ and $D(F_1[v], F_2[w])$ can be computed recursively:

1. *initialisation:*

$$\begin{aligned} D(\theta, \theta) &= 0 \\ D(F_1[v], \theta) &= \sum_{v_k \in \text{son}[v]} D(T_1[v_k], \theta), & D(T_1[v], \theta) &= D(F_1[v], \theta) + d(v, \lambda) \\ D(\theta, F_2[w]) &= \sum_{w_k \in \text{son}[w]} D(\theta, T_2[w_k]), & D(\theta, T_2[w]) &= D(\theta, F_2[w]) + d(\lambda, w) \end{aligned}$$

2. *Distance between trees:*

$$D(T_1[v], T_2[w]) = \min \left\{ \begin{array}{l} D(\theta, T_2[w]) + \min_{w_k \in \text{son}[w]} \{D(T_1[v], T_2[w_k]) - D(\theta, T_2[w_k])\} \\ D(T_1[v], \theta) + \min_{v_k \in \text{son}[v]} \{D(T_1[v_k], T_2[w]) - D(T_1[v_k], \theta)\} \\ D(F_1[v], F_2[w]) + d(v, w) \end{array} \right.$$

3. Distance between forests:

$$D(F_1[v], F_2[w]) = \min \begin{cases} D(\theta, F_2[w]) + \min_{w_k \in \text{son}[w]} \{D(F_1[v], F_2[w_k]) - D(\theta, F_2[w_k])\} \\ D(F_1[v], \theta) + \min_{v_k \in \text{son}[v]} \{D(F_1[v_k], F_2[w]) - D(F_1[v_k], \theta)\} \\ \min_{M \in \mathcal{R}(v,w)} \{\gamma(M)\} \end{cases}$$

Zhang [12] models the computation of $\min_{M \in \mathcal{R}(v,w)} \{\gamma(M)\}$ as a problem of minimum cost maximum flow, which mainly determines the overall complexity of the final algorithm. The cost of optimal restricted EDMs is studied in section 4.4. The complexity of this algorithm is:

$$O(|T_1| \times |T_2| \times (\deg(T_1) + \deg(T_2)) \times \log_2(\deg(T_1) + \deg(T_2)))$$

4 Distance between quotiented trees

4.1 Valid EDMs

Let us consider two quotiented trees $G_1 = (T_1, W_1, \pi_1)$ and $G_2 = (T_2, W_2, \pi_2)$ such that the roots of T_1 and T_2 are respectively v and w (if no confusion is possible, π_1 and π_2 are denoted by π). Let M be an EDM from $T_1[v]$ to $T_2[w]$. M induces an EDM from tree $\mathcal{Q}(G_1)$ to tree $\mathcal{Q}(G_2)$, called *the quotient EDM*, denoted by $\mathcal{Q}(M)$, composed of pairs of vertices in $W_1 \times W_2$ and defined as:

$$(a, b) \in \mathcal{Q}(M) \Leftrightarrow \exists (z, t) \in M \text{ such that } \begin{cases} \pi(z) = a \\ \pi(t) = b \end{cases}$$

In the following, $\mathcal{Q}(M)$ will be denoted by N .

Let x be a vertex of $T_1[v]$ (resp. of $T_2[w]$), $\mathcal{Q}(M_x)$ denotes the set of vertices of quotient graphs $\mathcal{Q}(G_1[x])$ or (resp. of $\mathcal{Q}(G_2[y])$) which have an image by $\mathcal{Q}(M)$:

$$\mathcal{Q}(M_x) = \{\pi(z) \mid z \in M_x\}$$

It should be noted that if M denotes a valid EDM from $T_1[v]$ to $T_2[w]$, $\mathcal{Q}(M)$ is not necessarily a valid EDM from tree $\mathcal{Q}(G_1)$ to tree $\mathcal{Q}(G_2)$ (see figure 3).

figure 3

Definition 5 (valid EDMs on quotiented trees) *Let $G_1[v] = (T_1, W_1, \pi_1)$ and $G_2[w] = (T_2, W_2, \pi_2)$ be two quotiented trees, a valid EDM M from $G_1[v]$ to $G_2[w]$ is a valid EDM from $T_1[v]$ to $T_2[w]$ such that $\mathcal{Q}(M)$ is also a valid EDM from $\mathcal{Q}(G_1)$ to $\mathcal{Q}(G_2)$.*

The set of valid mappings from $T_1[\pi(v)]$ to $T_2[\pi(w)]$ is denoted by $\mathcal{T}(\pi(v), \pi(w))$. Thus by definition M is a valid EDM from $G_1[v]$ to $G_2[w]$ if, and only if, M is in $\mathcal{T}(v, w)$ and $\mathcal{Q}(M)$ is in $\mathcal{T}(\pi(v), \pi(w))$. The set of valid EDMs from $G_1[v]$ to $G_2[w]$ is denoted by $\mathcal{G}(v, w)$.

Similarly to unordered tree comparison, we need to consider a set of valid EDMs between the quotiented trees $G_1[v]$ and $G_2[w]$ in which v and w do not have any image by the EDM, *i.e.* EDMs between forests $F_1[v]$ and $F_2[w]$. This set will be denoted by $\mathcal{H}(v, w)$:

$$\mathcal{H}(v, w) = \mathcal{G}(v, w) \cap \mathcal{F}(v, w)$$

A dissimilarity measure between quotiented trees is then defined by the following optimization problem.

Problem 3 Finding $\Gamma_{v,w}(M)$ minimum, such that M is a valid EDM from $G_1[v]$ to $G_2[w]$ satisfying definition 5:

$$D(G_1[v], G_2[w]) = \min_{M \in \mathcal{G}(v,w)} \{\Gamma_{v,w}(M)\}$$

Lemma 1 D is a distance.

4.2 Properties of valid EDMs

According to the properties described in section 3.4, the efficient computation of a distance between unordered trees relies on the possibility to apply the dynamic programming principle using recursive relations between the set of EDMs $\mathcal{T}(v, w)$ and the sets $\{\mathcal{T}(v, w_k), \mathcal{T}(v_i, w), \mathcal{T}(v_i, w_k), \mathcal{F}(v, w)\}$ where v_i and w_k are respectively sons of v and w . In a similar fashion, we wish to determine a recursive expression between the set $\mathcal{G}(v, w)$ and $\{\mathcal{G}(v, w_k), \mathcal{G}(v_i, w), \mathcal{G}(v_i, w_k), \mathcal{H}(v, w)\}$ of valid EDMs on quotiented trees which will enable us to solve efficiently the optimization problem 3 using a dynamic programming-based algorithm.

figure 4

In the sequel, we will show that such recursive expressions can be obtained by defining an adequate partition of $\mathcal{G}(v, w)$. This partition is based on a two level scheme Figure 4: a first partition of $\mathcal{G}(v, w)$ is made, based on the partitioning of $\mathcal{T}(v, w)$ into subsets $\mathcal{T}(v, w)_{=,=}$, $\mathcal{T}(v, w)_{\subset,=}$, $\mathcal{T}(v, w)_{\subset,=}$, $\mathcal{T}(v, w)_{\subset,\subset} = \mathcal{F}(\pi(v), \pi(w))$ and $\mathcal{T}(v, w)_{\theta,\theta}$ (Figure 4.1); $\mathcal{T}(v, w)_{\theta,\theta}$ is not represented on the figure). Note that, as explained in the previous section, $\mathcal{F}(\pi(v), \pi(w))$ is further decomposed into $\mathcal{F}(v, w)_{=,=}$, $\mathcal{F}(v, w)_{\subset,=}$, $\mathcal{F}(v, w)_{\subset,=}$, $\mathcal{F}(v, w)_{\subset,\subset}$ and $\mathcal{F}(v, w)_{\theta,\theta}$ (Figure 4.2; $\mathcal{F}(v, w)_{\theta,\theta}$ is not represented on the figure). Then, at a second level, each set of the resulting partition is itself decomposed into a partition based on configurations of valid EDM on quotient graphs, *i.e.* partitions of $\mathcal{T}(\pi(v), \pi(w))$ and $\mathcal{F}(\pi(v), \pi(w))$. Each set is decomposed into five new sub-sets depending on the image of $T_1[\pi(v)]$ and $T_2[\pi(w)]$ by \widehat{N} (Figure 4.3):

- a: $\widehat{N}(T_1[\pi(v)]) \subset T_2[\pi(w)]$ and $\widehat{N}(T_2[\pi(w)]) = T_1[\pi(v)]$;
- b: $\widehat{N}(T_1[\pi(v)]) = T_2[\pi(w)]$ and $\widehat{N}(T_2[\pi(w)]) \subset T_1[\pi(v)]$;
- c: $\widehat{N}(T_1[\pi(v)]) = T_2[\pi(w)]$ and $\widehat{N}(T_2[\pi(w)]) = T_1[\pi(v)]$;
- d: $\widehat{N}(T_1[\pi(v)]) \subset T_2[\pi(w)]$ and $\widehat{N}(T_2[\pi(w)]) \subset T_1[\pi(v)]$.

Note that the case corresponding to $\widehat{N}(T_1[\pi(v)]) = \theta$ and $\widehat{N}(T_2[\pi(w)]) = \theta$ is not represented on the figure.

Finally, the sub-sets corresponding to case d (labelled d on Figure 4.3) are further decomposed into five new sub-sets depending on the image of $F_1[\pi(v)]$ and $F_2[\pi(w)]$ by \widehat{N} (Figure 4.4):

- e: $\widehat{N}(F_1[\pi(v)]) \subset F_2[\pi(w)]$ and $\widehat{N}(F_2[\pi(w)]) = F_1[\pi(v)]$;
- f: $\widehat{N}(F_1[\pi(v)]) = F_2[\pi(w)]$ and $\widehat{N}(F_2[\pi(w)]) \subset F_1[\pi(v)]$;
- g: $\widehat{N}(F_1[\pi(v)]) = F_2[\pi(w)]$ and $\widehat{N}(F_2[\pi(w)]) = F_1[\pi(v)]$;
- h: $\widehat{N}(F_1[\pi(v)]) \subset F_2[\pi(w)]$ and $\widehat{N}(F_2[\pi(w)]) \subset F_1[\pi(v)]$.

Note that the case corresponding to $\widehat{N}(F_1[\pi(v)]) = \theta$ and $\widehat{N}(F_2[\pi(w)]) = \theta$ is not represented on the figure.

Hence the combinatorics of the different configurations of interest of valid EDMs both at microscopic level (tree level) and at macroscopic level (quotient tree level) results in a partition of $\mathcal{G}(v, w)$ into 51 sub-sets. To establish recursive relations between these sub-sets, we study in the next section their properties

4.2.1 Properties of \widehat{N}

In the following, if no confusion is possible, $\mathcal{G}(v, w)$ and $\mathcal{H}(v, w)$ are both denoted by $\mathcal{M}(v, w)$. If M belongs to $\mathcal{M}(v, w)$, $N = \mathcal{Q}(M)$ is a valid EDM from $S_1[\pi(v)]$ to $S_2[\pi(w)]$. Then, according to proposition 3, M belongs to one and only one of the following sets, depending on the respective images of $S_1[\pi(v)]$ and $S_2[\pi(w)]$ by \widehat{N} :

- $[\mathcal{M}(v, w)]_{C,=} = \{M \in \mathcal{M}(v, w) \mid N \in \mathcal{T}(\pi(v), \pi(w))_{C,=}\}$ is the set of valid EDMs of $\mathcal{M}(v, w)$ such that $\mathcal{Q}(M)$ is in $\mathcal{T}(\pi(v), \pi(w))_{C,=}$, this means that $S_1[\pi(v)]$ has an image included in $S_2[\pi(w)]$ and the image of $S_2[\pi(w)]$ is $S_1[\pi(v)]$:

Similarly:

- $[\mathcal{M}(v, w)]_{=,C} = \{M \in \mathcal{M}(v, w) \mid N \in \mathcal{T}(\pi(v), \pi(w))_{=,C}\}$;
- $[\mathcal{M}(v, w)]_{=,=} = \{M \in \mathcal{M}(v, w) \mid N \in \mathcal{T}(\pi(v), \pi(w))_{=,=}\}$;
- $[\mathcal{M}(v, w)]_{C,C} = \{M \in \mathcal{M}(v, w) \mid N \in \mathcal{T}(\pi(v), \pi(w))_{C,C}\}$;
- $[\mathcal{M}(v, w)]_{\theta,\theta} = \{M \in \mathcal{M}(v, w) \mid N \in \mathcal{T}(\pi(v), \pi(w))_{\theta,\theta}\}$.

$N \setminus \{(\pi(v), \pi(w))\}$ is denoted by N^* . Note here that according to proposition 4, for any valid EDM N from $T_1[\pi(v)]$ to $T_2[\pi(w)]$, N is in $\mathcal{T}(\pi(v), \pi(w))_{=,=}$ if and only if $(\pi(v), \pi(w)) \in N$ and if N^* is a valid EDM in $\mathcal{F}(\pi(v), \pi(w))$. Similarly N is in $\mathcal{T}(\pi(v), \pi(w))_{C,C}$, if and only if $N = N^*$ is a valid EDM in $\mathcal{F}(\pi(v), \pi(w))$. Therefore, to express recursive relations based on sets $\mathcal{T}(\pi(v), \pi(w))_{=,=}$ and $\mathcal{T}(\pi(v), \pi(w))_{C,C}$, we need to consider partitions of $[\mathcal{M}(v, w)]_{=,=}$ and $[\mathcal{M}(v, w)]_{C,C}$ based respectively on the membership of N^* or N in the different subsets of $\mathcal{F}(\pi(v), \pi(w))$:

- $\left([\mathcal{M}(v, w)]_{=,=}\right)_{C,=} = \left\{M \in [\mathcal{M}(v, w)]_{=,=} \mid N^* \in \mathcal{F}(\pi(v), \pi(w))_{C,=}\right\}$;
- $\left([\mathcal{M}(v, w)]_{C,C}\right)_{C,=} = \left\{M \in [\mathcal{M}(v, w)]_{C,C} \mid N^* \in \mathcal{F}(\pi(v), \pi(w))_{C,=}\right\}$;

Note that in the case of $[\mathcal{M}(v, w)]_{C,C}$, $N^* = N$. Sets $\left([\mathcal{M}(v, w)]_{=,=}\right)_{=,=}$, $\left([\mathcal{M}(v, w)]_{=,=}\right)_{=,C}$, $\left([\mathcal{M}(v, w)]_{=,=}\right)_{\theta,\theta}$, $\left([\mathcal{M}(v, w)]_{C,C}\right)_{=,=}$, $\left([\mathcal{M}(v, w)]_{C,C}\right)_{=,C}$ and $\left([\mathcal{M}(v, w)]_{C,C}\right)_{\theta,\theta}$ are defined similarly. The different types of EDMs corresponding to the partition of $[\mathcal{M}(v, w)]_{=,=}$ are represented graphically in figure 5.

figure 5

To find recursive relations between these sets for a pair of vertices (v, w) , we need to study how such sets can be computed from similar sets associated with the descendants of v and w . To compute recursively M in $\mathcal{M}(v, w)$ from M_i s in $\mathcal{M}(x_i, y_i)$, where x_i and y_i are descendants of v and w respectively, we need to study two kinds of situation: either M is identical to one of the M_i s (lemma 2) or M is a union of the M_i s (lemma 3).

Lemma 2 *Let x and y both be descendants of v and w respectively, for any valid EDM M in $\mathcal{M}(x, y)$, then M is a valid EDM in $\mathcal{M}(v, w)$ such that:*

$$\widehat{N}(S_1[\pi(x)]) = \theta \Rightarrow \widehat{N}(S_1[\pi(v)]) = \theta \quad (4)$$

$$\widehat{N}(S_1[\pi(x)]) \subset S_2[\pi(y)] \Rightarrow \widehat{N}(S_1[\pi(v)]) \subset S_2[\pi(w)] \quad (5)$$

$$\widehat{N}(S_1[\pi(x)]) = S_2[\pi(y)] \Rightarrow \begin{cases} \widehat{N}(S_1[\pi(v)]) = S_2[\pi(w)] & \text{if } \pi(y) = \pi(w) \\ \widehat{N}(S_1[\pi(v)]) \subset S_2[\pi(w)] & \text{otherwise} \end{cases} \quad (6)$$

Here it should be recalled that S represents either a tree or a forest and that x and y can be respectively equal to v and w . By symmetry, the same proposition holds if the roles of S_1 and S_2 are inverted. This proposition can be used to compute the sets of the partition of $\mathcal{M}(v, w)$, i.e. $[\mathcal{M}(v, w)]_{C,=}$, $[\mathcal{M}(v, w)]_{=,C}$, $[\mathcal{M}(v, w)]_{=,=}$, $[\mathcal{M}(v, w)]_{C,C}$, from those of the partition of $\mathcal{M}(x, y)$.

For example, if M denotes a valid EDM in $[\mathcal{M}(x, y)]_{C,=}$, by definition $N = \mathcal{Q}(M)$ is in $\mathcal{T}(\pi(x), \pi(y))_{C,=}$, which means that $\widehat{N}(T_1[\pi(x)]) \subset T_2[\pi(y)]$ and $\widehat{N}(T_2[\pi(y)]) = T_1[\pi(x)]$. Then according to lemma 2, M is a valid EDM in $\mathcal{M}(v, w)$ such that:

- if $\pi(x) = \pi(v)$ then $\widehat{N}(T_1[\pi(v)]) \subset T_2[\pi(w)]$ (from (5)) and $\widehat{N}(T_2[\pi(w)]) = T_1[\pi(v)]$ (from (6)), which means that N is in $\mathcal{T}(\pi(v), \pi(w))_{C,=}$.
- Otherwise $\widehat{N}(T_1[\pi(v)]) \subset T_2[\pi(w)]$ (from (5)) and $\widehat{N}(T_2[\pi(w)]) \subset T_1[\pi(v)]$ (from (6)), which means that N is in $\mathcal{T}(\pi(v), \pi(w))_{C,C}$.

In other terms, M is in $[\mathcal{M}(v, w)]_{C,=}$ if $\pi(x) = \pi(v)$, otherwise M is in $[\mathcal{M}(v, w)]_{C,C}$.

Proposition 8 details similar relationships between sets $[\mathcal{M}(x, y)]_{\alpha,\beta}$ and $[\mathcal{M}(v, w)]_{\alpha',\beta'}$, where α, β, α' and β' are in $\{C, =, \theta\}$.

Proposition 8 *Let x and y both be descendants of v and w respectively, for any valid EDM M in $[\mathcal{M}(x, y)]_{\alpha,\beta}$, then M is a valid EDM in $[\mathcal{M}(v, w)]_{\alpha',\beta'}$ as detailed in table 1.*

In section 4.4, we shall need to further analyse sets $[\mathcal{M}(v, w)]_{=,=}$ and $[\mathcal{M}(v, w)]_{C,C}$ to derive a complete and sound recursive expression of $\mathcal{M}(v, w)$. To achieve this, lemma 2 can also be applied to derive inclusions between sets of the partition of $[\mathcal{M}(x, y)]_{C,C}$ (resp. $[\mathcal{M}(x, y)]_{=,=}$), i.e. $\left([\mathcal{M}(x, y)]_{C,C}\right)_{=,C}$, $\left([\mathcal{M}(x, y)]_{C,C}\right)_{C,=}$, $\left([\mathcal{M}(x, y)]_{C,C}\right)_{=,=}$, ... (resp. $\left([\mathcal{M}(x, y)]_{=,=}\right)_{=,C}$, $\left([\mathcal{M}(x, y)]_{=,=}\right)_{C,=}$, ...), and those of the partition of $[\mathcal{M}(v, w)]_{=,=}$ and $[\mathcal{M}(v, w)]_{C,C}$.

For example, consider M in $\left([\mathcal{M}(x, y)]_{C,C}\right)_{C,=}$. By definition N^* is in $\mathcal{F}(\pi(x), \pi(y))_{C,=}$, then according to lemma 2, M is a valid EDM in $\mathcal{M}(v, w)$ such that :

$\pi(x)$	$\pi(y)$	$[\mathcal{M}(x, y)]_{C,=}$	$[\mathcal{M}(x, y)]_{=,C}$	$[\mathcal{M}(x, y)]_{=,=}$	$[\mathcal{M}(x, y)]_{C,C}$	$[\mathcal{M}(x, y)]_{\theta,\theta}$
$= \pi(v)$	$= \pi(w)$	$[\mathcal{M}(v, w)]_{C,=}$	$[\mathcal{M}(v, w)]_{=,C}$	$[\mathcal{M}(v, w)]_{=,=}$	$[\mathcal{M}(v, w)]_{C,C}$	$[\mathcal{M}(v, w)]_{\theta,\theta}$
$= \pi(v)$	$\neq \pi(w)$	$[\mathcal{M}(v, w)]_{C,=}$	$[\mathcal{M}(v, w)]_{C,C}$	$[\mathcal{M}(v, w)]_{C,=}$	$[\mathcal{M}(v, w)]_{C,C}$	$[\mathcal{M}(v, w)]_{\theta,\theta}$
$\neq \pi(v)$	$= \pi(w)$	$[\mathcal{M}(v, w)]_{C,C}$	$[\mathcal{M}(v, w)]_{=,C}$	$[\mathcal{M}(v, w)]_{=,C}$	$[\mathcal{M}(v, w)]_{C,C}$	$[\mathcal{M}(v, w)]_{\theta,\theta}$
$\neq \pi(v)$	$\neq \pi(w)$	$[\mathcal{M}(v, w)]_{C,C}$	$[\mathcal{M}(v, w)]_{C,C}$	$[\mathcal{M}(v, w)]_{C,C}$	$[\mathcal{M}(v, w)]_{C,C}$	$[\mathcal{M}(v, w)]_{\theta,\theta}$

Table 1: Membership of a valid EDM in $\mathcal{M}(x, y)$ into $\mathcal{M}(v, w)$ depending on $\pi(x)$ and $\pi(y)$.

$\pi(x)$	$\pi(y)$	$([\mathcal{M}(x, y)]_{C,C})_{C,=}$	$([\mathcal{M}(x, y)]_{C,C})_{=,C}$	$([\mathcal{M}(x, y)]_{C,C})_{=,=}$	$([\mathcal{M}(x, y)]_{C,C})_{C,C}$	$([\mathcal{M}(x, y)]_{C,C})_{\theta,\theta}$
$= \pi(v)$	$= \pi(w)$	$([\mathcal{M}(v, w)]_{C,C})_{C,=}$	$([\mathcal{M}(v, w)]_{C,C})_{=,C}$	$([\mathcal{M}(v, w)]_{C,C})_{=,=}$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	\emptyset
$= \pi(v)$	$\neq \pi(w)$	$([\mathcal{M}(v, w)]_{C,C})_{C,=}$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	$([\mathcal{M}(v, w)]_{C,C})_{C,=}$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	\emptyset
$\neq \pi(v)$	$= \pi(w)$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	$([\mathcal{M}(v, w)]_{C,C})_{=,C}$	$([\mathcal{M}(v, w)]_{C,C})_{=,C}$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	\emptyset
$\neq \pi(v)$	$\neq \pi(w)$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	\emptyset

Table 2: Membership of a valid EDM in $[\mathcal{M}(x, y)]_{C,C}$ into $\mathcal{M}(v, w)$ depending on $\pi(x)$ and $\pi(y)$.

- if $\pi(x) = \pi(v)$ then N is in $\mathcal{F}(\pi(v), \pi(w))_{C,=}$;
- otherwise N is in $\mathcal{F}(\pi(v), \pi(w))_{C,C}$.

Finally, M is in $([\mathcal{M}(v, w)]_{C,C})_{C,=}$, if $\pi(x) = \pi(v)$, otherwise M is in $([\mathcal{M}(v, w)]_{C,C})_{C,C}$.

All these results are summarised by the following propositions.

Proposition 9 *Let x and y both be descendants of v and w respectively, for any valid EDM M in $[\mathcal{M}(x, y)]_{C,C}$, then M is a valid EDM in one of set of the partition of $[\mathcal{M}(v, w)]_{C,C}$, as detailed in table 2.*

Proposition 10 *Let x and y both be descendants of v and w respectively, for any valid EDM M in $[\mathcal{M}(x, y)]_{=,=}$, then M is a valid EDM in $\mathcal{M}(v, w)$ as detailed in table 3.*

In the same way, the propositions below, give the relations between an EDM M in $\mathcal{M}(v, w)$ and EDMs M_i s in $\mathcal{M}(x_i, y_i)$, where x_i and y_i are descendants of v and w , when M can be considered as a union of the M_i s.

$\pi(x)$	$\pi(y)$	$([\mathcal{M}(x, y)]_{=,=})_{C,=}$	$([\mathcal{M}(x, y)]_{=,=})_{=,C}$	$([\mathcal{M}(x, y)]_{=,=})_{=,=}$	$([\mathcal{M}(x, y)]_{=,=})_{C,C}$	$([\mathcal{M}(x, y)]_{=,=})_{\theta,\theta}$
$= \pi(v)$	$= \pi(w)$	$([\mathcal{M}(v, w)]_{=,=})_{C,=}$	$([\mathcal{M}(v, w)]_{=,=})_{=,C}$	$([\mathcal{M}(v, w)]_{=,=})_{=,=}$	$([\mathcal{M}(v, w)]_{=,=})_{C,C}$	$([\mathcal{M}(v, w)]_{=,=})_{\theta,\theta}$
$= \pi(v)$	$\neq \pi(w)$	$[\mathcal{M}(v, w)]_{C,=}$	$[\mathcal{M}(v, w)]_{C,=}$	$[\mathcal{M}(v, w)]_{C,=}$	$[\mathcal{M}(v, w)]_{C,=}$	$[\mathcal{M}(v, w)]_{C,=}$
$\neq \pi(v)$	$= \pi(w)$	$[\mathcal{M}(v, w)]_{=,C}$	$[\mathcal{M}(v, w)]_{=,C}$	$[\mathcal{M}(v, w)]_{=,C}$	$[\mathcal{M}(v, w)]_{=,C}$	$[\mathcal{M}(v, w)]_{=,C}$
$\neq \pi(v)$	$\neq \pi(w)$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	$([\mathcal{M}(v, w)]_{C,C})_{\theta,\theta}$

Table 3: Membership of a valid EDM in $[\mathcal{M}(x, y)]_{=,=}$ into $\mathcal{M}(v, w)$ depending on $\pi(x)$ and $\pi(y)$.

Lemma 3 Let v_1, v_2, \dots, v_{n_v} be the sons of v and let w_1, w_2, \dots, w_{n_w} be the sons of w . Consider n EDMs $(M_k)_{k \in \{1..n\}}$ in $\mathcal{M}(v_{p_k}, w_{q_k})$ such that for any M_i and M_j , $i \neq j$ if and only if $v_{p_i} \neq v_{p_j}$ and $w_{p_i} \neq w_{p_j}$. Let M be the union $\bigcup_{k \in \{1..n\}} \{M_k\}$:

$$\begin{aligned}
N \in \mathcal{T}(\pi(v), \pi(w))_{C,=} &\Leftrightarrow \exists i \mid N_i \in \mathcal{T}(\pi(v), \pi(w))_{C,=} \text{ and } \forall j \neq i, N_j \in \mathcal{T}(\pi(v), \pi(w))_{\theta, \theta} \\
N \in \mathcal{T}(\pi(v), \pi(w))_{=,=} &\Leftrightarrow \exists i \mid N_i \in \mathcal{T}(\pi(v), \pi(w))_{=,=} \text{ and } \forall j \neq i : \begin{cases} N_j \in \mathcal{T}(\pi(v), \pi(w))_{=,=} \\ N_j \in \mathcal{T}(\pi(v), \pi(w))_{C,C} \\ N_j \in \mathcal{T}(\pi(v), \pi(w))_{\theta, \theta} \end{cases} \\
N \in \mathcal{T}(\pi(v), \pi(w))_{C,C} &\Leftrightarrow \exists i \mid N_i \in \mathcal{T}(\pi(v), \pi(w))_{C,C} \text{ and } \forall j \neq i : \begin{cases} N_i \in \mathcal{T}(\pi(v), \pi(w))_{C,C} \\ N_i \in \mathcal{T}(\pi(v), \pi(w))_{\theta, \theta} \end{cases} \\
N^* \in \mathcal{F}(\pi(v), \pi(w))_{C,=} &\Leftrightarrow \exists i \mid N_i^* \in \mathcal{F}(\pi(v), \pi(w))_{C,=} \text{ and } \forall j \neq i, N_j^* \in \mathcal{F}(\pi(v), \pi(w))_{\theta, \theta} \\
N^* \in \mathcal{F}(\pi(v), \pi(w))_{=,=} &\Leftrightarrow \exists i \mid N_i^* \in \mathcal{F}(\pi(v), \pi(w))_{=,=} \text{ and } \forall j \neq i : \begin{cases} N_j^* \in \mathcal{F}(\pi(v), \pi(w))_{=,=} \\ N_j^* \in \mathcal{F}(\pi(v), \pi(w))_{C,C} \\ N_j^* \in \mathcal{F}(\pi(v), \pi(w))_{\theta, \theta} \end{cases} \\
N^* \in \mathcal{F}(\pi(v), \pi(w))_{C,C} &\Leftrightarrow \exists i \mid N_i^* \in \mathcal{F}(\pi(v), \pi(w))_{C,C} \text{ and } \forall j \neq i : \begin{cases} N_i^* \in \mathcal{F}(\pi(v), \pi(w))_{C,C} \\ N_i^* \in \mathcal{F}(\pi(v), \pi(w))_{\theta, \theta} \end{cases}
\end{aligned}$$

where N_i^* denotes the valid EDM $N_i \setminus \{(\pi(v_{p_i}), \pi(w_{q_i}))\}$.

According to this proposition it is possible to find relationships between the sets of the partition of $\mathcal{M}(v_{p_i}, w_{q_i})$, *i.e.* $[\mathcal{M}(v_{p_i}, w_{q_i})]_{C,=}$, $[\mathcal{M}(v_{p_i}, w_{q_i})]_{=,=}$, $[\mathcal{M}(v_{p_i}, w_{q_i})]_{=,=}$, $[\mathcal{M}(v_{p_i}, w_{q_i})]_{C,C}$, and those of the partition of $\mathcal{M}(v, w)$.

For example, consider n valid EDMs M_1, M_2, \dots and M_n in $\mathcal{M}(v_{p_1}, w_{q_1})$, $\mathcal{M}(v_{p_2}, w_{q_2})$, \dots and $\mathcal{M}(v_{p_n}, w_{q_n})$ respectively, if there exists i in $\{1..n\}$ such that M_i denotes a valid EDM in $[\mathcal{M}(x, y)]_{C,=}$, where $\pi(x) = \pi(v)$ and $\pi(y) = \pi(w)$, according to lemma 2, N_i is in $\mathcal{T}(\pi(v), \pi(w))_{C,=}$. As established in lemma 3, $N = \bigcup_{k \in \{1..n\}} \{N_k\}$ is in $N \in \mathcal{T}(\pi(v), \pi(w))_{C,=}$ if and only if for any $j \neq i$, $N_j \in \mathcal{T}(\pi(v), \pi(w))_{\theta, \theta}$. In other terms, M is a valid EDM in $[\mathcal{M}(v, w)]_{C,=}$ if and only if M_j is in $[\mathcal{M}(x, y)]_{\theta, \theta}$.

The next proposition summarize results which allow us to determine the union of sets $(M_k)_{k \in \{1..n\}}$ in the other cases.

Let v_1, v_2, \dots, v_{n_v} be the sons of v and let w_1, w_2, \dots, w_{n_w} be the sons of w . Consider n EDMs $(M_k)_{k \in \{1..n\}}$ in $\mathcal{M}(v_{p_k}, w_{q_k})$ such that for any M_i and M_j , $i \neq j$ if and only if $v_{p_i} \neq v_{p_j}$ and $w_{p_i} \neq w_{p_j}$. Let M be the union $\bigcup_{k \in \{1..n\}} \{M_k\}$, then M is in $\mathcal{A}(v, w)$ if there exists M_i in $\mathcal{B}(v, w)$ and for any M_j , $j \neq i$, M_j is in $\mathcal{C}(v, w)$, where $\mathcal{A}(v, w)$, $\mathcal{B}(v, w)$ and $\mathcal{C}(v, w)$ are sets of valid EDMs as detailed in table 4.

4.2.2 Recursive expression of EDM sets

These results are used to determine the sets of the partition of $\mathcal{G}(v, w)$ and $\mathcal{H}(v, w)$.

Proposition 11 Let M be a valid EDM from $T_1[v]$ to $T_2[w]$ of $[\mathcal{G}(v, w)]_{C,=}$, then M satisfies one and only one of the follow assertions:

1. $\exists w_k \in \text{son}[w]$ such that:

$M \in \mathcal{A}(v, w)$	$M_i \in \mathcal{B}(v, w)$	$M_j \in \mathcal{C}(v, w)$		
$[\mathcal{M}(v, w)]_{C,=}$	$[\mathcal{M}(v, w)]_{C,=}$	$[\mathcal{M}(v, w)]_{\theta,\theta}$		
$[\mathcal{M}(v, w)]_{=,C}$	$[\mathcal{M}(v, w)]_{=,C}$	$[\mathcal{M}(v, w)]_{\theta,\theta}$		
$[\mathcal{M}(v, w)]_{=,=}$	$[\mathcal{M}(v, w)]_{=,=}$	$[\mathcal{M}(v, w)]_{=,=}$	$[\mathcal{M}(v, w)]_{C,C}$	$[\mathcal{M}(v, w)]_{\theta,\theta}$
$([\mathcal{M}(v, w)]_{=,=})_{C,=}$	$([\mathcal{M}(v, w)]_{=,=})_{C,=}$	$([\mathcal{M}(v, w)]_{=,=})_{\theta,\theta}$		
$([\mathcal{M}(v, w)]_{=,=})_{=,C}$	$([\mathcal{M}(v, w)]_{=,=})_{=,C}$	$([\mathcal{M}(v, w)]_{=,=})_{\theta,\theta}$		
$([\mathcal{M}(v, w)]_{=,=})_{=,=}$	$([\mathcal{M}(v, w)]_{=,=})_{=,=}$	$([\mathcal{M}(v, w)]_{=,=})_{=,=}$	$([\mathcal{M}(v, w)]_{C,C})_{=,=}$	$([\mathcal{M}(v, w)]_{=,=})_{\theta,\theta}$
$([\mathcal{M}(v, w)]_{=,=})_{C,C}$	$([\mathcal{M}(v, w)]_{=,=})_{C,C}$	$([\mathcal{M}(v, w)]_{=,=})_{C,C}$	$([\mathcal{M}(v, w)]_{=,=})_{\theta,\theta}$	
$([\mathcal{M}(v, w)]_{=,=})_{\theta,\theta}$	$([\mathcal{M}(v, w)]_{=,=})_{\theta,\theta}$	$([\mathcal{M}(v, w)]_{=,=})_{\theta,\theta}$		
$[\mathcal{M}(v, w)]_{C,C}$	$[\mathcal{M}(v, w)]_{C,C}$	$[\mathcal{M}(v, w)]_{C,C}$	$[\mathcal{M}(v, w)]_{\theta,\theta}$	
$([\mathcal{M}(v, w)]_{C,C})_{C,=}$	$([\mathcal{M}(v, w)]_{C,C})_{C,=}$	$([\mathcal{M}(v, w)]_{C,C})_{\theta,\theta}$		
$([\mathcal{M}(v, w)]_{C,C})_{=,C}$	$([\mathcal{M}(v, w)]_{C,C})_{=,C}$	$([\mathcal{M}(v, w)]_{C,C})_{\theta,\theta}$		
$([\mathcal{M}(v, w)]_{C,C})_{=,=}$	$([\mathcal{M}(v, w)]_{C,C})_{=,=}$	$([\mathcal{M}(v, w)]_{C,C})_{=,=}$	$([\mathcal{M}(v, w)]_{C,C})_{\theta,\theta}$	
$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	$([\mathcal{M}(v, w)]_{C,C})_{C,C}$	$([\mathcal{M}(v, w)]_{C,C})_{\theta,\theta}$	
$([\mathcal{M}(v, w)]_{C,C})_{\theta,\theta}$	$([\mathcal{M}(v, w)]_{C,C})_{\theta,\theta}$	$([\mathcal{M}(v, w)]_{C,C})_{\theta,\theta}$		
$[\mathcal{M}(v, w)]_{\theta,\theta}$	$[\mathcal{M}(v, w)]_{\theta,\theta}$	$[\mathcal{M}(v, w)]_{\theta,\theta}$		

Table 4: Membership of an union of valid EDMs $M = \bigcup_{k \in \{1..n\}} \{M_k\}$ to $\mathcal{M}(v, w)$. M_i is a necessary EDM, while M_j is not a necessary EDM. An empty cell denotes the empty set.

- $\pi(w_k) \neq \pi(w)$ and $M \in [\mathcal{G}(v, w_k)]_{C,=} \cup [\mathcal{G}(v, w_k)]_{=,=}$;
 - $\pi(w_k) = \pi(w)$ and $M \in [\mathcal{G}(v, w_k)]_{C,=}$;
2. $\exists v_k \in \text{son}[v]$ such that $\pi(v_k) = \pi(v)$ and $M \in [\mathcal{G}(v_k, w)]_{C,=}$.
 3. $M \in [\mathcal{H}(v, w)]_{C,=}$;
 4. $M = \emptyset$.

SKETCH OF THE PROOF. The proof of this proposition is based on propositions 7 and lemmas 3 and 2. Let M be a valid EDM of $[\mathcal{G}(v, w)]_{C,=}$, then according to proposition 7, M satisfies one and only one of the five assertions:

1. $\exists w_k \in \text{son}[w]$ such that $M \in \mathcal{G}(v, w_k)$;
2. $\exists v_k \in \text{son}[v]$ such that $M \in \mathcal{G}(v_k, w)$;
3. $(v, w) \in M$ and $M^* \in \mathcal{H}(v, w)$;
4. $M \in \mathcal{H}(v, w)$;
5. $M = \emptyset$.

Suppose that there exists w_k a son of w such that M is in $\mathcal{G}(v, w_k)$. As established by proposition 2, if M is in $[\mathcal{G}(v, w_k)]_{\mathcal{C},=}$ then M is in $[\mathcal{G}(v, w)]_{\mathcal{C},=}$ (corresponds to cases 1.a and 2.a of table 1, *i.e.* first line, first column and second line, first column), if M is in $[\mathcal{G}(v, w_k)]_{=,=}$ then M is in $[\mathcal{G}(v, w)]_{\mathcal{C},=}$ if $\pi(w_k) \neq \pi(w)$ (corresponds to the case 1.c of table 1). In the other case, M is in $[\mathcal{G}(v, w_k)]_{=,\mathcal{C}} \cup [\mathcal{G}(v, w_k)]_{\mathcal{C},\mathcal{C}}$, and necessarily M is not in $[\mathcal{G}(v, w)]_{\mathcal{C},=}$ (corresponds to cases 1.b, 2.b, 1.d and 2.d of table1). The other assertions are determined using the same scheme.

□

The detailed proof is given in the appendix section. This recursive expression of EDM sets is used to derive a recursive expression of the distance between two quotiented trees.

4.3 Recursive expression of the distance between quotiented trees

We respectively denote by $[D(S_1[v], S_2[w])]_{\mathcal{C},=}$, $[D(S_1[v], S_2[w])]_{=,\mathcal{C}}$, $[D(S_1[v], S_2[w])]_{=,=}$ and $[D(S_1[v], S_2[w])]_{\mathcal{C},\mathcal{C}}$ the minimum cost $\Gamma_{v,w}(M)$ of EDMs M of sets $[\mathcal{M}(v, w)]_{\mathcal{C},=}$, $[\mathcal{M}(v, w)]_{=,\mathcal{C}}$, $[\mathcal{M}(v, w)]_{=,=}$ and $[\mathcal{M}(v, w)]_{\mathcal{C},\mathcal{C}}$.

Theorem 2 $D(G_1[v], G_2[w])$ can be computed recursively:

1. *Initialisation:*

$$\begin{aligned} D(\theta, \theta) &= 0 \\ D(H_1[v], \theta) &= \sum_{v_k \in \text{son}[v]} D(G_1[v_k], \theta), & D(G_1[v], \theta) &= D(H_1[v], \theta) + d(v, \lambda) \\ D(\theta, H_2[w]) &= \sum_{w_k \in \text{son}[w]} D(\theta, G_2[w_k]), & D(\theta, G_2[w]) &= D(\theta, H_2[w]) + d(\lambda, w) \end{aligned}$$

2. *Computation of the distance between quotiented trees:*

$$D(G_1[v], G_2[w]) = \min \begin{cases} D(\theta, G_2[w]) + \min_{w_k \in \text{son}[w]} \{D(G_1[v], G_2[w_k]) - D(\theta, G_2[w_k])\} \\ D(G_1[v], \theta) + \min_{v_k \in \text{son}[v]} \{D(G_1[v_k], G_2[w]) - D(G_1[v_k], \theta)\} \\ D(H_1[v], H_2[w]) + d(v, \lambda) + d(\lambda, w) \\ [D(H_1[v], H_2[w])]_{=,=} + d(v, w) \\ [D(H_1[v], H_2[w])]_{\mathcal{C},\mathcal{C}} + d(v, w) \end{cases}$$

The recursive relation for computing the partial distance which appears in the previous equation is given, with proof in the appendix. These results are summarized in figure 6 by a dependency graph showing how quantities are recursively dependent one upon the other in the computation of $D(G_1[v], G_2[w])$. This graph shows that the computation of $D(G_1[v], G_2[w])$ ultimately relies on the computation of special restricted EDMs, detailed in the following section.

figure 6

4.4 Restricted EDMs with minimum cost

From the definition of restricted EDM (proposition 6) the problem of finding the restricted EDM with minimum cost is related to the minimum cost bipartite matching problem. However, we have to compute several optimal restricted EDM depending on the image of $\pi(v)$ and $\pi(w)$. For each different case, we give a method for computing the optimal EDM, based on the modelling of Zhang [12, 8] as a minimum cost flow problem.

4.4.1 Modelling as a minimum cost flow problem

If $n_v = n_w$, this is exactly the minimum cost bipartite matching problem. If $n_v \neq n_w$, we have to consider the extra trees in one of the forests. Suppose that $n_v > n_w$. One way to solve this problem is to add $n_v - n_w$ null trees to $F_2[w]$ and then use a bipartite matching. However, this results in redundant computation. We can reduce this problem directly to the minimum cost maximum flow problem by adding only one null tree to $F_2[w]$.

Given two forests $F_1[v]$ and $F_2[w]$, we assume that $n_v > n_w$. Let $I = \{v_1, v_2, \dots, v_{n_v}\}$ and $J = \{w_1, w_2, \dots, w_{n_w}\}$, where v_k , $1 \leq k \leq n_v$, represents the tree $T_1[v_k]$ and w_k , $1 \leq k \leq n_w$, represents the tree $T_2[w_k]$. Let us construct a graph $R = (S, A)$ as follows :

- vertex set: $S = \{s, t, e\} \cup I \cup J$, where s is the source, t is the sink and e represents a null tree;
- edge set:

$$A = \left\{ \bigcup_{v_k \in \text{son}(v)} (s, v_k) \cup \bigcup_{v_k \in \text{son}(v)} (v_k, e) \cup \bigcup_{w_k \in \text{son}(w)} (w_k, t) \cup \bigcup_{v_k \in \text{son}(v)} \left(\bigcup_{w_l \in \text{son}(w)} (v_k, w_l) \right) \right\}$$

All the edges have capacity one, except (e, t) whose capacity is $n_v - n_w$.

R is a network with integer capacities and the maximum flow $f^* = n_v = \max\{n_v, n_w\}$.

In the original graph proposed by Zhang [12, 8], the cost $D(T_1[v_k], T_2[w_l])$ is attached to each edge (v_k, w_l) . However, finding an optimal matching in this case does not ensure that constraints (2) and (3) are satisfied for the quotient EDM. The sub-sections below show how we can modify the cost of edges to compute the optimal restricted EDM of $[\mathcal{R}(v, w)]_{C,=}$, $[\mathcal{R}(v, w)]_{=,C}$, $[\mathcal{R}(v, w)]_{=,=}$ and $[\mathcal{R}(v, w)]_{C,C}$. A representation of the network is given in figure 7 and 8.

figure 7 and 8

According to proposition 6, for any restricted EDM M , there exists a partition R of M and a matching K of $\mathcal{K}(v, w)$. We show in propositions 26-32 that there exists one and only one element M_i of R and a pair $(v_k, w_l) \in K$ such that M_i is a valid EDM in $\mathcal{G}(v_k, w_l)$ and for any other element $M_j \neq M_i$ of the partition R there exists a pair $(v_p, w_q) \neq (v_k, w_l)$ such that M_j is a valid EDM in $\mathcal{G}(v_p, w_q)$; both M_i and M_j belonging respectively to particular set of valid EDMs of the partition. The results of these propositions are summarized in table 5 and detailed in appendix section. Note that according to these propositions, if no M_i satisfies the condition (the particular sets of valid EDMs $\mathcal{G}(v_k, w_l)$ is empty) then $[\mathcal{R}(v, w)]_{\alpha,\beta}$ is necessarily empty.

According to these results, a cost is attached to each edge of the previous network. Next list gives the assigned cost to each edge of the network flow for computing $\left([\mathcal{R}(v, w)]_{C,C}\right)_{=,=}$:

- for any v_k and w_l such that $\pi(v_k) = \pi(v)$ and $\pi(w_l) = \pi(w)$:

$$\gamma(v_k, w_l) = \min \left\{ \left([D(T_1[v_k], T_2[w_l])]_{C,C} \right)_{=,=}, \left([D(T_1[v_k], T_2[w_l])]_{C,C} \right)_{C,C} \right\}$$

$M \in$	$M_i \in$			
	$\pi(v_k) = \pi(v)$ $\pi(w_l) = \pi(w)$	$\pi(v_k) = \pi(v)$ $\pi(w_l) \neq \pi(w)$	$\pi(v_k) \neq \pi(v)$ $\pi(w_l) = \pi(w)$	$\pi(v_k) \neq \pi(v)$ $\pi(w_l) \neq \pi(w)$
$[\mathcal{R}(v, w)]_{C,=}$	$[\mathcal{G}(v_k, w_l)]_{C,=}$	$[\mathcal{G}(v_k, w_l)]_{C,=}$ $[\mathcal{G}(v_k, w_l)]_{=,=}$		
$[\mathcal{R}(v, w)]_{=,C}$	$[\mathcal{G}(v_k, w_l)]_{=,C}$		$[\mathcal{G}(v_k, w_l)]_{=,C}$ $[\mathcal{G}(v_k, w_l)]_{=,=}$	
$([\mathcal{R}(v, w)]_{=,=})_{C,=}$	$([\mathcal{G}(v_k, w_l)]_{=,=})_{C,=}$	$([\mathcal{G}(v_k, w_l)]_{=,=})_{C,=}$ $([\mathcal{G}(v_k, w_l)]_{=,=})_{=,=}$		
$([\mathcal{R}(v, w)]_{=,=})_{=,C}$	$([\mathcal{G}(v_k, w_l)]_{=,=})_{=,C}$		$([\mathcal{G}(v_k, w_l)]_{=,=})_{=,C}$ $([\mathcal{G}(v_k, w_l)]_{=,=})_{=,=}$	
$([\mathcal{R}(v, w)]_{=,=})_{=,=}$	$([\mathcal{G}(v_k, w_l)]_{=,=})_{=,=}$			
$([\mathcal{R}(v, w)]_{=,=})_{C,C}$	$([\mathcal{G}(v_k, w_l)]_{=,=})_{C,C}$			
$([\mathcal{R}(v, w)]_{C,C})_{C,=}$	$([\mathcal{G}(v_k, w_l)]_{C,C})_{C,=}$	$([\mathcal{G}(v_k, w_l)]_{C,C})_{C,=}$ $([\mathcal{G}(v_k, w_l)]_{C,C})_{=,=}$		
$([\mathcal{R}(v, w)]_{C,C})_{=,C}$	$([\mathcal{G}(v_k, w_l)]_{C,C})_{=,C}$		$([\mathcal{G}(v_k, w_l)]_{C,C})_{=,C}$ $([\mathcal{G}(v_k, w_l)]_{C,C})_{=,=}$	
$([\mathcal{R}(v, w)]_{C,C})_{=,=}$	$([\mathcal{G}(v_k, w_l)]_{C,C})_{=,=}$	$[\mathcal{G}(v_p, w_q)]_{=,C}$ $([\mathcal{G}(v_p, w_q)]_{C,C})_{C,=}$ $([\mathcal{G}(v_p, w_q)]_{C,C})_{C,C}$	$[\mathcal{G}(v_p, w_q)]_{C,=}$ $([\mathcal{G}(v_p, w_q)]_{C,C})_{=,C}$ $([\mathcal{G}(v_p, w_q)]_{C,C})_{C,C}$	$\mathcal{G}(v_p, w_q)$
$([\mathcal{R}(v, w)]_{C,C})_{C,C}$	$([\mathcal{G}(v_k, w_l)]_{C,C})_{C,C}$	$[\mathcal{G}(v_p, w_q)]_{=,C}$ $([\mathcal{G}(v_p, w_q)]_{C,C})_{C,=}$ $([\mathcal{G}(v_p, w_q)]_{C,C})_{C,C}$	$[\mathcal{G}(v_p, w_q)]_{C,=}$ $([\mathcal{G}(v_p, w_q)]_{C,C})_{=,C}$ $([\mathcal{G}(v_p, w_q)]_{C,C})_{C,C}$	$\mathcal{G}(v_p, w_q)$

Table 5: For any M in $[\mathcal{R}(v, w)]_{\alpha, \beta}$, this table gives the memberships of M_i . An empty cell denotes the empty set. A cell which is composed with several sets, denotes the union of these sets.

$M \in$	$M_j \in$			
	$\pi(v_k) = \pi(v)$ $\pi(w_l) = \pi(w)$	$\pi(v_k) = \pi(v)$ $\pi(w_l) \neq \pi(w)$	$\pi(v_k) \neq \pi(v)$ $\pi(w_l) = \pi(w)$	$\pi(v_k) \neq \pi(v)$ $\pi(w_l) \neq \pi(w)$
$[\mathcal{R}(v, w)]_{C,=}$				
$[\mathcal{R}(v, w)]_{=,C}$				
$\left([\mathcal{R}(v, w)]_{=,=}\right)_{C,=}$	$\left([\mathcal{G}(v_p, w_q)]_{=,=}\right)_{\theta,\theta}$			
$\left([\mathcal{R}(v, w)]_{=,=}\right)_{=,C}$	$\left([\mathcal{G}(v_p, w_q)]_{=,=}\right)_{\theta,\theta}$			
$\left([\mathcal{R}(v, w)]_{=,=}\right)_{=,=}$	$\left([\mathcal{G}(v_p, w_q)]_{=,=}\right)_{=,=}$ $\left([\mathcal{G}(v_p, w_q)]_{=,=}\right)_{C,C}$ $\left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{=,=}$ $\left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{C,C}$ $\left([\mathcal{G}(v_p, w_q)]_{=,=}\right)_{\theta,\theta}$	$[\mathcal{G}(v_p, w_q)]_{=,C}$ $[\mathcal{G}(v_p, w_q)]_{C,C}$	$[\mathcal{G}(v_p, w_q)]_{C,=}$ $[\mathcal{G}(v_p, w_q)]_{C,C}$	$\mathcal{G}(v_p, w_q)$
$\left([\mathcal{R}(v, w)]_{=,=}\right)_{C,C}$	$\left([\mathcal{G}(v_p, w_q)]_{=,=}\right)_{\theta,\theta}$			
$\left([\mathcal{R}(v, w)]_{C,C}\right)_{C,=}$				
$\left([\mathcal{R}(v, w)]_{C,C}\right)_{=,C}$				
$\left([\mathcal{R}(v, w)]_{C,C}\right)_{=,=}$	$\left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{=,=}$ $\left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{C,C}$	$[\mathcal{G}(v_p, w_q)]_{=,C}$ $\left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{C,=}$ $\left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{C,C}$	$[\mathcal{G}(v_p, w_q)]_{C,=}$ $\left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{=,C}$ $\left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{C,C}$	$\mathcal{G}(v_p, w_q)$
$\left([\mathcal{R}(v, w)]_{C,C}\right)_{C,C}$				

Table 6: For any M in $[\mathcal{R}(v, w)]_{\alpha,\beta}$, this table gives the memberships of M_j . An empty cell denotes the empty set. A cell which is composed with several sets, denotes the union of these sets.

- for any v_k and w_l such that $\pi(v_k) = \pi(v)$ and $\pi(w_l) \neq \pi(w)$:

$$\gamma(v_k, w_l) = \min \left\{ [D(T_1[v_k], T_2[w_l])]_{=,C}, \left([D(T_1[v_k], T_2[w_l])]_{C,C} \right)_{C,C}, \left([D(T_1[v_k], T_2[w_l])]_{C,C} \right)_{=,C} \right\}$$

- for any v_k and w_l such that $\pi(v_k) \neq \pi(v)$ and $\pi(w_l) = \pi(w)$:

$$\gamma(v_k, w_l) = \min \left\{ [D(T_1[v_k], T_2[w_l])]_{C,=}, \left([D(T_1[v_k], T_2[w_l])]_{C,C} \right)_{C,C}, \left([D(T_1[v_k], T_2[w_l])]_{C,C} \right)_{C,=} \right\}$$

- for any v_k and w_l such that $\pi(v_k) \neq \pi(v)$ and $\pi(w_l) \neq \pi(w)$:

$$\gamma(v_k, w_l) = D(T_1[v_k], T_2[w_l])$$

- for any v_k in $son(v)$:

$$\gamma(v_k, \theta_1) = D(T_1[v_k], \theta_1)$$

- the other edge costs are null.

4.5 Algorithm and complexity

The following algorithm computes a distance between two quotiented trees:

Input: G_1 and G_2 .

Output : $D(G_1[x], G_2[y])$ for any $x \in G_1$ and $y \in G_2$.

$$D(\theta, \theta) = 0$$

For $v \in G_1$

For $w \in G_2$

$$\begin{aligned} D(\theta, \theta) &= 0 \\ D(H_1[v], \theta) &= \sum_{v_k \in son[v]} D(G_1[v_k], \theta), & D(G_1[v], \theta) &= D(H_1[v], \theta) + d(v, \lambda) \\ D(\theta, H_2[w]) &= \sum_{w_k \in son[w]} D(\theta, G_2[w_k]), & D(\theta, G_2[w]) &= D(\theta, H_2[w]) + d(\lambda, w) \end{aligned}$$

$$\text{computation of } \begin{cases} [D(H_1[v], H_2[w])]_{C,=} \\ [D(H_1[v], H_2[w])]_{=,C} \\ [D(H_1[v], H_2[w])]_{=,=} \text{ (according to propositions 22 and 23),} \\ [D(H_1[v], H_2[w])]_{C,C} \\ [D(H_1[v], H_2[w])]_{\emptyset, \emptyset} \end{cases}$$

$$\text{and then computation of } \begin{cases} [D(G_1[v], G_2[w])]_{C,=} \\ [D(G_1[v], G_2[w])]_{=,C} \\ [D(G_1[v], G_2[w])]_{=,=} \\ [D(G_1[v], G_2[w])]_{C,C} \\ [D(G_1[v], G_2[w])]_{\emptyset, \emptyset} \end{cases},$$

$$D(G_1[v], G_2[w]) = \min \begin{cases} D(\theta, G_2[w]) + \min_{w_k \in son[w]} \{D(G_1[v], G_2[w_k]) - D(\theta, G_2[w_k])\} \\ D(G_1[v], \theta) + \min_{v_k \in son[v]} \{D(G_1[v_k], G_2[w]) - D(G_1[v_k], \theta)\} \\ D(H_1[v], H_2[w]) + d(v, \lambda) + d(\lambda, w) \\ [D(H_1[v], H_2[w])]_{=,=} + d(v, w) \\ [D(H_1[v], H_2[w])]_{C,C} + d(v, w) \end{cases}$$

At one step of the recursion, *i.e.* for given v and w , the computation of terms 1 and 2 in $D(G_1[v], G_2[w])$ takes a time proportional to $n_v + n_w$. The computation of terms $D(H_1[v], H_2[w])$, $[D(H_1[v], H_2[w])]_{=,=}$, $[D(H_1[v], H_2[w])]_{C,C}$ relies on computation of minima which also takes a time proportional to $n_v + n_w$ and on the computation of costs of restricted EDMs.

The computation of the cost of a restricted EDM in $[\mathcal{R}(v, w)]_{C,C}$, uses a graph with integer capacities, nonnegative edge costs, and maximum flow $f^* = n_v + n_w$. The complexity of finding minimum cost maximum flow for such a graph, using the improvement proposed by Tarjan [19], is $O(m \times |f^*| \times \log_2(n))$ where m is the number of edges and n is the number of vertices. Here, $n = n_v + n_w + 4$ and $m = n_v \times n_w + 2n_v + 2n_w + 3$; therefore the complexity is $O(n_v \times n_w \times (n_v + n_w) \times \log_2(n_v + n_w))$. The case of $[\mathcal{R}(v, w)]_{=,=}$ is similar: as discussed in proposition 29, a total number of $\min\{|\Pi(v)|, |\Pi(w)|\}$ graphs of flow are used, where the number of edges is $m = (n_v - 1) \times n_w + 2n_v + 2n_w + 3 + |\Pi(w)|$ if $|\Pi(w)| \geq |\Pi(v)|$ and $m = n_v \times (n_w - 1) + 2n_v + 2n_w + 3 + |\Pi(v)|$ otherwise. The total complexity of the minimum cost maximum flow computation (here split into several sub-graphs) is $O(\min\{|\Pi(v)|, |\Pi(w)|\} \times n_v \times n_w \times (n_v + n_w) \times \log_2(n_v + n_w))$.

The overall complexity of the algorithm is thus :

$$O(|T_1| \times |T_2| \times (deg(T_1) + deg(T_2)) \times \min\{deg_{\pi_1}(T_1), deg_{\pi_2}(T_2)\} \times \log_2(deg(T_1) + deg(T_2)))$$

5 Conclusion

In this paper, we have extended an algorithm to compute a distance between unordered trees [8] and thus we have defined a distance between quotiented trees. The resulting algorithm computes this distance recursively in polynomial time, using the dynamic programming principle. The highest source of the complexity is due to a bipartite matching problem which occurs when comparing the forests rooted at two given vertices. We adapted a minimum cost maximum flow algorithm to take account of constraints derived from the quotiented structures. The final algorithm has the same complexity as Zhang's algorithm, multiplied by a factor $\min\{deg_{\pi_1}(T_1), deg_{\pi_2}(T_2)\}$ which expresses the mean number of components of a complex (*i.e.* of a macro-constituent).

This work is part of a project to develop computer tools for studying plant architecture [20, 21]. The proposed algorithm is currently integrated within tools dealing with the quantitative evaluation of plant similarity [9]. This algorithm opens new perspectives for the comparison of plant architectures by considering extensions of the algorithm to multiscale tree graphs [10, 9].

References

- [1] R. A. Wagner and M. J. Fisher, "The string-to-string correction problem," *Journal of the association for computing machinery*, vol. 21, pp. 168–173, 1974.
- [2] S. M. Selkow, "The tree-to-tree editing problem," *Information processing letters*, pp. 184–186, 1977.

- hal-00307409, version 1 - 8 Dec 2008
- [3] K.-C. Tai, "The tree-to-tree correction problem," *Journal of the Association for Computing Machinery*, pp. 422–433, 1979.
 - [4] S.-Y. Lu, "A tree-to-tree distance and its application to cluster analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, pp. 219–224, 1979.
 - [5] M. Farach and M. Thorup, "Sparse dynamic programming for evolutionary-tree comparison," *SIAM journal of computing*, vol. 26, no. 1, pp. 210–230, 1997.
 - [6] Y. Takajashi, Y. Satoh, H. Suzuki, and S. Sasaki, "Recognition of largest common structural fragment among a variety of chemical structures," *Anal. Sci*, vol. 3, pp. 23–28, 1987.
 - [7] B. A. Shapiro and K. Zhang, "Comparing multiple rna secondary structures using trees comparisons," *Cabios*, vol. 6, pp. 309–318, 1990.
 - [8] K. Zhang, "A constrained edit distance between unordered labeled trees," *Algorithmica*, vol. 15, pp. 205–222, 1996.
 - [9] P. Ferraro and C. Godin, "A distance measure between plant architectures," *Annals of Forest Science*, vol. 57, pp. 445–461, june 2000.
 - [10] C. Godin and Y. Caraglio, "A multiscale model of plant topological structures," *Journal of theoretical biology*, vol. 191, pp. 1–46, 1998.
 - [11] A. Ohmori and E. Tanaka, "A unified view on tree metrics," *Syntactic and Structural Pattern Recognition*, vol. 45, pp. 85–100, 1988.
 - [12] K. Zhang, "A new editing-based distance between unordered trees," in *Combinatorial Pattern Matching, 4th Ann. Symp.*, (Padala (Italy)), pp. 254–265, CPM'93, 1993.
 - [13] P. Ferraro, "Propriétés des alignements valides entre arborescences quotientées," tech. rep., Cirad, Plant modelling Program - Montpellier - France, 2000.
 - [14] K. Zhang and T. Jiang, "Some max snp-hard results concerning unordered labeled trees," *Information Processing Letters*, vol. 49, pp. 249–254, 1994.
 - [15] P. Kilpellainen and H. Mannila, "The tree inclusion problem," in *Proc. Internat. Joint Conf. on the Theory and Practice of Software*, vol. 1, pp. 202–214, 1991.
 - [16] E. Tanaka and K. Tanaka, "The tree-to-tree editing problem," *International journal Pattern Recognition And Atificial Intelligency*, vol. 2, no. 2, pp. 221–240, 1988.
 - [17] M. Minoux, *Mathematical Programming. Theory and algorithms*. A Wiley-Interscience Publication, 1986.
 - [18] C. Berge, *Hypergraphs*. Elsevier Science Publishers, 1989.
 - [19] R. E. Tarjan, *Data Structures and Network Algorithms*. CBMS-NFS - Regional Conference Series In Applied Mathematics, 1983.
 - [20] C. Godin, E. Costes, and Y. Caraglio, "Exploring plant topological structure with the amapmod software: an outline," *Silva Fennica*, vol. 31, pp. 355–366, 1997.

- [21] C. Godin, Y. Guédon, and E. Costes, "Exploration of a plant architecture database with the amamod software illustrated on an apple tree hybrid family," *Agronomie*, vol. 19, p. 163-184, March-May 1999.

Appendix: proof of propositions

Note that $\mathcal{G}(v, w) \subseteq \mathcal{T}(v, w)$, and then $\mathcal{G}(v, w)$ can be partitionned according to proposition 3 into sets $(\mathcal{G}(v, w))_{C,=}$, $(\mathcal{G}(v, w))_{=,C}$, $(\mathcal{G}(v, w))_{=,=}$, $(\mathcal{G}(v, w))_{C,C}$ and $(\mathcal{G}(v, w))_{\theta,\theta}$.

PROOF OF PROPOSITION 1. For any $S_1[x]$ and $S_1[y]$ in $\mathcal{S}_1(v)$ such that $S_1[x] \subseteq S_1[y]$, by definition, $y \leq x$: then

$$M_{w/y} = M_{w/x} \cup \{z_2 \in V_1[w] \mid \exists z_1 \in V_1[y] \setminus V_1[x]; (z_1, z_2) \in M\}$$

and thus $M_{w/x} \subseteq M_{w/y}$. Finally $\hat{y} \leq \hat{x}$:

$$\widehat{M}(S_1[x]) \subseteq \widehat{M}(S_1[y])$$

□

PROOF OF PROPOSITION 2. Let M be a valid EDM from $S_1[v]$ to $S_2[w]$, such that $\widehat{M}(S_1[x]) \subseteq \widehat{M}(S_1[y])$, following the definition 4 of \widehat{M} , there is two cases depending on the image $S_1[y]$:

1. $\widehat{M}(S_1[y])$ is a tree $T_2[t_2]$: there exists a vertex t_1 in $S_1[y]$ such that $(t_1, t_2) \in M$. If $\widehat{M}(S_1[x]) \neq \theta$ then for any vertex z_2 in $\widehat{M}(S_1[x])$, such that z_2 has an image z_1 in $S_1[x]$, z_2 is in $\widehat{M}(S_1[y])$, that is in $T_2[t_2]$. And according to constraint (2):

$$t_2 \leq z_2 \Leftrightarrow t_1 \leq z_1$$

Thus, $S_1[x] \subseteq S_1[y]$;

2. $\widehat{M}(S_1[y])$ is a forest $F_2[t_1 \wedge u_1]$: where t_1 and u_1 are two vertices t_1 and u_1 in $S_1[y]$ such that t_1 and u_1 have respectively an image t_2 and u_2 by M . If $\widehat{M}(S_1[x]) \neq \theta$ then for any vertex z_2 in $\widehat{M}(S_1[x])$, such that z_2 has an image z_1 in $S_1[x]$, z_2 is in $\widehat{M}(S_1[y])$, that is in $F_2[t_2]$. And according to constraint (3):

$$t_2 \wedge u_1 < z_2 \Leftrightarrow t_1 \wedge u_1 < z_1$$

Thus, $S_1[x] \subseteq S_1[y]$.

The reciprocal is due to proposition 1.

□

PROOF OF PROPOSITION 3.

1. If $M = \emptyset$ then $\widehat{M}(S_1[v]) = \theta$ and $\widehat{M}(S_2[w]) = \theta$;
2. else, $M \neq \emptyset$ and then necessarily $\widehat{M}(S_1[v]) \neq \theta$ and $\widehat{M}(S_2[w]) \neq \theta$.
According to proposition 2, $\widehat{M}(S_1[v]) \subseteq \widehat{M}(S_2[w])$ and $\widehat{M}(S_2[w]) \subseteq \widehat{M}(S_1[v])$, thus there is four cases:

(a) $\widehat{M}(S_1[v]) \subset S_2[w]$ and $\widehat{M}(S_2[w]) = S_1[v]$;

- (b) $\widehat{M}(S_1[v]) = S_2[w]$ and $\widehat{M}(S_2[w]) \subset S_1[v]$;
- (c) $\widehat{M}(S_1[v]) = S_2[w]$ and $\widehat{M}(S_2[w]) = S_1[v]$;
- (d) $\widehat{M}(S_1[v]) \subset S_2[w]$ and $\widehat{M}(S_2[w]) \subset S_1[v]$.

□

PROOF OF PROPOSITION 4. Let M be a valid EDM from $T_1[v]$ to $T_2[w]$, M is in $\mathcal{T}(v, w)_{=,=}$ if and only if $\widehat{M}(T_1[v]) = T_2[w]$ and $\widehat{M}(T_2[w]) = T_1[v]$. Then by the definition of \widehat{M} , v and w have both an image by M . By the the condition (2) of valid EDM, v and w are necessarily image on each other.

□

PROOF OF PROPOSITION 5. Obvious following definition of valid EDM from $F_1[v]$ to $F_2[w]$.

□

PROOF OF PROPOSITION 6. See [8].

□

PROOF OF PROPOSITION 7. This proposition is decomposed in two parts:

1. Let M be a valid EDM from $T_1[v]$ to $T_2[w]$, according to proposition 3 applied to trees:

- either $\widehat{M}(T_1[v]) \neq \theta$ and $\widehat{M}(T_2[w]) \neq \theta$ and:

- (a) $\widehat{M}(T_1[v]) \subset T_2[w]$ and $\widehat{M}(T_2[w]) = T_1[v]$, then there exists w_k , a son of w , such that $\widehat{M}(T_1[v]) \subseteq T_2[w_k]$. If there is a son w_l of w such that $\widehat{M}(T_2[w_l]) \neq \theta$, then necessarily $\widehat{M}(T_1[v]) = T_2[w]$ (contradiction with ou hypothesis), thus for any sons $w_l \neq w_k$ of w , $\widehat{M}(T_2[w_l]) = \theta$ and $\widehat{M}(T_2[w_k]) \neq \theta$: $M \in \mathcal{T}(v, w_k)$
- (b) $\widehat{M}(T_1[v]) = T_2[w]$ and $\widehat{M}(T_2[w]) \subset T_1[v]$, similar to the previous case;
- (c) $\widehat{M}(T_1[v]) = T_2[w]$ and $\widehat{M}(T_2[w]) = T_1[v]$, then proposition 4 can be applied and $M^* \in \mathcal{F}(v, w)$;
- (d) $\widehat{M}(S_1[v]) \subset S_2[w]$ and $\widehat{M}(S_2[w]) \subset S_1[v]$, then proposition 5 can be applied and $M \in \mathcal{F}(v, w)$.
- (e) or either $\widehat{M}(T_1[v]) = \theta$ and $\widehat{M}(T_2[w]) = \theta$ and then $M = \emptyset$;

2. Let M be a valid EDM from $F_1[v]$ to $F_2[w]$, according to proposition 3 applied to forests:

- either $\widehat{M}(F_1[v]) \neq \theta$ and $\widehat{M}(F_2[w]) \neq \theta$ and:

- (a) $\widehat{M}(F_1[v]) \subset F_2[w]$ and $\widehat{M}(F_2[w]) = F_1[v]$, then there exists w_k , a son of w , such that $\widehat{M}(F_1[v]) \subseteq F_2[w_k]$. If there is a son w_l of w such that $\widehat{M}(F_2[w_l]) \neq \theta$, then necessarily $\widehat{M}(F_1[v]) = F_2[w]$ (contradiction with ou hypothesis), thus for any sons $w_l \neq w_k$ of w , $\widehat{M}(F_2[w_l]) = \theta$ and $\widehat{M}(F_2[w_k]) \neq \theta$: $M \in \mathcal{F}(v, w_k)$
- (b) $\widehat{M}(F_1[v]) = F_2[w]$ and $\widehat{M}(F_2[w]) \subset F_1[v]$, similar to the previous case;

- (c) $\widehat{M}(F_1[v]) = F_2[w]$ and $\widehat{M}(F_2[w]) = F_1[v]$, or $\widehat{M}(F_1[v]) \subset F_2[w]$ and $\widehat{M}(F_2[w]) \subset F_1[v]$, then according to proposition 6, $M \in \mathcal{R}(v, w)$;
- (d) or either $\widehat{M}(F_1[v]) = \theta$ and $\widehat{M}(F_2[w]) = \theta$ and then $M = \emptyset$;

□

PROOF OF THEOREM 1. This theorem was firstly proved by Zhang in [12, 8].

□

PROOF OF LEMMA 1. The proof is similar to proof of Theorem 2 proposed by Zhang in [8]. A composition of EDMs can be defined as follow. Let M_1 be a valid EDM from T_1 to T_2 and let M_2 be a valid EDM from T_2 to T_3 , then:

$$M_1 \circ M_2 = \{(x, y) | \exists z \text{ s.t. } (x, z) \in M_1 \text{ and } (z, y) \in M_2\}$$

This definition is exactly similar to the definition proposed by Zhang. To proof that $M_1 \circ M_2$, we can use the same scheme given by Zhang in his lemma 2 [8]. To check conditions (4), (5) and (6), we thus just need to chose three pairs (x_1, z_1) , (x_2, z_2) and (x_3, z_3) in $M_1 \circ M_2$ such that $\pi(x_1)$, $\pi(x_2)$, $\pi(x_3)$ are different. Since M_1 and M_2 are valid EDMs, conditions (4) and (5) are obviously verified. Condition (6) is also verified if we applied the proof of Zhang on complexes $\pi(x_1)$, $\pi(x_2)$, $\pi(x_3)$. Furthermore, the proof of Zhang to show $\gamma(M_1 \circ M_2) \leq \gamma(M_1) + \gamma(M_2)$ can also be directly applied to quotiented tree graphs. Thus the same results can be applied in that case.

Now to show that D is a distance we need to prove the following relations:

1. $D(T_1[v], T_1[v]) = 0$;
2. $D(T_1[v], T_2[w]) = D(T_2[w], T_1[v])$;
3. $D(T_1[v], T_3[x]) \leq D(T_1[v], T_2[w]) + D(T_2[w], T_3[x])$.

Relations (1) and (2) are direct consequences of definition of valid EDMs. Furthermore, relation (3) is a result obtained from the definition of composition between EDMs (cf Zhang [8], theorem 2).

PROOF OF LEMMA 2. For any x and y in $S_1[v] \times S_2[w]$, let M be a valid EDM from $S_1[x]$ to $S_2[y]$ then M is a valid EDM from $S_1[v]$ to $S_2[w]$. Furthermore $\widehat{N}(S_1[\pi(v)]) = \widehat{N}(S_1[\pi(x)])$ and $\widehat{N}(S_2[\pi(w)]) = \widehat{N}(S_2[\pi(y)])$. According to proposition 3, there is three cases depending on the position of $S_1[\pi(x)]$ front of $S_2[\pi(y)]$:

1. $\widehat{N}(S_1[\pi(x)]) = \theta$ then necessarily, $\widehat{N}(S_1[\pi(v)]) = \theta$;
2. $\widehat{N}(S_1[\pi(x)]) \subset S_2[\pi(y)]$, then necessarily, $\widehat{N}(S_1[\pi(v)]) \subset S_2[\pi(y)] \subseteq S_2[\pi(w)] \Rightarrow \widehat{N}(S_1[\pi(v)]) \subset S_2[\pi(w)]$;
3. $\widehat{N}(S_1[\pi(x)]) = S_2[\pi(y)]$, then necessarily, $\widehat{N}(S_1[\pi(v)]) = S_2[\pi(y)]$ and if $\pi(y) = \pi(w)$ then $S_2[\pi(y)] = S_2[\pi(w)] \Rightarrow \widehat{N}(S_1[\pi(v)]) = S_2[\pi(w)]$. Else $\pi(y) > \pi(w)$ then $S_2[\pi(y)] \subset S_2[\pi(w)] \Rightarrow \widehat{N}(S_1[\pi(v)]) \subset S_2[\pi(w)]$.

□

PROOF OF PROPOSITION 8. Obvious from lemma 2.

□

PROOF OF PROPOSITION 9. Obvious from lemma 2.

□

PROOF OF PROPOSITION 10. Obvious from lemma 2.

□

PROOF OF LEMMA 3. For any x and y in $S_1[v] \times S_2[w]$, and for any valid EDM M in $\mathcal{M}(x, y)$ $\widehat{N}(S_1(\pi(v))) = \widehat{N}(S_1(\pi(x)))$. Furthermore, $\pi(w) \leq \pi(y)$ and then according to proposition 2, $S_2(\pi(y)) \subseteq S_2(\pi(w))$, thus if $\widehat{N}(S_1(\pi_1(x))) \subseteq S_2(\pi_2(y))$ then $\widehat{N}(S_1(\pi_1(v))) \subseteq S_2(\pi_2(w))$.

If $\pi(y) = \pi(w)$, then $S_2(\pi(y)) = S_2(\pi(w))$ and if $\widehat{N}(S_1(\pi_1(x))) = S_2(\pi_2(y))$ then $\widehat{N}(S_1(\pi_1(v))) = S_2(\pi_2(w))$.

Otherwise, $S_2(\pi(y)) \subset S_2(\pi(w))$, $\widehat{N}(S_1(\pi(x))) = S_2(\pi(y)) \subset S_2(\pi(w)) \Rightarrow \widehat{N}(S_1(\pi_1(v))) \subset S_2(\pi_2(w))$.

Let $M_1 \in \mathcal{M}(x, y)$ such that $\widehat{N}_1(S_1(\pi(v))) \subseteq S_2(\pi(w))$ and $\widehat{N}_1(S_2(\pi(w))) = S_1(\pi(v))$, then there exists $y' \geq y$ such that $\pi(y') \neq \pi(w)$ and $\widehat{N}_1(S_1(\pi(v))) \subseteq S_2(\pi(y'))$ and $\widehat{N}_1(S_2(\pi(y'))) = S_1(\pi(v))$. Let $M_2 \in \mathcal{M}(z, t)$ and let M be the union $M_1 \cup M_2$. Remark that $M_1 \cap M_2 = \emptyset$, then the image of $S_2(\pi(y'))$ by \widehat{N} is given by \widehat{N}_1 : $\widehat{N}(S_2(\pi(y'))) = \widehat{N}_1(S_2(\pi(y')))$. According to previous result $\widehat{N}(S_2(\pi(y'))) = S_1(\pi(v)) \Rightarrow \widehat{N}(S_1(\pi(v))) \subseteq S_2(\pi(y'))$. Let z' be a descendant of z such that $\pi(z') \neq \pi(v)$ and $\widehat{N}(S_1(\pi(z'))) \neq \theta$. \widehat{N} is an increasing function, $S_1(\pi(z')) \subseteq S_1(\pi(v)) \Leftrightarrow \widehat{N}(S_1(\pi(z'))) \subseteq \widehat{N}(S_1(\pi(v))) \subseteq S_2(\pi(y'))$, however $\widehat{N}(S_1(\pi(z'))) = \widehat{N}_2(S_1(\pi(z'))) \subseteq S_2(\pi(t'))$ where $t' \geq t$ such that $\pi(t') \neq \pi(w)$ or $\widehat{N}(S_1(\pi(z'))) = \widehat{N}_2(S_1(\pi(z'))) \subseteq S_2(\pi(w))$. In both cases : $\widehat{N}(S_1(\pi(z'))) \not\subseteq S_2(\pi(y'))$, this is a contradiction, thus $\widehat{N}(S_1(\pi(z'))) = \theta$ and $\widehat{N}_2(S_1(\pi(z'))) = \theta$. Finally $M = M_1$ and then $\widehat{N}(S_1(\pi(v))) \subseteq S_2(\pi(w))$ and $\widehat{N}(S_2(\pi(w))) = S_1(\pi(v))$, this means $M \in \mathcal{M}_{C,=}(v, w)$.

Following previous results, necessarily, M_2 is not a valid EDM from $S_1[z]$ to $S_2[t]$ such that $\widehat{N}_2(S_1(\pi(v))) \subseteq S_2(\pi(w))$ and $\widehat{N}_2(S_2(\pi(w))) = S_1(\pi(v))$ (and conversely). And M_2 must to satisfy one of the previous properties.

Reciprocal : obvious.

□

PROOF OF PROPOSITION 11. The proof of this proposition is based on propositions 7 and lemma 2 and 3. Let M be a valid EDM of $[\mathcal{G}(v, w)]_{C,=}$, then according to proposition 7 ($\mathcal{G}(v, w) \subseteq \mathcal{T}(v, w)$), M satisfies one and only one of the five assertions:

1. $\exists w_k \in \text{son}[w]$ such that $M \in \mathcal{G}(v, w_k)$. As established by proposition 2, M is in $[\mathcal{G}(v, w)]_{C,=}$ if and only if M is in $[\mathcal{G}(v, w_k)]_{C,=}$ (corresponds to cases 1.a and 2.a of the table 2), or if M is in $[\mathcal{G}(v, w_k)]_{=,=}$ and $\pi(w_k) \neq \pi(w)$ (corresponds to the case 1.c of the array). In the other cases, M is in $[\mathcal{G}(v, w_k)]_{=,C} \cup [\mathcal{G}(v, w_k)]_{C,C}$, and necessarily M is not in $[\mathcal{G}(v, w)]_{C,=}$ (corresponds to cases 1.b, 2.b, 1.d and 2.d of the array);

2. $\exists v_k \in \text{son}[v]$ such that $M \in \mathcal{G}(v_k, w)$. As established by proposition 2, M is in $[\mathcal{G}(v, w)]_{C,=}$ if and only if M is in $[\mathcal{G}(v_k, w)]_{C,=}$ and $\pi(v_k) = \pi(v)$;
3. $(v, w) \in M$ and $M^* \in \mathcal{H}(v, w)$, then $\widehat{N}(T_1(\pi(v))) = T_2(\pi(w))$ and $\widehat{N}(T_2(\pi(w))) = T_1(\pi(v))$, and then M is not in $[\mathcal{G}(v, w)]_{C,=}$;
4. $M^* \in \mathcal{H}(v, w)$, and necessarily M is in $[\mathcal{G}(v, w)]_{C,=}$ if and only if M^* is in $[\mathcal{H}(v, w)]_{C,=}$;
5. $M = \emptyset$.

□

$[\mathcal{G}(v, w)]_{=,C}$ is determined symmetrically from $[\mathcal{G}(v, w)]_{C,=}$.

Proposition 12 *Let M be a valid EDM from $G_1[v]$ to $G_2[w]$ of $[\mathcal{G}(v, w)]_{=,=}$, then M satisfies one of the follows assertions :*

1. $\exists w_k \in \text{son}[w]$ such that $\pi(w_k) = \pi(w)$ and $M \in [\mathcal{G}(v, w_k)]_{=,=}$;
2. $\exists v_k \in \text{son}[v]$ such that $\pi(v_k) = \pi(v)$ and $M \in [\mathcal{G}(v_k, w)]_{=,=}$;
3. $(v, w) \in M$ and $M^* \in [\mathcal{H}(v, w)]_{=,=} \cup [\mathcal{H}(v, w)]_{C,C}$;
4. $M \in [\mathcal{H}(v, w)]_{=,=}$;
5. $M = \emptyset$.

PROOF. the proof is similar to the proof of proposition 11.

□

Proposition 13 *Let M be a valid EDM from $G_1[v]$ to $G_2[w]$ of $\left([\mathcal{G}(v, w)]_{=,=}\right)_{C,=}$, then M satisfies one of the follows assertions :*

1. $\exists w_k \in \text{son}[w]$ such that $\pi(w_k) = \pi(w)$ and $M \in \left([\mathcal{G}(v, w_k)]_{=,=}\right)_{C,=}$;
2. $\exists v_k \in \text{son}[v]$ such that $\pi(v_k) = \pi(v)$ and $M \in \left([\mathcal{G}(v_k, w)]_{=,=}\right)_{C,=}$;
3. $(v, w) \in M$ and $M^* \in \left([\mathcal{H}(v, w)]_{=,=}\right)_{C,=} \cup \left([\mathcal{H}(v, w)]_{C,C}\right)_{C,=}$;
4. $M \in \left([\mathcal{H}(v, w)]_{=,=}\right)_{C,=}$;
5. $M = \emptyset$.

PROOF. the proof is similar to the proof of proposition 11.

□

$\left([\mathcal{G}(v, w)]_{=,=}\right)_{=, \mathcal{C}}$ is determined symmetrically from $\left([\mathcal{G}(v, w)]_{=,=}\right)_{\mathcal{C}, =}$.

In the following, $\left([\mathcal{G}(v, w)]_{=,=}\right)_{=, =}$ denotes $\left([\mathcal{G}(v, w)]_{=,=}\right)_{=, =} \cup \left([\mathcal{G}(v, w)]_{=,=}\right)_{\mathcal{C}, \mathcal{C}}$.

Proposition 14 *Let M be a valid EDM from $G_1[v]$ to $G_2[w]$ of $\left([\mathcal{G}(v, w)]_{=,=}\right)_{=, =}$, then M satisfies one of the follows assertions :*

1. $\exists w_k \in \text{son}[w]$ such that $\pi(w_k) = \pi(w)$ and $M \in \left([\mathcal{G}(v, w_k)]_{=,=}\right)_{=, =}$;
2. $\exists v_k \in \text{son}[v]$ such that $\pi(v_k) = \pi(v)$ and $M \in \left([\mathcal{G}(v_k, w)]_{=,=}\right)_{=, =}$;
3. $(v, w) \in M$ and $M^* \in \left([\mathcal{H}(v, w)]_{=,=}\right)_{=, =} \cup \left([\mathcal{H}(v, w)]_{\mathcal{C}, \mathcal{C}}\right)_{=, =}$;
4. $M \in \left([\mathcal{H}(v, w)]_{=,=}\right)_{=, =}$;
5. $M = \emptyset$.

PROOF. the proof is similar to the proof of proposition 11.

□

Proposition 15 *Let M be a valid EDM from $G_1[v]$ to $G_2[w]$ of $[\mathcal{G}(v, w)]_{\mathcal{C}, \mathcal{C}}$, then M satisfies one of the follows assertions :*

1. $M \in [\mathcal{H}(v, w)]_{\mathcal{C}, \mathcal{C}}$;
2. $M = \emptyset$.

PROOF. the proof is similar to the proof of proposition 11.

□

Proposition 16 *Let M be a valid EDM from $G_1[v]$ to $G_2[w]$ of $\left([\mathcal{G}(v, w)]_{\mathcal{C}, \mathcal{C}}\right)_{\mathcal{C}, =}$, then M satisfies one of the follows assertions :*

1. $M \in \left([\mathcal{H}(v, w)]_{\mathcal{C}, \mathcal{C}}\right)_{\mathcal{C}, =}$;
2. $M = \emptyset$.

PROOF. the proof is similar to the proof of proposition 11.

□

$([\mathcal{G}(v, w)]_{\mathcal{C}, \mathcal{C}})_{=, \mathcal{C}}$ is determined symmetrically from $([\mathcal{G}(v, w)]_{\mathcal{C}, \mathcal{C}})_{\mathcal{C}, =}$.

In the following, $([\mathcal{G}(v, w)]_{\mathcal{C}, \mathcal{C}})_{=, =}$ denotes $([\mathcal{G}(v, w)]_{\mathcal{C}, \mathcal{C}})_{=, =} \cup ([\mathcal{G}(v, w)]_{\mathcal{C}, \mathcal{C}})_{\mathcal{C}, \mathcal{C}}$.

Proposition 17 *Let M be a valid EDM from $G_1[v]$ to $G_2[w]$ of $([\mathcal{G}(v, w)]_{\mathcal{C}, \mathcal{C}})_{=, =}$, then M satisfies one of the follows assertions :*

1. $M \in ([\mathcal{H}(v, w)]_{\mathcal{C}, \mathcal{C}})_{=, =}$;
2. $M = \emptyset$.

PROOF. the proof is similar to the proof of proposition 11.

□

Proposition 18 *Let M be a valid EDM from $H_1[v]$ to $H_2[w]$ of $[\mathcal{H}(v, w)]_{\mathcal{C}, =}$, then M satisfies one of the follows assertions:*

1. $\exists w_k \in \text{son}[w]$ such that:
 - (a) $\pi(w_k) \neq \pi(w)$ and $M \in [\mathcal{H}(v, w_k)]_{\mathcal{C}, =} \cup [\mathcal{H}(v, w_k)]_{=, =}$ or;
 - (b) $\pi(w_k) = \pi(w)$ and $M \in [\mathcal{H}(v, w_k)]_{\mathcal{C}, =}$;
2. $\exists v_k \in \text{son}[v]$ such that $\pi(v_k) = \pi(v)$ and $M \in [\mathcal{H}(v_k, w)]_{\mathcal{C}, =}$;
3. $M \in [\mathcal{R}(v, w)]_{\mathcal{C}, =}$
4. $M = \emptyset$

PROOF. the proof is similar to the proof of proposition 11.

□

$[\mathcal{H}(v, w)]_{=, \mathcal{C}}$ is determined symmetrically from $[\mathcal{H}(v, w)]_{\mathcal{C}, =}$.

Proposition 19 *Let M be a valid EDM from $H_1[v]$ to $H_2[w]$ of $[\mathcal{H}(v, w)]_{=, =}$, then M satisfies one of the follows assertions :*

1. $\exists w_k \in \text{son}[w]$ such that $\pi(w_k) = \pi(w)$ and $M \in [\mathcal{H}(v, w_k)]_{=, =}$;
2. $\exists v_k \in \text{son}[v]$ such that $\pi(v_k) = \pi(v)$ and $M \in [\mathcal{H}(v_k, w)]_{=, =}$;
3. $M \in [\mathcal{R}(v, w)]_{=, =}$
4. $M = \emptyset$

PROOF. the proof is similar to the proof of proposition 11.

□

Proposition 20 Let M be a valid EDM from $H_1[v]$ to $H_2[w]$ of $([\mathcal{H}(v, w)]_{=,=})_{C,=}$, then M satisfies one of the follows assertions :

1. $\exists w_k \in \text{son}[w]$ such that $\pi(w_k) = \pi(w)$ and $M \in ([\mathcal{H}(v, w_k)]_{=,=})_{C,=}$;
2. $\exists v_k \in \text{son}[v]$ such that $\pi(v_k) = \pi(v)$ and $M \in ([\mathcal{H}(v_k, w)]_{=,=})_{C,=}$;
3. $M \in ([\mathcal{R}(v, w)]_{=,=})_{C,=}$
4. $M = \emptyset$

PROOF. the proof is similar to the proof of proposition 11.

□

Proposition 21 Let M be a valid EDM from $H_1[v]$ to $H_2[w]$ of $([\mathcal{H}(v, w)]_{=,=})_{=,=}$, then M satisfies one of the follows assertions :

1. $\exists w_k \in \text{son}[w]$ such that $\pi(w_k) = \pi(w)$ and $M \in ([\mathcal{H}(v, w_k)]_{=,=})_{=,=}$;
2. $\exists v_k \in \text{son}[v]$ such that $\pi(v_k) = \pi(v)$ and $M \in ([\mathcal{H}(v_k, w)]_{=,=})_{=,=}$;
3. $M \in ([\mathcal{R}(v, w)]_{=,=})_{=,=}$;
4. $M = \emptyset$.

PROOF. the proof is similar to the proof of proposition 11.

□

Proposition 22 Let M be a valid EDM from $H_1[v]$ to $H_2[w]$ of $[\mathcal{H}(v, w)]_{C,C}$, then M satisfies one of the follows assertions :

1. $\exists w_k \in \text{son}[w]$ such that:
 - (a) $M \in [\mathcal{H}(v, w_k)]_{C,C}$ or;
 - (b) $\pi(w_k) \neq \pi(w)$ and $M \in [\mathcal{H}(v, w_k)]_{=,C}$;
2. $\exists v_k \in \text{son}[v]$ such that:
 - (a) $M \in [\mathcal{H}(v_k, w)]_{C,C}$ or;

(b) $\pi(v_k) \neq \pi(v)$ and $M \in [\mathcal{H}(v_k, w)]_{C,=}$;

3. $M \in [\mathcal{R}(v, w)]_{C,C}$;

4. $M = \emptyset$.

PROOF. the proof is similar to the proof of proposition 11.

□

Proposition 23 Let M be a valid EDM from $H_1[v]$ to $H_2[w]$ of $([\mathcal{H}(v, w)]_{C,C})_{C,=}$, then M satisfies one of the follows assertions :

1. $\exists w_k \in \text{son}[w]$ such that:

(a) $\pi(w_k) = \pi(w)$ and $M \in ([\mathcal{H}(v, w_k)]_{C,C})_{C,=}$ or;

(b) $\pi(w_k) \neq \pi(w)$ and $M \in ([\mathcal{H}(v, w_k)]_{C,C})_{C,=} \cup ([\mathcal{H}(v, w_k)]_{C,C})_{=,=}$;

2. $\exists v_k \in \text{son}[v]$ such that $\pi(v_k) = \pi(v)$ and $M \in ([\mathcal{H}(v_k, w)]_{C,C})_{C,=}$;

3. $M \in ([\mathcal{R}(v, w)]_{C,C})_{C,=}$;

4. $M = \emptyset$.

PROOF. the proof is similar to the proof of proposition 11.

□

Proposition 24 Let M be a valid EDM from $H_1[v]$ to $H_2[w]$ of $([\mathcal{H}(v, w)]_{C,C})_{=,=}$, then M satisfies one of the follows assertions :

1. $\exists w_k \in \text{son}[w]$ such that:

(a) $\pi(w_k) = \pi(w)$ and $M \in ([\mathcal{H}(v, w_k)]_{C,C})_{=,=}$ or;

(b) $\pi(w_k) \neq \pi(w)$ and $M \in ([\mathcal{H}(v, w_k)]_{C,C})_{=,C} \cup [\mathcal{H}(v, w_k)]_{=,C}$;

2. $\exists v_k \in \text{son}[v]$ such that:

(a) $\pi(v_k) = \pi(v)$ and $M \in ([\mathcal{H}(v_k, w)]_{C,C})_{=,=}$;

(b) $\pi(v_k) \neq \pi(v)$ and $M \in ([\mathcal{H}(v_k, w)]_{C,C})_{C,=} \cup [\mathcal{H}(v_k, w)]_{C,=}$

$$3. M \in \left([\mathcal{R}(v, w)]_{C, C} \right)_{=, =};$$

$$4. M = \emptyset.$$

PROOF. the proof is similar to the proof of proposition 11.

□

Theorem 3 Let v_1, v_2, \dots, v_n be the sons of v and let w_1, w_2, \dots, w_p be the sons of w , then:

$$D(G_1[v], G_2[w]) = \min \begin{cases} D(\theta, G_2[w]) + \min_{w_k \in \text{son}[w]} \{D(G_1[v], G_2[w_k]) - D(\theta, G_2[w_k])\} \\ D(G_1[v], \theta) + \min_{v_k \in \text{son}[v]} \{D(G_1[v_k], G_2[w]) - D(G_1[v_k], \theta)\} \\ [D(H_1[v], H_2[w])]_{C, =} + d(v, \lambda) + d(\lambda, w) \\ [D(H_1[v], H_2[w])]_{=, C} + d(v, \lambda) + d(\lambda, w) \\ [D(H_1[v], H_2[w])]_{C, C} + d(v, w) \\ [D(H_1[v], H_2[w])]_{=, =} + d(v, w) \end{cases}$$

Lemma 4 Let v_1, v_2, \dots, v_n be the sons of v and let w_1, w_2, \dots, w_p be the sons of w , then:

$$[D(H_1[v], H_2[w])]_{C, =} = \min \begin{cases} D(H_1[v], \theta) + \min_{w_k \in \text{son}[w]} \left\{ [D(H_1[v], H_2[w_k])]_{C, =} - D(\theta, H_2[w_k]) \right\} \\ D(H_1[v], \theta) + \min_{w_k | \pi(w_k) \neq \pi(w)} \left\{ [D(H_1[v], H_2[w_k])]_{=, =} - D(\theta, H_2[w_k]) \right\} \\ D(\theta, H_2[w]) + \min_{v_k | \pi(v_k) \neq \pi(v)} \left\{ [D(H_1[v_k], H_2[w])]_{C, =} - D(H_1[v_k], \theta) \right\} \\ \min_{R \in [\mathcal{R}(v, w)]_{=, =}} \{ \Gamma_{v, w}(R) \} \end{cases}$$

Lemma 5 Let v_1, v_2, \dots, v_n be the sons of v and let w_1, w_2, \dots, w_p be the sons of w , then:

$$[D(H_1[v], H_2[w])]_{=, =} = \min \begin{cases} \left[[D(H_1[v], H_2[w])]_{=, =} \right]_{C, =} \\ \left[[D(H_1[v], H_2[w])]_{=, =} \right]_{=, C} \\ \left[[D(H_1[v], H_2[w])]_{=, =} \right]_{=, =} \\ \left[[D(H_1[v], H_2[w])]_{=, =} \right]_{\theta, \theta} \end{cases}$$

$$\left([D(H_1[v], H_2[w])]_{=, =} \right)_{C, =} = \min \begin{cases} D(\theta, H_2[w]) + \min_{w_k | \pi(w_k) = \pi(w)} \left\{ \left([D(H_1[v], H_2[w_k])]_{=, =} \right)_{C, =} - D(\theta, H_2[w_k]) \right\} \\ D(H_1[v], \theta) + \min_{v_k | \pi(v_k) = \pi(v)} \left\{ \left([D(H_1[v_k], H_2[w])]_{=, =} \right)_{C, =} - D(H_1[v_k], \theta) \right\} \\ \min_{R \in [\mathcal{R}(v, w)]_{=, =}} \{ \Gamma_{v, w}(R) \} \end{cases}$$

$$\begin{aligned}
& \left([D(H_1[v], H_2[w])]_{=,=} \right)_{=,=} \\
&= \min \left\{ \begin{array}{l} D(\theta, H_2[w]) + \min_{w_k | \pi(w_k) = \pi(w)} \left\{ \left([D(H_1[v], H_2[w_k])]_{=,=} \right)_{=,=} - D(\theta, H_2[w_k]) \right\} \\ D(H_1[v], \theta) + \min_{v_k | \pi(v_k) = \pi(v)} \left\{ \left([D(H_1[v_k], H_2[w])]_{=,=} \right)_{=,=} - D(H_1[v_k], \theta) \right\} \\ \min_{R \in ([\mathcal{R}(v,w)]_{=,=})_{=,=}} \{ \Gamma_{v,w}(R) \} \end{array} \right. \\
& \left([D(H_1[v], H_2[w])]_{=,=} \right)_{\theta, \theta} \\
&= \min \left\{ \begin{array}{l} D(\theta, H_2[w]) + \min_{w_k | \pi(w_k) = \pi(w)} \left\{ \left([D(H_1[v], H_2[w_k])]_{=,=} \right)_{\theta, \theta} - D(\theta, H_2[w_k]) \right\} \\ D(H_1[v], \theta) + \min_{v_k | \pi(v_k) = \pi(v)} \left\{ \left([D(H_1[v_k], H_2[w])]_{=,=} \right)_{\theta, \theta} - D(H_1[v_k], \theta) \right\} \\ \min_{R \in ([\mathcal{R}(v,w)]_{=,=})_{\theta, \theta}} \{ \Gamma_{v,w}(R) \} \end{array} \right.
\end{aligned}$$

Lemma 6 Let v_1, v_2, \dots, v_n be the sons of v and let w_1, w_2, \dots, w_p be the sons of w , then:

$$\begin{aligned}
& [D(H_1[v], H_2[w])]_{C,C} = \min \left\{ \begin{array}{l} \left([D(H_1[v], H_2[w])]_{C,C} \right)_{C,=} \\ \left([D(H_1[v], H_2[w])]_{C,C} \right)_{=,C} \\ \left([D(H_1[v], H_2[w])]_{C,C} \right)_{=,=} \end{array} \right. \\
& \left([D(H_1[v], H_2[w])]_{C,C} \right)_{C,=} \\
&= \min \left\{ \begin{array}{l} D(\theta, H_2[w]) + \min_{w_k \in \text{son}[w]} \left\{ \left([D(H_1[v], H_2[w_k])]_{C,C} \right)_{C,=} - D(\theta, H_2[w_k]) \right\} \\ D(\theta, H_2[w]) + \min_{\pi(w_k) \neq \pi(w)} \left\{ \left([D(H_1[v], H_2[w_k])]_{C,C} \right)_{=,=} - D(\theta, H_2[w_k]) \right\} \\ D(H_1[v], \theta) + \min_{\pi(v_k) \neq \pi(v)} \left\{ \left([D(H_1[v_k], H_2[w])]_{C,C} \right)_{C,=} - D(H_1[v_k], \theta) \right\} \\ \min_{R \in ([\mathcal{R}(v,w)]_{C,C})_{C,=}} \{ \Gamma_{v,w}(R) \} \end{array} \right.
\end{aligned}$$

$$\begin{aligned}
& \left([D(H_1[v], H_2[w])]_{C,C} \right)_{=,=} \\
&= \min \left\{ \begin{array}{l} D(\theta, H_2[w]) + \min_{\pi(w_k) = \pi(w)} \left\{ \left([D(H_1[v], H_2[w_k])]_{C,C} \right)_{=,=} - D(\theta, H_2[w_k]) \right\} \\ D(\theta, H_2[w]) + \min_{\pi(w_k) \neq \pi(w)} \left\{ \left([D(H_1[v], H_2[w_k])]_{C,C} \right)_{=,C} - D(\theta, H_2[w_k]) \right\} \\ D(\theta, H_2[w]) + \min_{\pi(w_k) \neq \pi(w)} \left\{ [D(H_1[v], H_2[w_k])]_{=,C} - D(\theta, H_2[w_k]) \right\} \\ D(H_1[v], \theta) + \min_{\pi(v_k) = \pi(v)} \left\{ \left([D(H_1[v_k], H_2[w])]_{C,C} \right)_{=,=} - D(H_1[v_k], \theta) \right\} \\ D(H_1[v], \theta) + \min_{\pi(v_k) \neq \pi(v)} \left\{ \left([D(H_1[v_k], H_2[w])]_{C,C} \right)_{C,=} - D(H_1[v_k], \theta) \right\} \\ D(H_1[v], \theta) + \min_{\pi(v_k) \neq \pi(v)} \left\{ [D(H_1[v_k], H_2[w])]_{C,C} - D(H_1[v_k], \theta) \right\} \\ \min_{R \in ([\mathcal{R}(v,w)]_{C,C})_{=,=}} \{ \Gamma_{v,w}(R) \} \end{array} \right.
\end{aligned}$$

Lemma 7 Let v_1, v_2, \dots, v_n be the sons of v and let w_1, w_2, \dots, w_p be the sons of w , then:

$$\begin{aligned}
& [D(G_1[v], G_2[w])]_{C,=} \\
&= \min \left\{ \begin{array}{l} D(\theta, G_2[w]) + \min_{w_k \in \text{son}[w]} \left\{ [D(G_1[v], T_2[w_k])]_{C,=} - D(\theta, G_2[w_k]) \right\} \\ D(\theta, G_2[w]) + \min_{w_k | \pi(w_k) \neq \pi(w)} \left\{ [D(G_1[v], G_2[w_k])]_{=,=} - D(\theta, G_2[w_k]) \right\} \\ [D(H_1[v], H_2[w])]_{C,=} + d(v, \lambda) + d(\lambda, w) \end{array} \right\} \\
& [D(G_1[v], G_2[w])]_{=,=} \\
&= \min \left\{ \begin{array}{l} D(\theta, G_2[w]) + \min_{w_k | \pi(w_k) = \pi(w)} \left\{ [D(G_1[v], G_2[w_k])]_{=,=} - D(\theta, G_2[w_k]) \right\} \\ D(G_1[v], \theta) + \min_{v_k | \pi(v_k) = \pi(v)} \left\{ [D(G_1[v_k], G_2[w])]_{=,=} - D(G_1[v_k], \theta) \right\} \\ [D(H_1[v], H_2[w])]_{=,=} + d(v, w) \\ [D(H_1[v], H_2[w])]_{C,C} + d(v, w) \end{array} \right\} \\
& \left([D(G_1[v], G_2[w])]_{=,=} \right)_{C,=} \\
&= \min \left\{ \begin{array}{l} D(\theta, G_2[w]) + \min_{w_k | \pi(w_k) = \pi(w)} \left\{ \left([D(G_1[v], G_2[w_k])]_{=,=} \right)_{C,=} - D(\theta, G_2[w_k]) \right\} \\ D(G_1[v], \theta) + \min_{v_k | \pi(v_k) = \pi(v)} \left\{ \left([D(G_1[v_k], G_2[w])]_{=,=} \right)_{C,=} - D(G_1[v_k], \theta) \right\} \\ \left([D(H_1[v], H_2[w])]_{=,=} \right)_{C,=} + d(v, w) \\ \left([D(H_1[v], H_2[w])]_{C,C} \right)_{C,=} + d(v, w) \end{array} \right\} \\
& \left([D(G_1[v], G_2[w])]_{=,=} \right)_{=,=} \\
&= \min \left\{ \begin{array}{l} D(\theta, G_2[w]) + \min_{w_k | \pi(w_k) = \pi(w)} \left\{ \left([D(G_1[v], G_2[w_k])]_{=,=} \right)_{=,=} - D(\theta, G_2[w_k]) \right\} \\ D(G_1[v], \theta) + \min_{v_k | \pi(v_k) = \pi(v)} \left\{ \left([D(G_1[v_k], G_2[w])]_{=,=} \right)_{=,=} - D(G_1[v_k], \theta) \right\} \\ \left([D(H_1[v], H_2[w])]_{=,=} \right)_{=,=} + d(v, w) \\ \left([D(H_1[v], H_2[w])]_{C,C} \right)_{=,=} + d(v, w) \end{array} \right\} \\
& [D(G_1[v], G_2[w])]_{C,C} = [D(H_1[v], H_2[w])]_{C,C} + d(v, \lambda) + d(\lambda, w) \\
& \left([D(G_1[v], G_2[w])]_{C,C} \right)_{C,=} = \left([D(H_1[v], H_2[w])]_{C,C} \right)_{C,=} + d(v, \lambda) + d(\lambda, w) \\
& \left([D(G_1[v], G_2[w])]_{C,C} \right)_{=,=} = \left([D(H_1[v], H_2[w])]_{C,C} \right)_{=,=} + d(v, \lambda) + d(\lambda, w)
\end{aligned}$$

Proves of following propositions are a direct consequence of propositions 8, 9 and 10.

Proposition 25 For any restricted EDM M in $[\mathcal{R}(v, w)]_{C,=}$, there exists v_k and w_l sons of v and w such that:

- $\pi(v_k) = \pi(v)$, $\pi(w_l) = \pi(w)$ and $M \in [\mathcal{G}(v_k, w_l)]_{C,=}$ or;

- $\pi(v_k) = \pi(v)$, $\pi(w_l) \neq \pi(w)$ and $M \in [\mathcal{G}(v_k, w_l)]_{C,=} \cup [\mathcal{G}(v_k, w_l)]_{=,=}$.

$[\mathcal{R}(v, w)]_{=,C}$ is determined by symmetry.

Proposition 26 For any restricted EDM M in $\left([\mathcal{R}(v, w)]_{C,C}\right)_{C,=}$, there exists v_k and w_l sons of v and w such that:

- $\pi(v_k) = \pi(v)$, $\pi(w_l) = \pi(w)$ and $M \in \left([\mathcal{G}(v_k, w_l)]_{C,C}\right)_{C,=}$ or;
- $\pi(v_k) = \pi(v)$, $\pi(w_l) \neq \pi(w)$ and $M \in \left([\mathcal{G}(v_k, w_l)]_{C,C}\right)_{C,=} \cup \left([\mathcal{G}(v_k, w_l)]_{C,C}\right)_{=,=}$.

$\left([\mathcal{R}(v, w)]_{C,C}\right)_{=,C}$ is determined by symmetry.

Proposition 27 For any restricted EDM M in $\left([\mathcal{R}(v, w)]_{C,C}\right)_{=,=}$, there exists a partition R of M and a matching K of $\mathcal{K}(v, w)$ such that there exists M' in $\left([\mathcal{G}(v_k, w_l)]_{C,C}\right)_{=,=}$ with $\pi(v_k) = \pi(v)$, $\pi(w_l) = \pi(w)$ and for any elements M'' of the partition R there exists a pair $(v_p, w_q) \in K$:

- if $\pi(v_q) = \pi(v)$ and $\pi(w_p) = \pi(w)$ then:

$$M' \in \left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{=,=} \cup \left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{C,C}$$

- if $\pi(v_q) = \pi(v)$ and $\pi(w_p) \neq \pi(w)$ then:

$$M'' \in [\mathcal{G}(v_p, w_q)]_{=,C} \cup \left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{C,C} \cup \left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{=,C}$$

- if $\pi(v_q) \neq \pi(v)$ and $\pi(w_p) = \pi(w)$ then:

$$M'' \in [\mathcal{G}(v_p, w_q)]_{C,=} \cup \left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{C,C} \cup \left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{C,=}$$

- if $\pi(v_q) \neq \pi(v)$ and $\pi(w_p) \neq \pi(w)$ then $M'' \in \mathcal{G}(v_p, w_q)$.

Proposition 28 For any restricted EDM M in $\left([\mathcal{R}(v, w)]_{C,C}\right)_{C,C}$, there exists a partition R of M and a matching K of $\mathcal{K}(v, w)$ such that for any elements M' of the partition R there exists a pair $(v_p, w_q) \in K$ such that:

- $\pi(v_p) = \pi(v)$, $\pi(w_q) = \pi(w)$ and $M' \in \left([\mathcal{G}(v_p, w_q)]_{C,C}\right)_{C,C}$;

- $\pi(v_p) = \pi(v)$, $\pi(w_q) \neq \pi(w)$ and :

$$M' \in [\mathcal{G}(v_p, w_q)]_{=, \mathcal{C}} \cup \left([\mathcal{G}(v_p, w_q)]_{\mathcal{C}, \mathcal{C}} \right)_{\mathcal{C}, \mathcal{C}} \cup \left([\mathcal{G}(v_p, w_q)]_{\mathcal{C}, \mathcal{C}} \right)_{=, \mathcal{C}}$$

- if $\pi(v_q) \neq \pi(v)$ and $\pi(w_p) = \pi(w)$ then:

$$M' \in [\mathcal{G}(v_p, w_q)]_{\mathcal{C}, =} \cup \left([\mathcal{G}(v_p, w_q)]_{\mathcal{C}, \mathcal{C}} \right)_{\mathcal{C}, \mathcal{C}} \cup \left([\mathcal{G}(v_p, w_q)]_{\mathcal{C}, \mathcal{C}} \right)_{\mathcal{C}, =}$$

- if $\pi(v_q) \neq \pi(v)$ and $\pi(w_p) \neq \pi(w)$ then $M' \in \mathcal{G}(v_p, w_q)$.

Proposition 29 For any restricted EDM M in $\left([\mathcal{R}(v, w)]_{=, =} \right)_{\mathcal{C}, =}$, there exists v_k and w_l sons of v and w such that $\pi(v_k) = \pi(v)$, $\pi(w_l) = \pi(w)$ and $M \in \left([\mathcal{G}(v_k, w_l)]_{=, =} \right)_{\mathcal{C}, =}$.

$\left([\mathcal{R}(v, w)]_{=, =} \right)_{=, \mathcal{C}}$ is determined by symmetry.

Proposition 30 For any restricted EDM M in $\left([\mathcal{R}(v, w)]_{=, =} \right)_{\mathcal{C}, \mathcal{C}}$, there exists a partition R of M and a matching K of $\mathcal{K}(v, w)$ such that there exists an element M' of R and a pair $(v_k, w_l) \in K$ such that $\pi(v_k) = \pi(v)$, $\pi(w_l) = \pi(w)$ and $M' \in \left([\mathcal{G}(v_k, w_l)]_{=, =} \right)_{\mathcal{C}, \mathcal{C}}$ and for any elements M'' of the partition R there exists a pair $(v_p, w_q) \in K$ satisfying one of these four properties:

- if $\pi(v_p) = \pi(v)$ and $\pi(w_q) = \pi(w)$ then $M'' \in \left([\mathcal{G}(v_p, w_q)]_{=, =} \right)_{\theta, \theta}$;
- otherwise, $M'' = \emptyset$.

Proposition 31 For any restricted EDM M in $\left([\mathcal{R}(v, w)]_{=, =} \right)_{\theta, \theta}$, there exists a partition R of M and a matching K of $\mathcal{K}(v, w)$ such that there exists an element M' of R and a pair $(v_k, w_l) \in K$ such that $\pi(v_k) = \pi(v)$, $\pi(w_l) = \pi(w)$ and $M' \in \left([\mathcal{G}(v_k, w_l)]_{=, =} \right)_{\theta, \theta}$ and for any elements M'' of the partition R there exists a pair $(v_p, w_q) \in K$ satisfying one of these four properties:

- if $\pi(v_p) = \pi(v)$ and $\pi(w_q) = \pi(w)$ then $M'' \in \left([\mathcal{G}(v_p, w_q)]_{=, =} \right)_{\theta, \theta}$;
- otherwise, $M'' = \emptyset$.

Proposition 32 For any restricted EDM M in $\left([\mathcal{R}(v, w)]_{=, =} \right)_{=, =}$, there exists a partition R of M and a matching K of $\mathcal{K}(v, w)$ such that there exists an element M' of R and a pair $(v_k, w_l) \in K$ such that $\pi(v_k) = \pi(v)$, $\pi(w_l) = \pi(w)$ and $M' \in \left([\mathcal{G}(v_k, w_l)]_{=, =} \right)_{=, =}$ and for any elements M'' of the partition R there exists a pair $(v_p, w_q) \in K$ satisfying one of these four properties:

- if $\pi(v_p) = \pi(v)$ and $\pi(w_q) = \pi(w)$ then:

$$M'' \in \left([\mathcal{G}(v_p, w_q)]_{=,=} \right)_{=,=} \cup \left([\mathcal{G}(v_p, w_q)]_{=,=} \right)_{C,C} \cup \left([\mathcal{G}(v_p, w_q)]_{=,=} \right)_{\theta,\theta} \\ \cup \left([\mathcal{G}(v_p, w_q)]_{C,C} \right)_{=,=} \cup \left([\mathcal{G}(v_p, w_q)]_{C,C} \right)_{C,C}$$

else;

- if $\pi(v_p) = \pi(v)$ and $\pi(w_q) \neq \pi(w)$ then $M'' \in [\mathcal{G}(v_p, w_q)]_{C,=} \cup [\mathcal{G}(v_p, w_q)]_{C,C}$ else;
- if $\pi(v_p) \neq \pi(v)$ and $\pi(w_q) = \pi(w)$ then $M'' \in [\mathcal{G}(v_p, w_q)]_{=,C} \cup [\mathcal{G}(v_p, w_q)]_{C,C}$ else;
- $M'' \in \mathcal{G}(v_p, w_q)$.

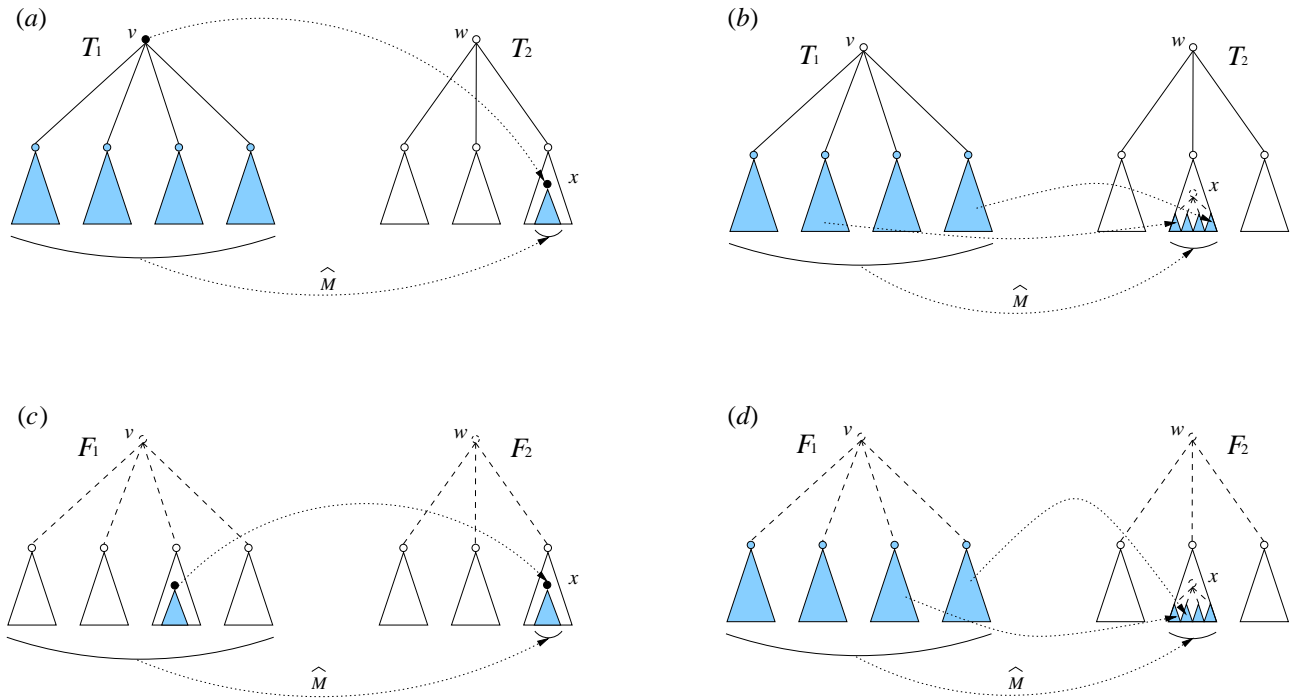


Figure 1: Definition of \widehat{M} : $S_1[v] \rightarrow \widehat{M}(S_1[v])$, (a) and (b) illustrates the image of one tree T_1 , (c) and (d) illustrates the image of one forest F_1 :

- (a) The image of T_1 is a sub-tree of T_2 ;
- (b) The image of T_1 is a forest included in T_2 , remark that the root of T_1 has no image by M ;
- (c) The image of F_1 is a tree included in F_2 , remark that vertices of F_1 which have an image by M belong to one tree of the forest F_2 ;
- (d) The image of F_1 is a forest of F_2 .

Color convention used throughout the paper. Vertices (or trees) represented in:

- black have an image by M ;
- grey may have or not an image by M ;
- white do not have any image.

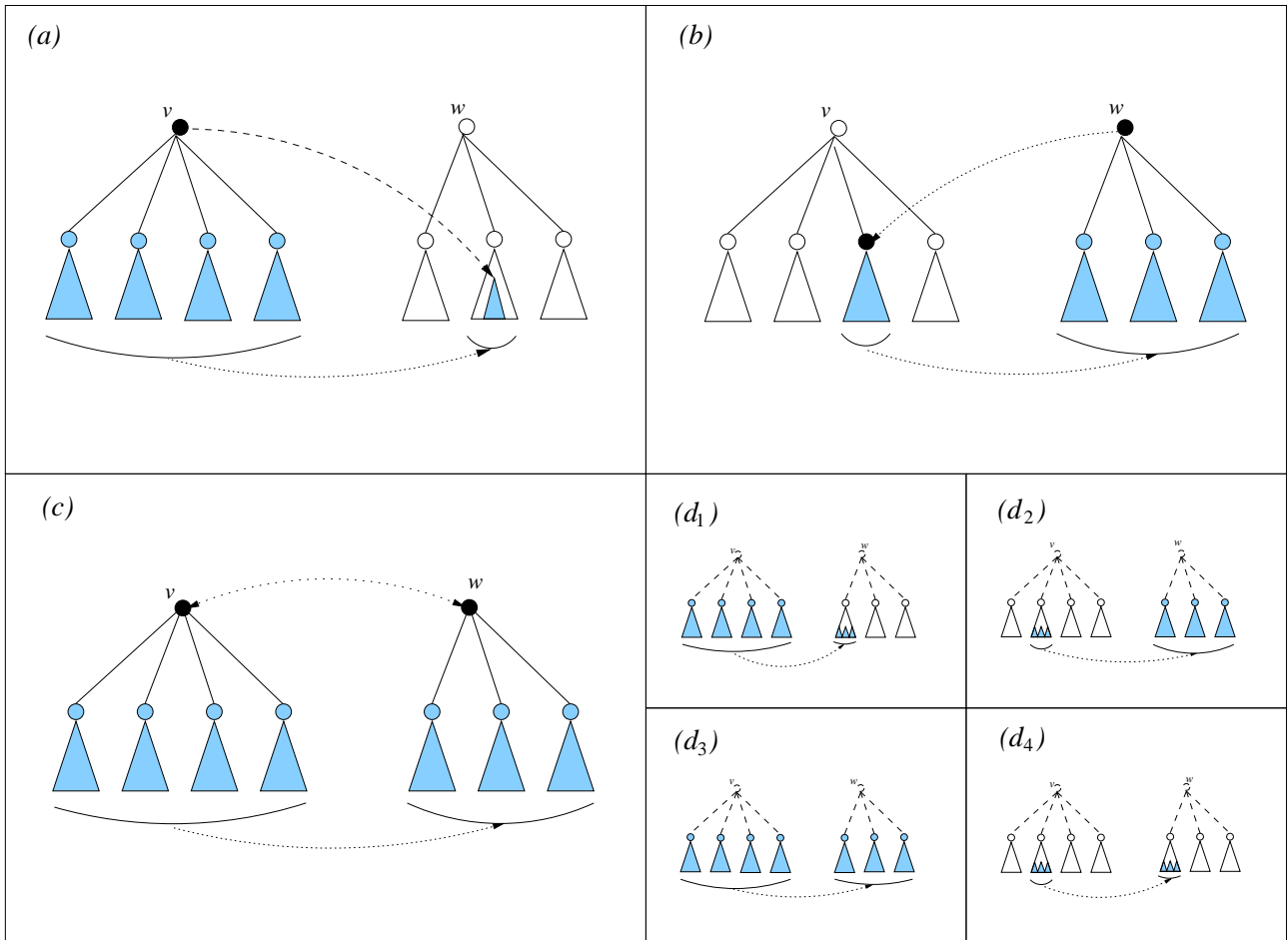


Figure 2: Partition of the set of valid EDMs from $T_1[v]$ to $T_2[w]$ using \widehat{M} , $\mathcal{T}(v, w)_{\theta, \theta}$ is not represented. General form of EDM in:

- (a) $\mathcal{T}(v, w)_{\mathcal{C}, =}$;
- (b) $\mathcal{T}(v, w)_{=, \mathcal{C}}$;
- (c) $\mathcal{T}(v, w)_{=, =}$;
- (d) $\mathcal{T}(v, w)_{\mathcal{C}, \mathcal{C}}$;
 - (d₁) $\mathcal{F}(v, w)_{\mathcal{C}, =}$;
 - (d₂) $\mathcal{F}(v, w)_{=, \mathcal{C}}$;
 - (d₃) $\mathcal{F}(v, w)_{=, =}$;
 - (d₄) $\mathcal{F}(v, w)_{\mathcal{C}, \mathcal{C}}$;

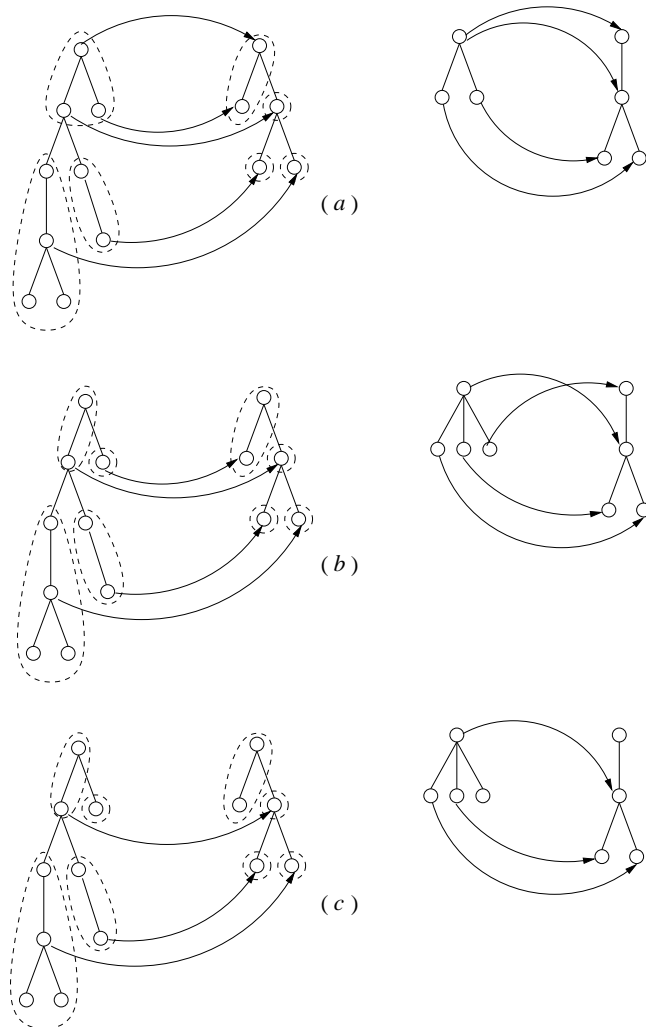


Figure 3: (a) and (b) illustrates valid EDMs for support tree comparison but they are not valid for quotient tree comparison:

- (a) since the one to one correspondance is not satisfied on quotient graph;
- (b) since the ancestor relationships is not satisfied on quotient graph.

(c) illustrates a valid EDM between quotiented trees, *i.e.* EDMs between both support and quotient graph are valid.

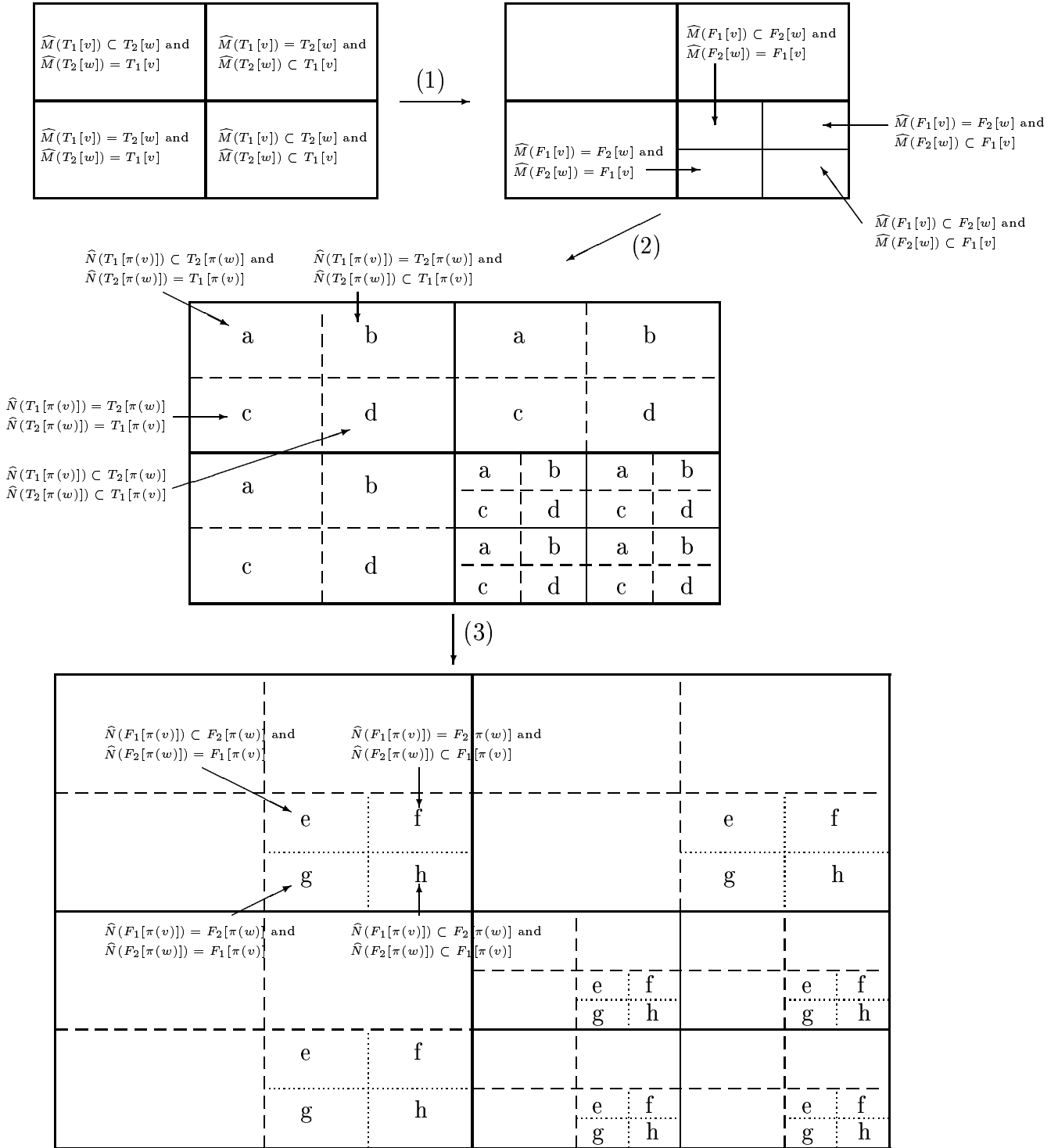


Figure 4: Different partition of $\mathcal{T}(v, w)$ of valid EDMs from $T_1[v]$ to $T_2[w]$. $\mathcal{T}(v, w)$ is firstly decomposed in five sub-sets ($\mathcal{T}(v, w)_{\theta, \theta}$ is not represented here) depending on the images of $T_1[v]$ to $T_2[w]$ by \widehat{M} . The sub-set corresponding to $\mathcal{T}(v, w)_{C, C}$ is then decomposed (1) in five new sub-sets ($\mathcal{F}(v, w)_{\theta, \theta}$ is neither represented here). Each sub-sets of this partition is then decomposed (2) in five new sub-sets depending on the image of $T_1[\pi(v)]$ and $T_2[\pi(w)]$ by \widehat{N} . So far, the sub-sets labelled d are decomposed in four new sub-sets depending on the image of $F_1[\pi(v)]$ and $F_2[\pi(w)]$ by \widehat{N} .

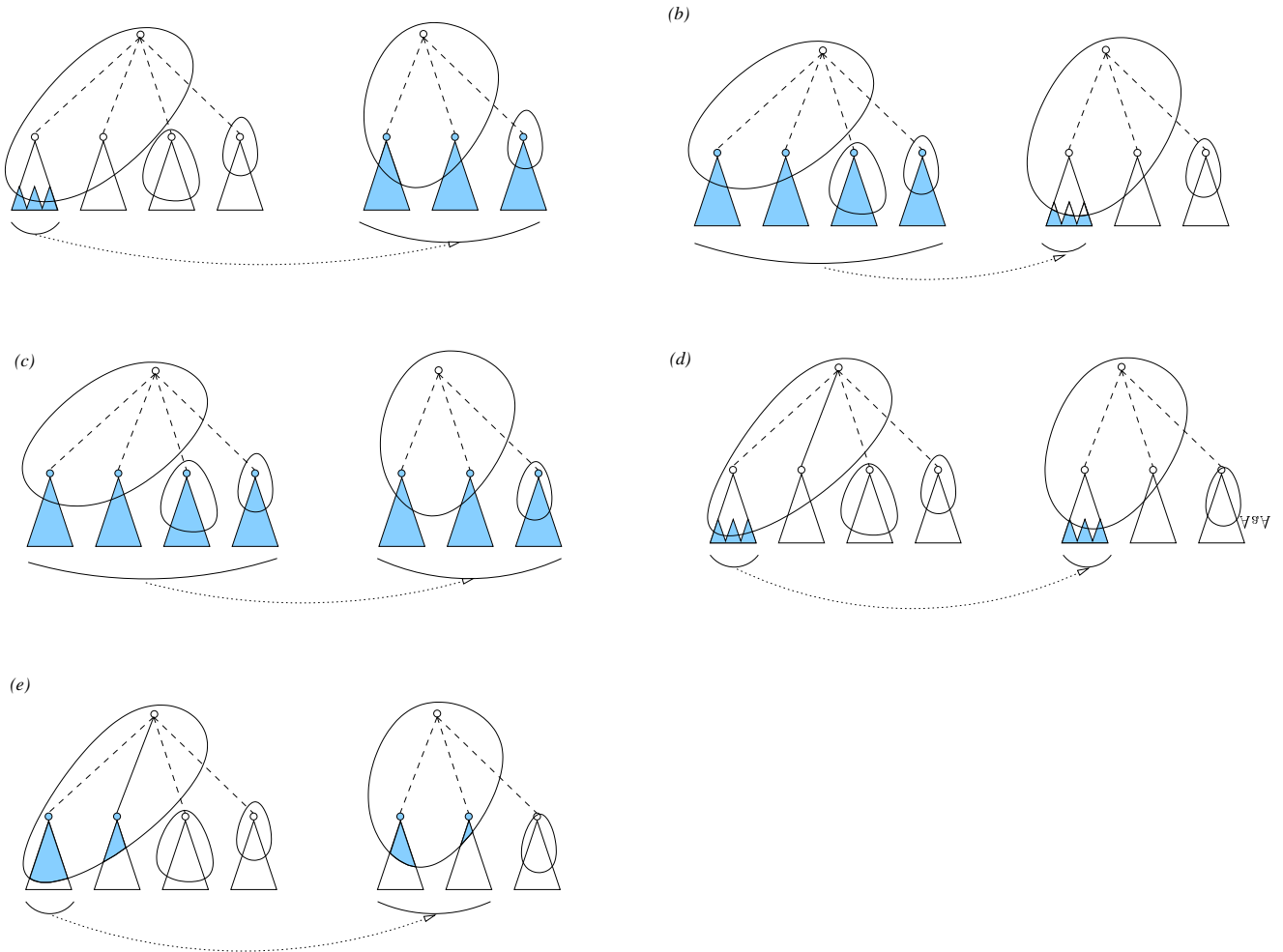


Figure 5: Partition of the set $[\mathcal{H}(v, w)]_{=,=}$ of valid EDMs from $F_1[v]$ to $F_2[w]$ using \widehat{M} such that $\widehat{M}(T_1[\pi(v)]) = T_2[\pi(w)]$ and $\widehat{M}(T_2[\pi(w)]) = T_1[\pi(v)]$. General form of EDM in:

- (a) $\left([\mathcal{H}(v, w)]_{=,=}\right)_{=, \mathcal{C}}$;
- (b) $\left([\mathcal{H}(v, w)]_{=,=}\right)_{\mathcal{C}, =}$;
- (c) $\left([\mathcal{H}(v, w)]_{=,=}\right)_{=, =}$;
- (d) $\left([\mathcal{H}(v, w)]_{=,=}\right)_{\mathcal{C}, \mathcal{C}}$;
- (e) $\left([\mathcal{H}(v, w)]_{=,=}\right)_{\theta, \theta}$.

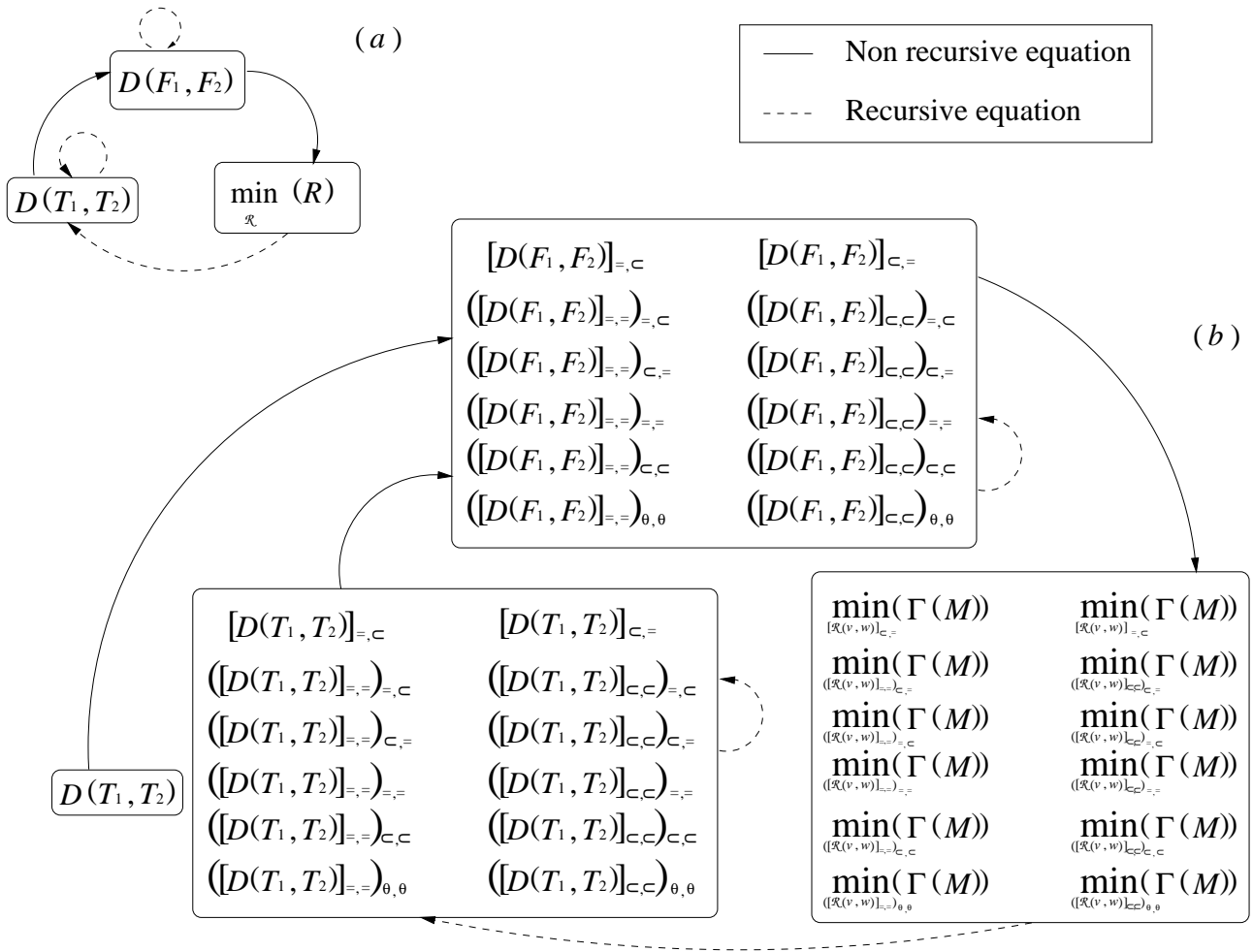


Figure 6: Dependency graphs of Zhang algorithm (a) and algorithm for comparing quotiented trees (b). In each graph an arrow from node A to node B means that quantities appearing in node A can be expressed in terms of quantities in node B . Plain arrows correspond to non-recursive equations between quantities of nodes A and B while dashed arrows corresponds to recursive equations.

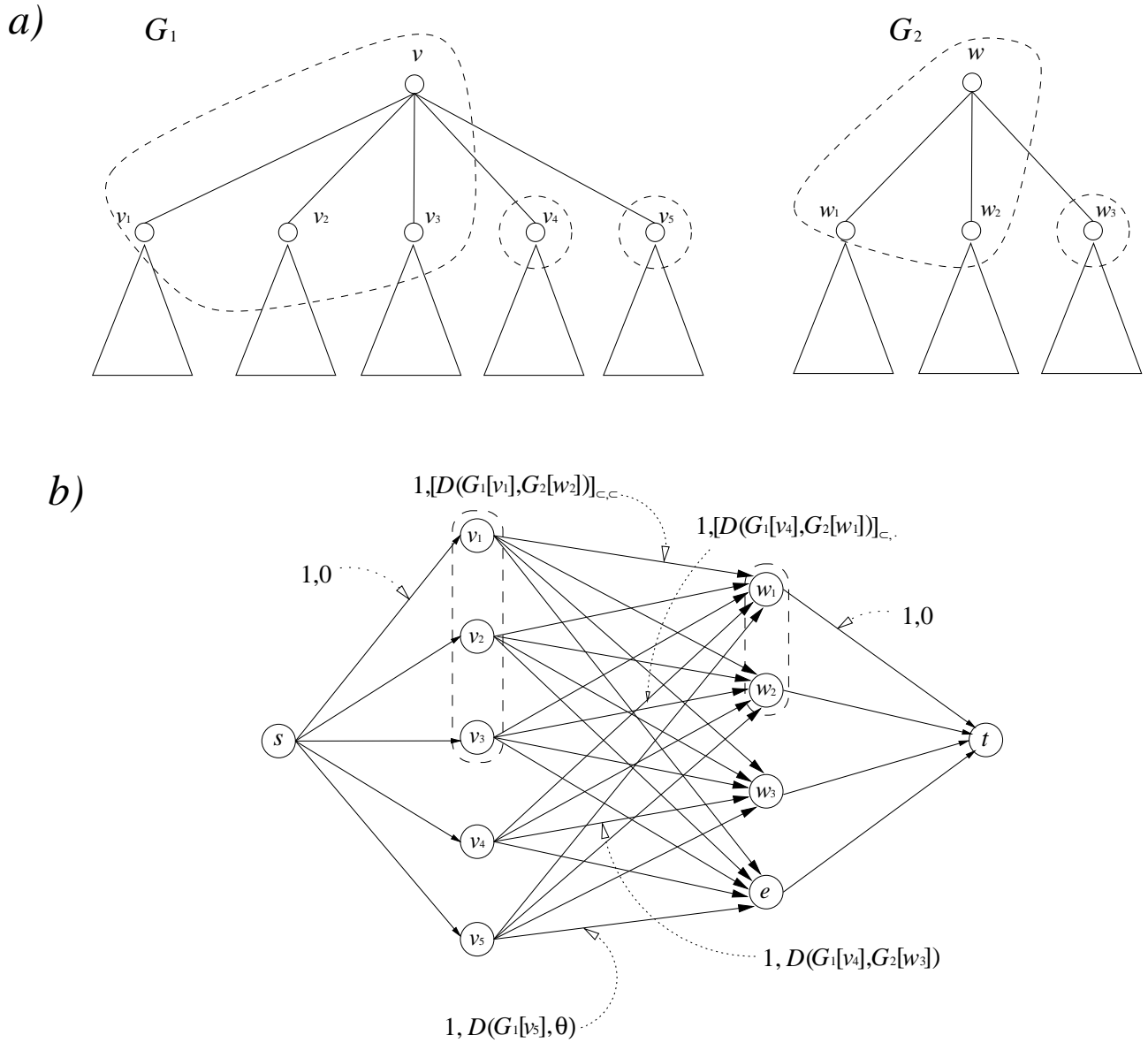


Figure 7: Reduction of optimal restricted EDM research to the minimum cost flow problem.

- (a) quotiented trees for which an optimal restricted EDM is searched;
- (b) Representation of the optimal restricted EDM problem as a minimum cost flow problem to find a restricted EDM in $[\mathcal{R}(v, w)]_{c,c}$.

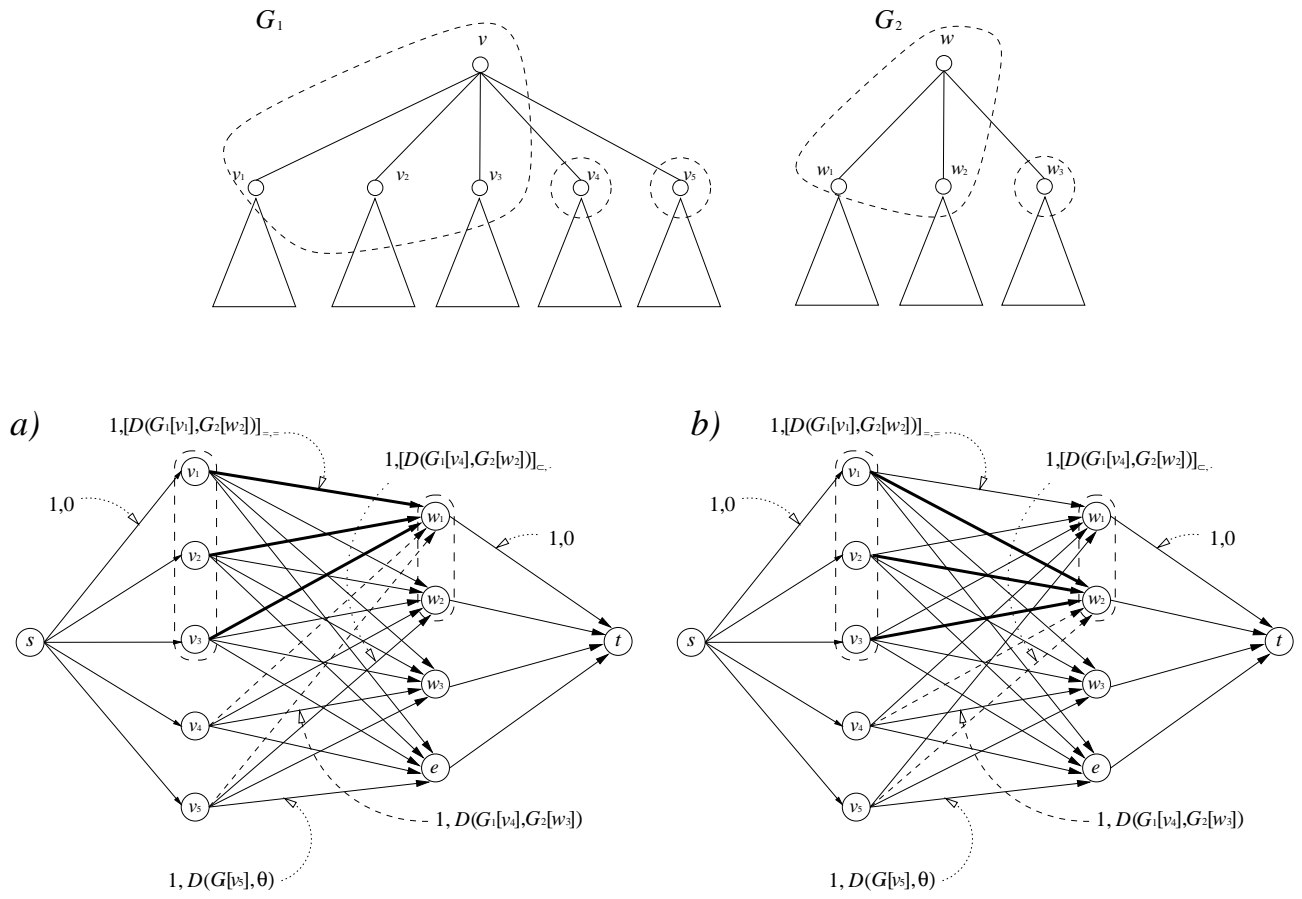


Figure 8: Reduction of optimal restricted EDM research to the minimum cost flow problem in $[\mathcal{R}(v, w)]_{=,=}$. In this case there must exist an edge between $\{w_1, w_2\}$ (in $\pi(w)$) and $\{v_1, v_2, v_3\}$ (in $\pi(v)$).

- (a) This edge reaches w_1 ;
- (b) This edge reaches w_2 .