Polynomial-Time Metrics for Attributed Trees [*]

Andrea Torsello [(1)] **Džena Hidović** [(2)] **Marcello Pelillo** [(1)]

(1) Dipartimento di Informatica

Università Ca' Foscari di Venezia

Via Torino 155, 30172 Venezia Mestre, Italy

(2) School of Computer Science

University of Birmingham

Edgbaston, Birmingham, B15 2TT, United Kingdom

Abstract

We address the problem of comparing attributed trees and propose four novel distance measures centered around the notion of a maximal similarity common subtree. The proposed measures are general and defined on trees endowed with either symbolic or continuous-valued attributes, and can be equally applied to ordered and unordered, rooted and unrooted trees. We prove that our measures satisfy the metric constraints and provide a polynomial-time algorithm to compute them. This is a remarkable and attractive property, since the computation of traditional edit-distance-based metrics is NP-complete, except for ordered structures. We experimentally validate the usefulness of our metrics on shape matching tasks, and compare them with edit-distance measures.

---

1   Introduction

Graph-based representations have long been used with considerable success in computer vision and pattern recognition in the abstraction and recognition of objects and scene structure. Concrete examples include the use of shock graphs to represent shape-skeletons [14, 26], the use of trees to represent articulated objects [12, 40] and the use of aspect graphs for 3D object representation [8]. The attractive feature of structural representations is that they concisely capture the relational arrangement of object primitives, in a manner which can be invariant to changes in object viewpoint. Using this framework we can transform a recognition problem into a relational matching problem. The problem of how to measure the similarity or distance of pictorial information using graph abstractions has been a widely researched topic of over twenty years.

Early work on the topic included Barrow and Burstall's idea [2] of characterizing the similarity of two graphs using the cardinality of their maximum common subgraphs, and the extension of the concept of string edit-distance to graph-matching by Fu and his co-workers [9]. Haralick and Shapiro [23] have described a relational distance measure between structural descriptions. There have also been attempts to use an information theoretic approach. Here, Wong and You [35] have computed the entropy for random graphs, while Boyer and Kak [4] have used mutual information. More recently, Christmas, Kittler and Petrou [7], and Wilson and Hancock [34] have developed probabilistic measures of graph-similarity. Unfortunately, with the notable exception of edit-distance, the resulting measures are not metrics, i.e. they are either non-symmetric, negative, or violate the triangular inequality. The lack of metric properties makes the undistorted embedding in a vector space impossible and does not provide a natural ordering within a database of graphs.

The idea behind edit-distance [9], which has become the standard metric approach to graph comparison, is that it is possible to identify a set of basic edit operations on nodes and edges of a structure, and to associate with these operations a cost. The edit-distance is found by searching for sequences of edit operations that will make the two graphs isomorphic with one-another, and the distance between the two graphs is then defined

1

to be the minimum over all the costs of these sequences. By making the evaluation of structural modification explicit, edit-distance provides a very effective way of measuring the similarity of relational structures. Moreover, the method has considerable potential for error tolerant object recognition and indexing problems. Unfortunately, the task of calculating edit-distance is a computationally hard problem [38], hence, goal-directed approximations are necessary to calculate it. The result is that the approximation almost invariably breaks the theoretical metric properties of the measure.

Recently, a new and more principled approach to the definition of distance measure has emerged. In [6], Bunke and Shearer introduce a distance measure on unattributed graphs based on the maximum common subgraph and prove that it is a metric. Wallis et al. [32] introduce a variant of this distance based on the size of the minimum common supergraph. Finally, Fernandez and Valiente [10] define a metric based on the difference in size between maximum common subgraph and minimum common supergraph. More recently, in [11] Hidović and Pelillo extend these metrics to the case of attributed graphs. Unfortunately all these metrics require the calculation of the maximum common subgraph, which is computationally equivalent to the calculation of edit-distance [5].

In many computer vision and pattern recognition applications, such as shape recognition [22, 40, 25, 29], pattern recognition [16], and image processing [21], the graphs at hand have a peculiar structure: they are connected and acyclic, i.e., they are *trees*, either rooted or unrooted, ordered or unordered, and frequently they are endowed with symbolic and/or continuous-valued attributes. Most metrics on trees found in the literature are defined in terms of edit-distance [30, 33]. Zhang and Shasha [37] have investigated a special case of edit-distance which involves trees with an order relation among sibling nodes in a rooted tree. This special case constrains the solution to maintain the order of the children of a node. They showed that this constrained tree-matching problem is solvable in polynomial time and gave an algorithm to solve it. Recently, Sebastian, Klein and Kimia [22] use a similar algorithm to compare shock trees. Other constrained solutions proved to be solvable in polynomial time have been proposed [36], yet in the general case the problem has been proven to be NP-complete both for rooted [38] and

unrooted trees [39]. Recently, Valiente [31] introduced a bottom-up distance measure between trees that is an extension to trees of the graph metric introduced by Bunke and Shearer [6], proving that the measure can be calculated in polynomial time on trees, but falling short of proving that the measure is a metric. While this measure can be calculated efficiently both on ordered and unordered trees, it is limited to rooted and unattributed trees. Another bottom-up method for computing the distance between trees has also been proposed by Tanaka [27].

Motivated by the work described in [11], in this paper we propose four distance measures, two normalized and two non-normalized, for trees equipped with either symbolic or continuous-valued attributes. We prove that all our measures fulfill the properties of a metric, and provide a polynomial-time algorithm to compute them. This is an important property which make them particularly attractive. In fact, as mentioned above, traditional metrics on trees, which are based on edit-distance, are computationally hard unless we confine ourselves to special cases. At an abstract level, our approach involves the computation of a maximum similarity common subtree. This allows us to define equivalent variations of the metrics on ordered and unordered, rooted and unrooted, and attributed and unattributed trees. They can also be viewed as variants of the metrics developed by Bunke and Shearer [6], Wallis et al. [32], and Fernandez and Valiente [10] on arbitrary graphs.

Since edit-distance on ordered trees can be computed in polynomial time, in the paper we focus on the unordered case where our approach provides a clear computational advantage. To show the validity of the proposed measures, we present experiments on various shape matching tasks and compare our results with those obtained using edit-distance metrics.

The outline of the paper is as follows. Section 2 introduces formalisms and concepts required throughout the paper. In Section 3, we present a polynomial-time algorithm to calculate the maximum similarity common subtree needed to compute all our metrics, and in Section 4 we define our measures and prove that they satisfy the metric properties. Finally section 6 provides some experimental validation of the usefulness of the metrics

and in Section 7 we draw our conclusions.

2 Preliminaries

Let $G = (V, E)$ be a graph, where V is the set of nodes (or vertices) and E is the set of undirected edges. Two nodes $u, v \in V$ are said to be *adjacent* (denoted $u \sim v$) if they are connected by an edge. A *path* is any sequence of distinct nodes $u_0 u_1 \ldots u_n$ such that for all $i = 1 \ldots n$, $u_{i-1} \sim u_i$; in this case, the *length* of the path is n. If $u_n \sim u_0$ the path is called a *cycle*. A graph is said to be *connected* if any two nodes are joined by a path. Given a subset of nodes $C \subseteq V$, the *induced subgraph $G[C]$* is the graph having C as its node set, and two nodes are adjacent in $G[C]$ if and only if they are adjacent in G. With the notation $|G|$ we shall refer to the cardinality of the node-set of graph G.

A connected graph with no cycles is called an unrooted tree. A rooted (or hierarchical) tree is a tree with a special node that can be identified as the root. In what follows, when using the word "tree" without qualification, we shall refer to both the rooted and unrooted cases. Trees have a number of interesting properties. One which turns out to be very useful is that in a tree any two nodes are connected by a unique path. Given two nodes $u, v \in V$ in a rooted tree, u is said to be an *ancestor* of v (and similarly v is said to be a *descendent* of u) if the path from the root node to u is a subpath of the path from the root to v. Furthermore, if $u \sim v$, u is said to be the *parent* of v and v is said to be a *child* of u. Both ancestor and descendent relations are order relations in V.

Let $T_1 = (V_1, E_1)$ and $T_2 = (V_2, E_2)$ be two trees. Any bijection $\phi : H_1 \to H_2$, with $H_1 \subseteq V_1$ and $H_2 \subseteq V_2$, is called a *subtree isomorphism* if it preserves both the adjacency relationships between the nodes and the connectedness of the matched subgraphs. Formally, this means that, given $u, v \in H_1$, we have $u \sim v$ if and only if $\phi(u) \sim \phi(v)$ and, in addition, the induced subgraphs $T_1[H_1]$ and $T_2[H_2]$ are connected. Two trees or rooted trees T_1 and T_2 are *isomorphic*, and we write $T_1 \cong T_2$, if there exists an isomorphism between them that maps every node in T_1 to every node in T_2. It is easy to verify that isomorphism is an equivalence relation. We shall use the notations $\mathrm{Dom}(\phi)$ and $\mathrm{Im}(\phi)$

to denote the domain and the image of ϕ, respectively.

A word of caution about terminology is in order here. Despite name similarity, we are not addressing the standard subtree isomorphism problem, which consists of determining whether a given tree is isomorphic to a subtree of a larger one. In fact, we are dealing with a generalization thereof, the maximum common subtree problem, which consists of determining the largest isomorphic subtrees of two given trees. We shall continue to use our own terminology, however, as it emphasizes the role of the isomorphism ϕ.

Formally, an *attributed tree* is a triple $T = (V, E, \alpha)$, where (V, E) is the "underlying" tree and α is a function which assigns an attribute vector $\alpha(u)$ to each node $u \in V$. It is clear that in matching two attributed trees, our objective is to find an isomorphism which pairs nodes having "similar" attributes. To this end, let σ be any similarity measure on the attribute space, i.e., any (symmetric) function which assigns a positive number to any pair of attribute vectors. If $\phi : H_1 \to H_2$ is a subgraph isomorphism between two attributed trees $T_1 = (V_1, E_1, \alpha_1)$ and $T_2 = (V_2, E_2, \alpha_2)$, the overall similarity between the induced subtrees $T_1[H_1]$ and $T_2[H_2]$ can be defined as follows:

$$W_\sigma(\phi) = \sum_{u \in H_1} \sigma(u, \phi(u)) \ . \tag{1}$$

where, for simplicity, we define $\sigma(u, \phi(u)) \equiv \sigma(\alpha_1(u), \alpha_2(\phi(u)))$. The isomorphism ϕ is called a *maximum similarity subtree isomorphism* if $W_\sigma(\phi)$ is largest among all subtree isomorphisms between T_1 and T_2. For the rest of the paper we will omit the subscript σ when the node-similarity used is clear from the context. Two isomorphic attributed trees $T_1 = (V_1, E_1, \alpha_1)$ and $T_2 = (V_2, E_2, \alpha_2)$, with isomorphism ϕ, are said to be *attribute-isomorphic* if for all $u \in V_1$ we have $\alpha_1(u) = \alpha_2(\phi(u))$. In this case we shall write $T_1 \cong_a T_2$. Attribute-isomorphism is clearly an equivalence relation.

We note that, although our treatment has started from the assumption that explicit attributes are available on each node, the framework is more general and can well be applied to situations where this is not the case but, rather, only pairwise measures are known.

Note that the problem of determining a maximum similarity subtree isomorphism is a direct extension of the standard problem of finding a maximum (cardinality) common

subtree, in fact the two problems are equivalent when the similarity σ is degenerate, i.e., $\sigma(u, v) = 1$.

Now, given a set S, a function $d : S \times S \to \mathbb{R}$ is a *metric* on S if the following properties hold for any $x, y, z \in S$.

1. $d(x, x) \geq 0$ (non-negativity)

2. $d(x, y) = 0 \Leftrightarrow x = y$ (identity and uniqueness)

3. $d(x, y) = d(y, x)$ (symmetry)

4. $d(x, y) + d(y, z) \geq d(x, z)$ (triangular inequality).

Furthermore, if the function satisfies $d(x, y) \leq 1$ it is said to be a *normalized metric*.

If $d : S \times S \to \mathbb{R}_+$ is a normalized metric, then the similarity function derived from δ, defined as

$$\sigma(x, y) = 1 - d(x, y) \tag{2}$$

fulfills the identity, uniqueness and similarity properties. Furthermore, it fulfills the following variant of the triangular inequality:

$$\sigma(x, y) + \sigma(y, z) - \sigma(x, z) \leq 1 . \tag{3}$$

This property can be obtained from the definition and the triangular inequality of the metric d:

$$\sigma(x, y) + \sigma(y, z) - \sigma(x, z) \leq 1 \Leftrightarrow (1 - d(x, y)) + (1 - d(y, z)) - (1 - d(x, z)) \leq 1 \Leftrightarrow$$
$$- d(x, y) - d(y, z) + d(x, z) \leq 0 \Leftrightarrow d(x, y) + d(y, z) \geq d(x, z) .$$

In the rest of the paper, we shall assume that all similarity functions are indeed derived from normalized metrics.

It is straightforward to show that, with this assumption, we have

$$T_1 \cong_a T_2 \Leftrightarrow |T_1| = |T_2| = W(\phi) \tag{4}$$

where ϕ is a maximum similarity isomorphism between T_1 and T_2.

3   Extracting a Maximum Similarity Common Subtree

In this section we give a polynomial-time algorithm for finding a maximum similarity subtree. The algorithm is based on the subtree identification algorithm presented by Matula [15], extending it in two ways. First, it generalizes it to deal with attributed trees and, second, it extends it to solve the more general problem of extracting the maximum (similarity) subtree and not merely to verify whether one tree is a subtree of the other. We give an algorithm to find the maximum similarity common subtree problem for rooted trees, and then we show how the same algorithm can be used for the unrooted tree case.

Let $T_1 = (V_1, E_1)$ and $T_2 = (V_2, E_2)$ be two rooted trees, and let $u \in V_1$ and $w \in V_2$. We say that a subtree isomorphism between T_1 and T_2 is *anchored* at nodes u and w, if the subtrees of T_1 and T_2 induced by the isomorphism are rooted at u and w, respectively. In this case, we shall write $\phi^{(u,w)}$ to refer to any isomorphism anchored at u and w. Clearly, if ϕ is a maximum similarity subtree isomorphism, we have

$$W(\phi) = \max_{(u,w) \in V_1 \times V_2} \max_{\phi^{(u,w)}} W(\phi^{(u,w)}).$$

In reality, since if neither u nor w is a root of T_1 or T_2, we can add the parents of u and w to the mapping without reducing the similarity, we have:

$$W(\phi) = \max_{(u,w) \in (\{r_1\} \times V_2) \cup (V_1 \times \{r_2\})} \max_{\phi^{(u,w)}} W(\phi^{(u,w)}).$$

where r_1 and r_2 are the roots of T_1 and T_2, respectively.

To determine the maximum similarity subtree isomorphism anchored at nodes u and w we adopt a divide-and-conquer approach. Let $u_1, \cdots, u_n$ be the children of node u in T_1, and $w_1, \cdots, w_m$ the children of node w in T_2. Without loss of generality, we can assume $n \leq m$. Moreover, let us assume that we know, for each $i = 1, \cdots, n$ and $j = 1, \cdots, m$, a maximum similarity subtree isomorphism $\widehat{\phi}^{(u_i, w_j)}$ anchored at u_i and w_j. Let W_{ij} be the similarity of $\widehat{\phi}^{(u_i, w_j)}$, then the computation of a maximum similarity

7

subtree isomorphism anchored at u and w can be reduced to an assignment problem on the children of u and w, i.e.,

$$W(\phi^{(u,w)}) = \sigma(u,w) + \max_{\pi \in \Sigma_n^m} \sum_{i=1}^n W_{i\pi(i)}, \tag{5}$$

where Σ_n^m is the space of all possible assignments between a set of cardinality n and one of cardinality m. As a consequence, if π is the optimal assignment, the function $\phi^{(u,w)}$ defined as:

$$\phi^{(u,w)}(x) = \begin{cases} w & \text{if } x = u \\ \widehat{\phi}^{(u_i,w_{\pi(i)})}(x) & \text{if } x \in \text{Dom}(\widehat{\phi}^{(u_i,w_{\pi(i)})}) \end{cases} \tag{6}$$

turns out to be a maximum similarity subtree isomorphism anchored at u and w.

Figure 1 shows the resulting algorithm for determining a maximum similarity subtree isomorphism of two rooted attributed trees. Since in the rest of the paper we only need the maximum similarity induced by an isomorphism, and not the isomorphism itself, for simplicity the main procedure `MaxSimilarity` accepts as input a pair of attributed rooted trees and returns only the similarity value. It makes use of a recursive procedure `AnchoredSimilarity` that accepts as input two vertices, one from T_1 and the other from T_2 and returns the similarity of the maximum isomorphism anchored at the input vertices, according to (5). To this end, it needs a procedure for solving an assignment (or, equivalently, a bipartite matching) problem, of which the algorithms literature abound (see., e.g., [17]). The calculation of the maximum similarity common subtree of two trees with N and M nodes respectively, is reduced to at most NM weighted assignments problems of dimension at most b, where b is the maximum branching factor of the two trees. The computational complexity of our algorithm heavily depends on the actual implementation of the assignment procedure. A popular way of solving it, and the one we actually employed, is the so-called Hungarian algorithm, which has complexity $O(n^2m)$, n and m being the number of children of u and v as used in (5), with $n \leq m$. It is simple to show that, using the Hungarian algorithm, our algorithm has overall complexity of $O(bNM)$. Of course, the algorithm can be sped up by using more sophisticated assignment procedures [1].

8

```
MaxSimilarity(T₁,T₂)

    maxsim=0

    for each node u in T₁

        sim=AnchoredSimilarity(u,root(T₂))

        if sim > maxsim

            maxsim=sim

    for each node w in T₂

        sim=AnchoredSimilarity(root(T₁),w)

        if sim > maxsim

            maxsim=sim

return maxsim
```

```
AnchoredSimilarity(u,w)

    Cᵤ=children(u)

    C_w=children(w)

    for each uᵢ in Cᵤ

        for each wⱼ in C_w

            wᵢⱼ=AnchoredSimilarity(uᵢ,wⱼ)

return σ(u,w) + Assign({wᵢⱼ})
```

Figure 1: A polynomial-time algorithm for computing the maximum similarity between two trees.

Finally, if we have two unrooted trees $T_1 = (V_1, E_1)$ and $T_2 = (V_2, E_2)$, we can still pick two nodes $r_1 \in V_2$ and $r_2 \in V_2$, and consider the trees $T_1^{r_1} = (V_1, E_1)$ and $T_2^{r_2} = (V_2, E_2)$ rooted at r_1 and r_2, respectively. Note that if ϕ is an isomorphism between $T_1^{r_1}$ and $T_2^{r_2}$ with similarity W, then it is an isomorphism between T_1 and T_2 with the same similarity. This yields a straightforward $O(bN^3M)$ algorithm for unrooted trees, which consists of iteratively calling `MaxSimilarity`(T_1^u, T_2^w) for all $u \in V_1$ and $w \in V_2$, and taking the maximum. However, we do not actually need to try all possible pairs of roots since by simply fixing the root in one tree and let the other vary among all possible vertices in the other tree, the algorithm is still guaranteed to achieve the maximum similarity. This follows easily from the observation made above that it always exists a maximum similarity (rooted) subtree isomorphism mapping at least one of the roots of the two trees. Indeed, without loss of generality, let us root T_1 on an arbitrary node u. Then, either u is mapped (say, to node $w \in V_2$) by a maximum similarity isomorphism, or it remains unmapped. In the former case, we clearly obtain the optimum by applying the rooted algorithm to T_1^u and T_2^w. In the latter case, a maximum similarity isomorphism ϕ will induce a subtree in T_1^u rooted at, say, $v \in V_1$. Clearly, the algorithm called on T_1^u and $T_1^{\phi(v)}$ will return the optimum. This yields an $O(bN^2M)$ algorithm for unrooted trees.

4 Distance Metrics

In this section, we present the main contribution of this paper: We define our measures for comparing attributed trees and prove that they fulfill the metric properties. First, we prove a lemma that turns out to be instrumental to prove our results. Second, we introduce two non-normalized metrics and, finally, we present the normalized versions of the previous measures.

Lemma 1 *Let T_1, T_2 and T_3 be three trees, and ϕ_{12} ϕ_{23}, and ϕ_{13} be maximum similarity subtrees isomorphisms between T_1 and T_2, T_2 and T_3, and T_1 and T_3, respectively. Then, we have: $|T_2| \geq W(\phi_{12}) + W(\phi_{23}) - W(\phi_{13})$.*

Proof. Let $V_2^1 = \text{Im}(\phi_{12}) \subseteq V_2$, $V_2^3 = \text{Dom}(\phi_{23}) \subseteq V_2$ be the sets of nodes in V_2 mapped by the isomorphisms ϕ_{12} and ϕ_{23}, respectively. Furthermore, let $\hat{V}_2 = V_2^1 \cap V_2^3$, be the set of vertices in V_2 that are mapped by both isomorphisms. It is clear that the subtrees $\hat{T}_1 = T_1[\phi_{12}^{-1}(\hat{V}_2)]$ and $\hat{T}_3 = T_3[\phi_{23}(\hat{V}_2)]$ are isomorphic to each-other, with isomorphism $\hat{\phi}_{13} = \phi_{12} \circ \phi_{23}$, where $\circ$ denotes the standard function composition operator, restricted to the nodes of $\hat{T}_1$. The similarity of this isomorphism is

$$W(\hat{\phi}_{13}) = \sum_{v \in \hat{V}_2} \sigma(\phi_{12}^{-1}(v), \phi_{23}(v)) \,.$$

Since ϕ_{13} is a maximum similarity subtree isomorphism between T_1 and T_3, we have $W(\phi_{13}) \geq W(\hat{\phi}_{13})$. Hence

$$W(\phi_{12}) + W(\phi_{23}) - W(\phi_{13}) \leq W(\phi_{12}) + W(\phi_{23}) - W(\hat{\phi}_{13}) =$$

$$\sum_{v \in V_2^1} \sigma(\phi_{12}^{-1}(v), v) + \sum_{v \in V_2^3} \sigma(v, \phi_{23}(v)) - \sum_{v \in \hat{V}_2} \sigma(\phi_{12}^{-1}(v), \phi_{23}(v)) =$$

$$\sum_{v \in V_2^1 \setminus V_2^3} \sigma(\phi_{12}^{-1}(v), v) + \sum_{v \in V_2^3 \setminus V_2^1} \sigma(v, \phi_{23}(v)) +$$

$$\sum_{v \in \hat{V}_2} \left[\sigma(\phi_{12}^{-1}(v), v) + \sigma(v, \phi_{23}(v)) - \sigma(\phi_{12}^{-1}(v), \phi_{23}(v)) \right] \leq$$

$$|V_2^1 \setminus V_2^3| + |V_2^3 \setminus V_2^1| + |V_2^1 \cap V_2^3| = |V_2^1 \cup V_2^3| \leq |T_2| \,,$$

where the inequality follows from the triangular inequality for metric-derived similarities (3). $\square$

4.1   Non-Normalized Metrics

Let $\mathcal{T}$ be the quotient set of trees modulo attribute-isomorphism, that is the set of trees on which two trees are considered the same if they are attribute-isomorphic.[1] For any

---

[1]The quotient set formalizes the intuitive idea that two attributed trees are indistinguishable when they are attribute-isomorphic. Furthermore, it is needed in order to fulfill the uniqueness property of a metric.

$T_1, T_2 \in \mathcal{T}$ we define the following distance functions

$$d_1(T_1, T_2) = \max(|T_1|, |T_2|) - W(\phi_{12}) \tag{7}$$

$$d_2(T_1, T_2) = |T_1| + |T_2| - 2W(\phi_{12}) \tag{8}$$

where ϕ_{12} is the maximum similarity common subtree isomorphism between T_1 and T_2. As noted before, the calculation of ϕ_{12} and, consequently, the optimal value of $W(\phi_{12})$, is going to be different for rooted and unrooted trees. Nevertheless, once the optimal similarity is at hand, the definition of the distance and the analysis of its properties are independent on whether the trees are rooted or not.

Theorem 1 *d_1 and d_2 are metrics in $\mathcal{T}$.*

Proof.

1. $d_1(T_1, T_2) \geq 0$ and $d_2(T_1, T_2) \geq 0$.

    We have $W(\phi_{12}) \leq \min(|T_1|, |T_2|) \leq \frac{|T_1| + |T_2|}{2} \leq \max(|T_1|, |T_2|)$. Hence, $d_1(T_1, T_2) = \max(|T_1|, |T_2|) - W(\phi_{12}) \geq 0$ and $d_2(T_1, T_2) = |T_1| + |T_2| - 2W(\phi_{12}) \geq 0$.

2. $d_1(T_1, T_2) = 0 \iff T_1 \cong_a T_2$ and $d_2(T_1, T_2) = 0 \iff T_1 \cong_a T_2$.

    Let us consider the direction of implication $\Leftarrow$ (identity). From (4), we have $T_1 \cong_a T_2 \Rightarrow |T_1| = |T_2| = W(\phi_{12})$. Hence $d_1(T_1, T_2) = \max(|T_1|, |T_2|) - W(\phi_{12}) = 0$ and $d_2(T_1, T_2) = |T_1| + |T_2| - 2W(\phi_{12}) = 0$.

    For the reverse implication (uniqueness), we have $d_1(T_1, T_2) = 0 \Rightarrow W(\phi_{12}) = \max(|T_1|, |T_2|)$. Since $W(\phi_{12}) \leq \min(|T_1|, |T_2|) \leq \max(|T_1|, |T_2|)$, we have $W(\phi_{12}) = \min(|T_1|, |T_2|) = \max(|T_1|, |T_2|)$. Hence, (4) yields $T_1 \cong_a T_2$.

    Similarly, $d_2(T_1, T_2) = 0 \Rightarrow 2W(\phi_{12}) = |T_1| + |T_2|$, and since $2W(\phi_{12}) \leq 2\min(|T_1|, |T_2|) \leq |T_1| + |T_2|$, we have $W(\phi_{12}) = |T_1| = |T_2|$, or $T_1 \cong_a T_2$.

3. $d_1(T_1, T_2) = d_1(T_2, T_1)$ and $d_2(T_1, T_2) = d_2(T_2, T_1)$

    This follows directly from the symmetry of the maximum similarity graph and of the function max.

4. $d_1(T_1, T_2) + d_1(T_2, T_3) \geq d_1(T_1, T_3)$ and $d_2(T_1, T_2) + d_2(T_2, T_3) \geq d_2(T_1, T_3)$.

To prove the triangular inequality of d_1, we need to separately analyze each of the six possible cases

1. $|T_1| \geq |T_2| \geq |T_3|$ 2. $|T_1| \geq |T_3| \geq |T_2|$ 3. $|T_2| \geq |T_1| \geq |T_3|$

4. $|T_2| \geq |T_3| \geq |T_1|$ 5. $|T_3| \geq |T_1| \geq |T_2|$ 6. $|T_3| \geq |T_2| \geq |T_1|$.

However, the roles of T_1 and T_3 in our proofs are symmetric, hence we can use this symmetry to reduce the analysis to three cases: $|T_2| \geq |T_1| \geq |T_3|$, $|T_1| \geq |T_2| \geq |T_3|$, and $|T_1| \geq |T_3| \geq |T_2|$.

(a) $|T_2| \geq |T_1| \geq |T_3|$

$$d_1(T_1, T_2) + d_1(T_2, T_3) - d_1(T_1, T_3) = |T_2| - W(\phi_{12}) + |T_2| - W(\phi_{23}) - |T_1| + W(\phi_{13}) =$$
$$(|T_2| - |T_1|) + [|T_2| - (W(\phi_{12}) + W(\phi_{23}) - W(\phi_{13}))] \geq 0.$$

(b) $|T_1| \geq |T_2| \geq |T_3|$

$$d_1(T_1, T_2) + d_1(T_2, T_3) - d_1(T_1, T_3) = |T_1| - W(\phi_{12}) + |T_2| - W(\phi_{23}) - |T_1| + W(\phi_{13}) =$$
$$|T_2| - (W(\phi_{12}) + W(\phi_{23}) - W(\phi_{13})) \geq 0.$$

(c) $|T_1| \geq |T_3| \geq |T_2|$

$$d_1(T_1, T_2) + d_1(T_2, T_3) - d_1(T_1, T_3) = |T_1| - W(\phi_{12}) + |T_3| - W(\phi_{23}) - |T_1| + W(\phi_{13}) =$$
$$|T_3| - (W(\phi_{12}) + W(\phi_{23}) - W(\phi_{13})) \geq |T_2| - (W(\phi_{12}) + W(\phi_{23}) - W(\phi_{13})) \geq 0.$$

On the other hand, for d_2, we have

$$d_2(T_1, T_2) + d_2(T_2, T_3) - d_2(T_1, T_3) = |T_1| + |T_2| - 2W(\phi_{12}) + |T_2| + |T_3| - 2W(\phi_{23})$$
$$- |T_1| - |T_3| + 2W(\phi_{13}) = 2[|T_2| - (W(\phi_{12}) + W(\phi_{23}) - W(\phi_{13}))] \geq 0. \quad \square$$

4.2   Normalized Metrics

The metrics introduced above are unbounded and provide an absolute measure of dissimilarity between two attributed trees, in the sense that a particular perturbation on

13

a tree "moves" it in tree-space by a distance which is independent of the whole tree mass. Therefore, it is sometimes useful to have a metric which is bounded from above and provides a measure of relative dissimilarity. For these reasons, we now introduce the following measures

$$d_3(T_1, T_2) = 1 - \frac{W(\phi_{12})}{\max(|T_1|, |T_2|)} \tag{9}$$

$$d_4(T_1, T_2) = 1 - \frac{W(\phi_{12})}{|T_1| + |T_2| - W(\phi_{12})} \tag{10}$$

which are the normalized counterparts of the metrics introduced in the previous section.

Theorem 2 *d_3 and d_4 are normalized metric in $\mathcal{T}$.*

Proof. We need to prove the properties defined in Section 2. Indeed, the normalization property is trivial, and the proof of the first three metric properties (non-negativity, identity and uniqueness, and symmetry) is similar to that of the non-normalized metrics, and therefore we omit them.

With simple algebraic operations, the triangular inequality $d_3(T_1, T_2) + d_3(T_2, T_3) \geq d_3(T_1, T_3)$ can be simplified to

$$\max(|T_1|, |T_2|) \max(|T_2|, |T_3|) \max(|T_1|, |T_3|) \geq$$
$$W(\phi_{12}) \max(|T_2|, |T_3|) \max(|T_1|, |T_3|) + W(\phi_{23}) \max(|T_1|, |T_2|) \max(|T_1|, |T_3|) -$$
$$W(\phi_{13}) \max(|T_1|, |T_2|) \max(|T_2|, |T_3|) \tag{11}$$

As was the case for the proof for metric d_1, due to the symmetry of our proof, we need to analyze the three cases: $|T_2| \geq |T_1| \geq |T_3|$, $|T_1| \geq |T_2| \geq |T_3|$, and $|T_1| \geq |T_3| \geq |T_2|$.

1. $|T_2| \geq |T_1| \geq |T_3|$

   The triangular inequality reduces to $|T_1||T_2| \geq W(\phi_{12})|T_1| + W(\phi_{23})|T_1| - W(\phi_{13})|T_2|$.

   $$|T_1||T_2| \geq |T_1|\big(W(\phi_{12}) + W(\phi_{23}) - W(\phi_{13})\big) \geq W(\phi_{12})|T_1| + W(\phi_{23})|T_1| - W(\phi_{13})|T_2|$$

2. $|T_1| \geq |T_2| \geq |T_3|$

Equation (11) reduces to $|T_1||T_2| \geq W(\phi_{12})|T_2| + W(\phi_{23})|T_1| - W(\phi_{13})|T_2|$.

$$|T_1||T_2| = |T_2|(|T_1| - |T_2|) + |T_2|^2 \geq W(\phi_{23})(|T_1| - |T_2|) + |T_2|^2 \geq$$
$$W(\phi_{23})(|T_1| - |T_2|) + |T_2|(W(\phi_{12}) + W(\phi_{23}) - W(\phi_{13})) =$$
$$W(\phi_{12})|T_2| + W(\phi_{23})|T_1| - W(\phi_{13})|T_2|$$

3. $|T_1| \geq |T_3| \geq |T_2|$

We have $|T_1||T_3| \geq W(\phi_{12})|T_3| + W(\phi_{23})|T_1| - |T_3|W(\phi_{13})$.

$$|T_1||T_3| \geq |T_1||T_2| - |T_2||T_3| + |T_2||T_3| \geq W(\phi_{23})(|T_1| - |T_3|) + |T_3||T_2| \geq$$
$$W(\phi_{23})(|T_1| - |T_3|) + |T_3|(W(\phi_{12}) + W(\phi_{23}) - W(\phi_{13})) =$$
$$W(\phi_{12})|T_3| + W(\phi_{23})|T_1| - |T_3|W(\phi_{13}).$$

In order to prove the triangular inequality for metric d_4, we define the quantity $w_{13} = \min(W(\phi_{13}), W(\phi_{12}) + W(\phi_{23}))$. Clearly, we have $|T_2| \geq W(\phi_{12}) + W(\phi_{23}) - w_{13} \geq 0$. Furthermore, we have $d_4(T_1, T_3) \leq 1 - \frac{w_{13}}{|T_1| + |T_2| - w_{13}}$. Hence, to prove the inequality, it is sufficient to prove:

$$1 - \frac{W(\phi_{12})}{|T_1| + |T_2| - W(\phi_{12})} - \frac{W(\phi_{23})}{|T_2| + |T_3| - W(\phi_{23})} - \frac{w_{13}}{|T_1| + |T_2| - w_{13}}.$$

Let us define the quantities

$$\begin{aligned} x &= |T_1| + |T_2| + |T_3| - W(\phi_{12}) - W(\phi_{23}) \geq 0 \\ x_1 &= |T_1| - W(\phi_{12}) \geq 0 \\ x_2 &= |T_2| - [W(\phi_{12}) + W(\phi_{23}) - w_{13}] \geq 0 \\ x_3 &= |T_3| - W(\phi_{23}) \geq 0 \end{aligned} \qquad (12)$$

Clearly, we can rewrite the triangular inequality as:

$$1 - \frac{W(\phi_{12})}{x - x_3} + 1 - \frac{W(\phi_{23})}{x - x_1} \geq 1 - \frac{w_{13}}{x - x_2}.$$

This inequality holds if and only if the following holds

$$(x - x_1)(x - x_2)(x - x_3) - W(\phi_{12})(x - x_1)(x - x_2)$$
$$-W(\phi_{23})(x - x_2)(x - x_3) + w_{13}(x - x_1)(x - x_3) \geq 0.$$

15

The left-hand side of this inequality can be expanded to the polynomial

$$x^2 \left[x - x_1 - x_2 - x_3 - W(\phi_{12}) - W(\phi_{23}) + w_{13}\right]$$
$$+ x \left[W(\phi_{12})(x_1 + x_2) + W(\phi_{23})(x_2 + x_3) - w_{13}(x_1 + x_3) + x_1 x_3\right]$$
$$+ x_1 x_2 \left[x - W(\phi_{12}) - x_3\right] + x_2 x_3 \left[x - W(\phi_{23})\right] + x_1 x_3 w_{13}.$$

This polynomial is the sum of non-negative terms and hence it will greater or equal to 0. In fact, by expanding the definition, we have:

$$x - x_1 - x_2 - x_3 - W(\phi_{12}) - W(\phi_{23}) + w_{13} = 0.$$

Furthermore, remembering that $|T_2| \geq W(\phi_{12}) + W(\phi_{23}) - w_{13} \geq 0$ and $|T_1| \geq W(\phi_{12}) + W(\phi_{13}) - W(\phi_{23}) \geq W(\phi_{12}) + w_{13} - W(\phi_{23})$, we have:

$$W(\phi_{12})(x_1 + x_2) + W(\phi_{23})(x_2 + x_3) - w_{13}(x_1 + x_3) + x_1 x_3 =$$
$$(W(\phi_{12}) + W(\phi_{23}) - w_{13}) \left[|T_2| - (W(\phi_{12}) + W(\phi_{23}) - w_{13})\right]$$
$$+ (|T_3| - W(\phi_{23})) \left[|T_1| - (W(\phi_{12}) + w_{13} - W(\phi_{23}))\right]$$
$$+ W(\phi_{12})(|T_1| - W(\phi_{12})) \geq 0.$$

Finally, we have:

$$x - W(\phi_{12}) - x_3 = (|T_1| - W(\phi_{12})) + (|T_2| - W(\phi_{12})) \geq 0,$$

and:

$$x - W(\phi_{23}) = (|T_1| - W(\phi_{12})) + (|T_2| - W(\phi_{23})) + (|T_3| - W(\phi_{23})) \geq 0.$$

Hence the triangular inequality holds. $\square$

5 Experimental Results

We evaluated the new metrics on three different tree-based shape representations. The first is the shock tree representation used by Pelillo, Siddiqi and Zucker in [19], which is based on the differential structure of the boundary of a 2D shape. It is obtained by extracting the skeleton of the shape, determined as the set of singularities (shocks) arising

from the inward evolution of the shape boundary, and then examining the differential behavior of the radius of the bitangent circle to the object boundary, as the skeleton is traversed. This yields a classification of local differential structure into four different classes [25]. The so-called shock-classes, distinguish between the cases where the local bitangent circle has maximum, minimum, constant, or monotonic radius. The labeled shock-groups are then abstracted using a rooted tree where two vertices are adjacent if the corresponding shock-groups are adjacent in the skeleton, and the distance from the root is related to the distance from the shape barycenter. Here, we used the same attributes and node-distances employed in [19]. Each shock was attributed with its coordinates, distance from the border, and propagation velocity and direction. The distance between two nodes, was defined as a convex combination of the (normalized) Euclidean distances of length, distance to the border, propagation speed, and curvature.

We compared our distance metrics with edit-distance. To approximate the edit-distance we used the relaxation labeling algorithm presented in [29] with the following costs: we defined the cost of matching node u to node w to be equal to the distance between their attributes, while the cost of removing any node to be equal to 1. Note that, with these costs, edit-distance is not normalized.

Our shape database contained 29 shapes from 8 different classes. Figure 2 shows the distance matrices obtained using our metrics and edit-distance. The first row contains our non-normalized matrices, the second row their normalized counterparts, and the last one edit-distance. Here, lighter colors represent lower distances while darker colors represent higher distances. As can be seen, the same block structure emerges in all five matrices. In particular, the main diagonal blocks are almost identical in all five cases, while the off-diagonal blocks present a wider variation. Essentially, the most significant differences among the five metrics are the dark bands clearly visible in the non-normalized matrices. To better visualize the distances we performed 2D multidimensional scaling (MDS) on the five matrices. The results can be observed in Figure 3.

In order to assess the ability of the distances to preserve class structure, we performed pairwise clustering. In particular, we used two pairwise clustering algorithms:
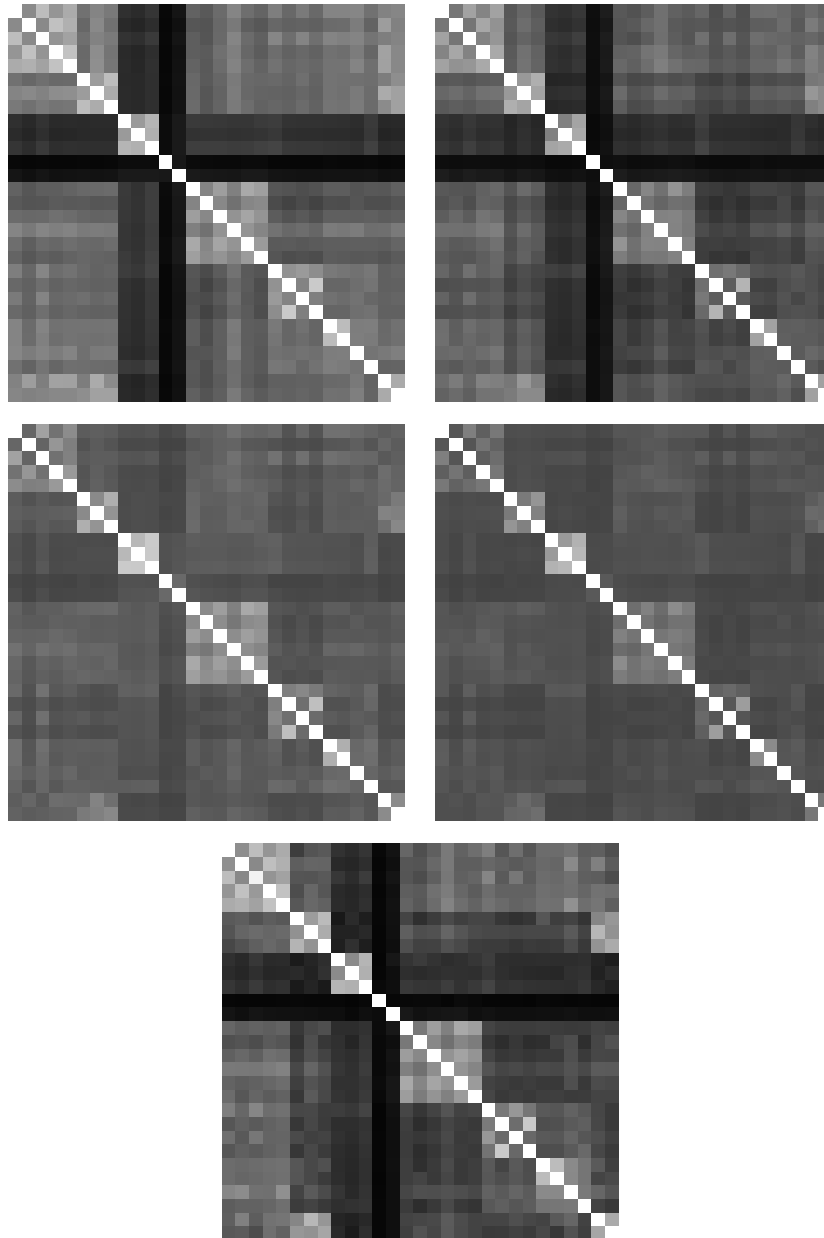
Figure 2: Distance matrices from the first experiment. Top to bottom, left to right: $d1$, $d2$, $d3$, $d4$, and edit-distance.

Shi and Malik's Normalized Cut [24], and Pavan and Pelillo's Dominant Sets [18]. Figure 4 shows the clusters obtained with Normalized Cut, displayed in order of extraction, while Figure 5 presents the clusters obtained with the Dominant Sets approach. While the performance of the clustering algorithms, on this shape recognition task, varied sig-
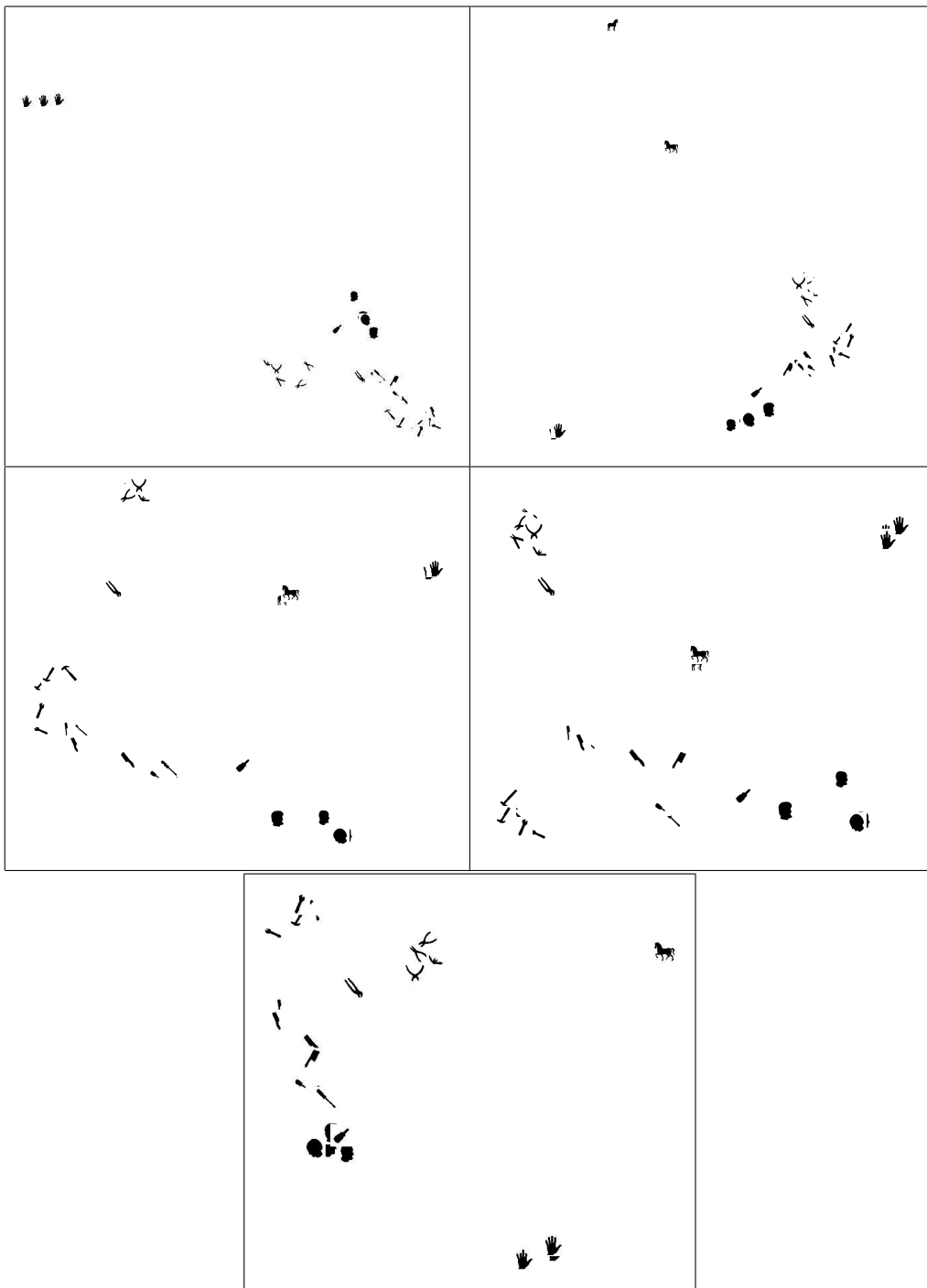
18

Figure 3: Multidimensional scaling from the first experiment. Top to bottom, left to right: $d1$, $d2$, $d3$, $d4$, and edit-distance.

nificantly, the dependency on the choice of the distance measure was less pronounced. Nonetheless, some differences can be observed. In particular, we notice how Normalized

d_1	d_2	d_3	d_4	Edit-distance

Figure 4: Clusters obtained with Normalized Cut in the first experiment.

Cut exhibits a well-known tendency to over-segment the data, a behavior particularly visible on the non-normalized metrics d_1 and d_2. A particularly interesting example is from the classification of the two horses: the shock-tree representation of the horses have the highest average number of nodes of all shape classes, and they present the highest variation in terms of number of nodes. For this reason, as can be seen by looking at the MDS results, the non-normalized measures strongly separate the two instances, while the normalized versions are able to keep them close together. The clusters obtained with the Dominant Sets approach are much better, with our normalized metrics providing results almost identical to edit-distance.

As for the running times, on a Pentium 4 2.5GHz PC, the maximum similarity algorithm presented in Section 3, took around 8 seconds to compute our metrics, while the relaxation labeling algorithm computed edit-distance in over 30 minutes.

d_1	d_2	d_3	d_4	Edit-distance

Figure 5: Clusters obtained with Dominant Sets in the first experiment.

Our second set of experiments used a larger database of shapes abstracted again in terms of shock-trees. Here, however, we used a different set of attributes introduced in [3] and recently analyzed in [28], i.e., the proportion of the shape boundary generating the corresponding shock-group. The database consisted of 150 shapes divided into 10 classes of 15 shapes each, and presented a higher structural noise than the previous one. Here the node distance and node-matching cost for edit-distance was defined as the absolute difference between the attributes, while the node removal cost was the value of the attribute itself. With this edit costs edit-distance is a normalized metric.

Figure 6 shows the distance matrices obtained using our metrics and edit-distance, and Figure 7 shows the results of MDS applied to them. Note that all measures extract the same block structure, with non-normalized metrics showing the same off-diagonal dark bands as in the previous experiments. In particular, the metrics d_1 and d_2 do not distribute the shapes uniformly, but, rather, on a tight band along a curve. There are two reasons for this behavior: First, the metrics are inherently non-Euclidean, while MDS performs an "optimal" embedding on a Euclidean space; Second, the metrics d_1 and d_3 take the tree-similarity, which is smaller than the cardinality of the smallest tree, and balances it against the cardinality of the maximum tree. The other two proposed

	Misclassification rate	Rand index
d_1	25.3%	90.1%
d_2	28.7%	90.1%
d_3	23.3%	90.3%
d_4	22.7%	90.5%
edit	22.7%	90.4%

Table 1: Validation measures of clusters obtained with Normalized Cut in the second experiment.

metrics balance the weight against the average cardinality, thereby providing a "tighter" measure.

Next, we applied the same clustering algorithms used in the previous series of experiments. In order to assess the quality of the groupings, we used two well-known cluster-validation measures [13]. The first is the standard misclassification rate. We assigned to each cluster the class that has most members in the cluster. The members of the cluster that belong to a different class are considered misclassified. The misclassification rate is the percentage of misclassified shapes over the total number of shapes. To avoid the bias towards higher segmentation that this measure exhibits, we also used a second validation measure, i.e., Rand index. We count the number of pairs of shapes that belong to the same class and that are clustered together and the number of pairs of shapes belonging to different classes that are in different clusters. The sum of these two figures divided by the total number of pairs gives us the Rand index. Here, the higher the value, the better the classification.

Table 1 summarizes the results obtained using Normalized Cut, while Table 2 presents the results obtained with the Dominant Sets approach. The Dominant Sets method provides better results in this case as well, while the different metrics generate clusters with comparable validation measures.

The last set of experiments was performed on a tree representation of Northern Lights [20]. As in the previous experiments, the representation used is derived from the
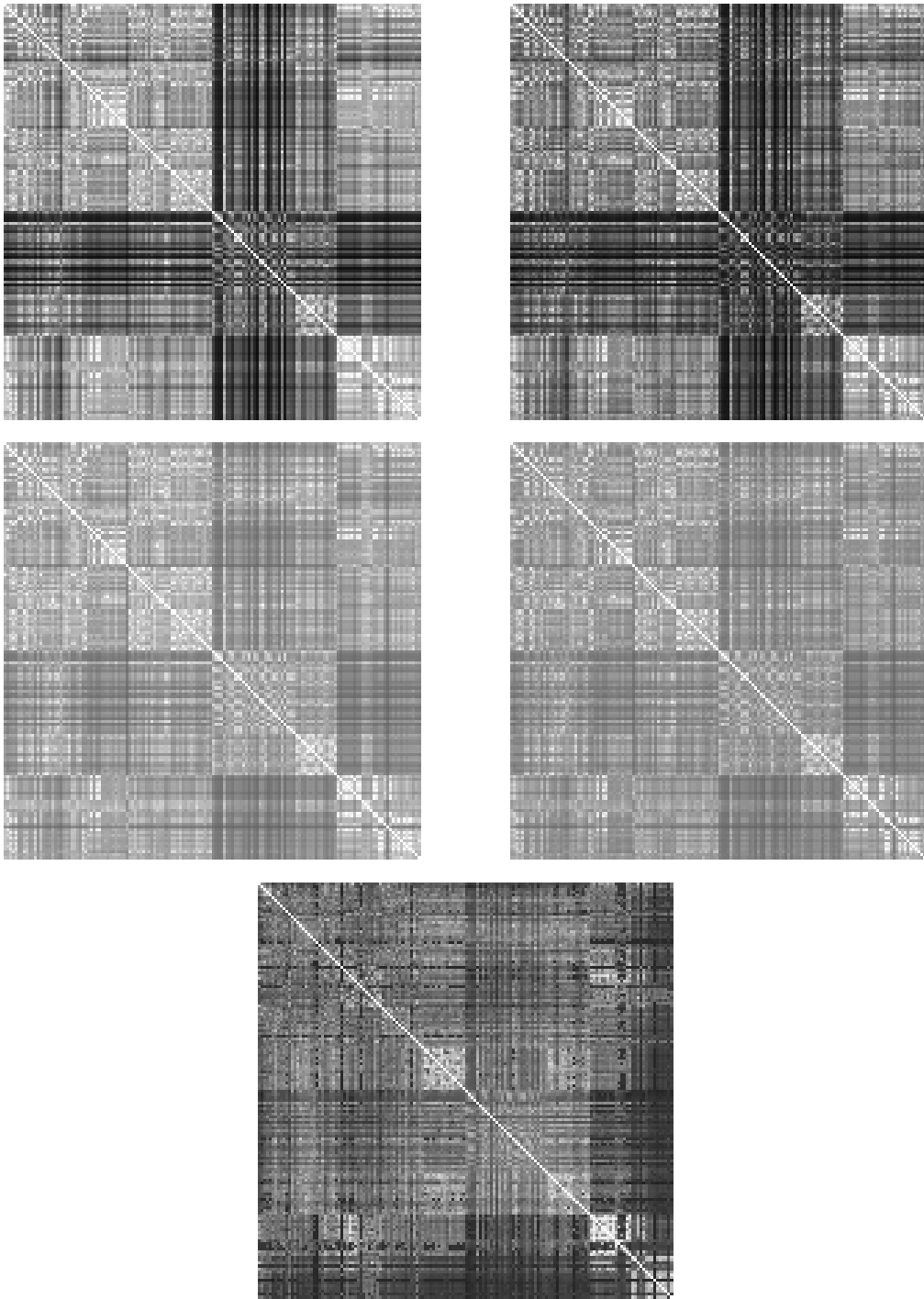
Figure 6: Distance matrices from the second experiment. Top to bottom, left to right: $d1$, $d2$, $d3$, $d4$, and edit-distance.
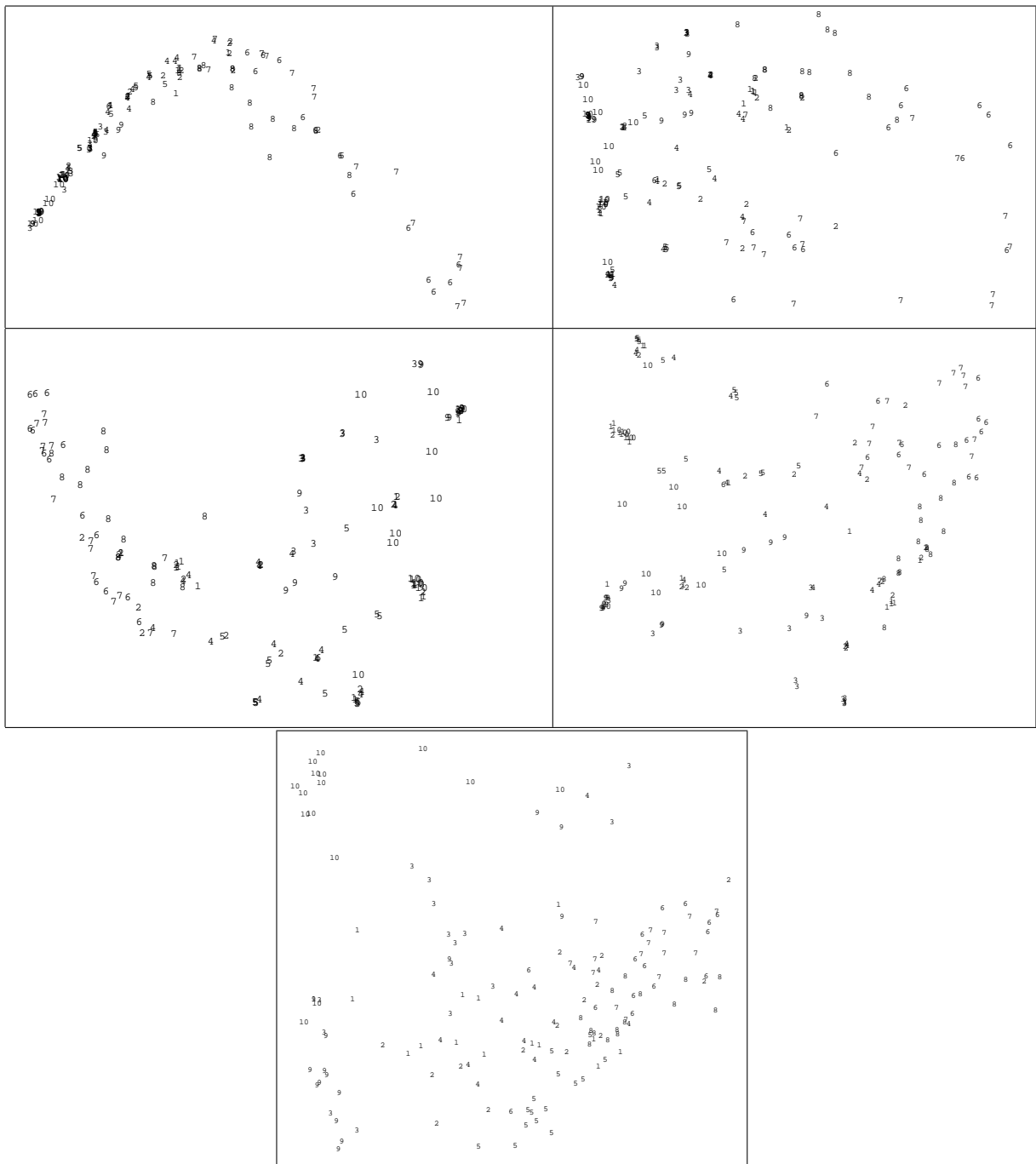
Figure 7: Multidimensional scaling from the second experiment. Top to bottom, left to right: $d1$, $d2$, $d3$, $d4$, and edit-distance. The numbers in each plot represent class labels.

morphological skeleton, but the choice of structural representation was different from the one adopted for shock-graphs, and the extracted trees tend to be larger.

The database consisted of 1440 shapes. Using our metrics we were able to extract the

	Misclassification rate	Rand index
d_1	20.7%	90.8%
d_2	22.7%	90.8%
d_3	21.3%	90.8%
d_4	20.7%	90.8%
edit	24.0%	90.8%

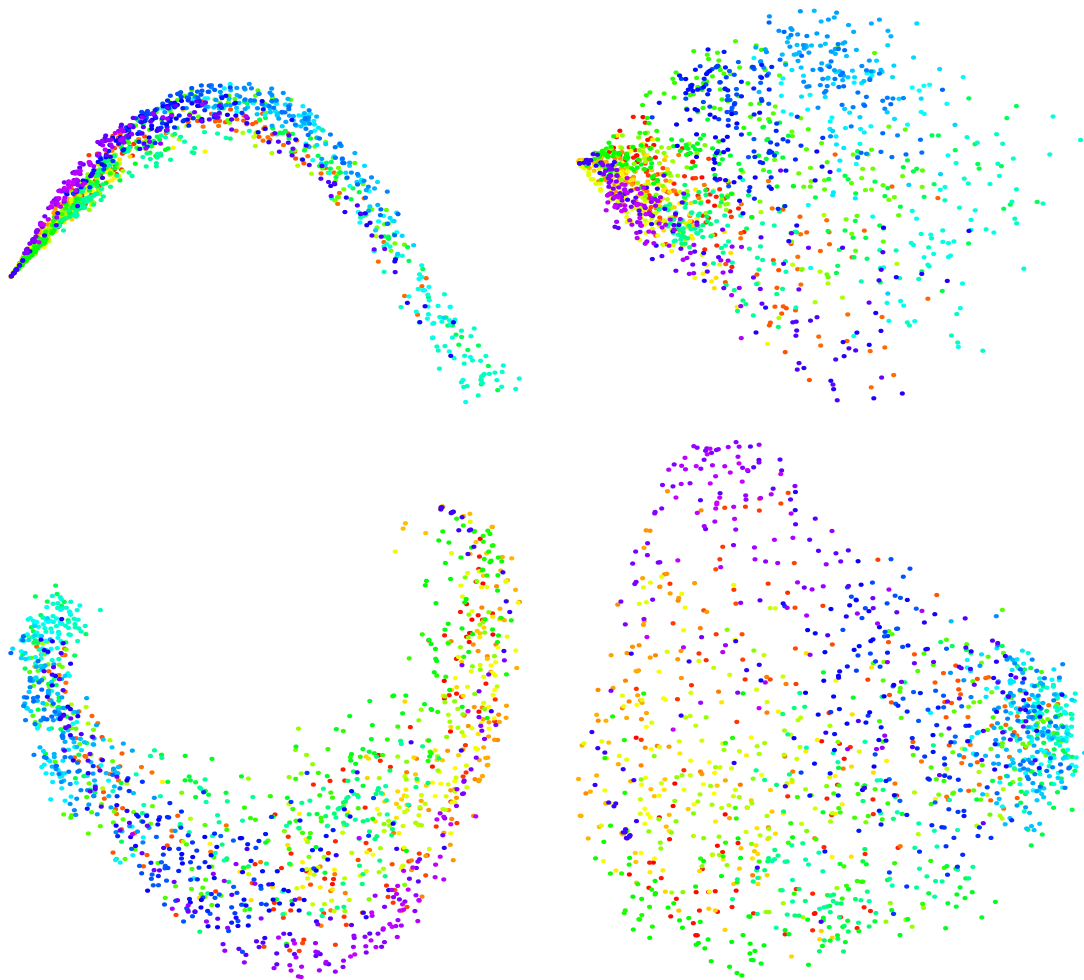Table 2: Validation measures of clusters obtained with Dominant Sets in the second experiment.

Figure 8: Multidimensional scaling of the distances obtained with our metrics from the third experiment. Top to bottom, left to right: $d1$, $d2$, $d3$, and $d4$.

	50 trees	1440 trees
d_1	0.0270	0.0159
d_2	0.0232	0.0135
d_3	0.0486	0.0165
d_4	0.0349	0.0155
edit	0.0232	—

Table 3: Davies-Bouldin index of clusters obtained with Normalized Cut in the third experiment.

full distance matrices within a few hours, but it was infeasible to compute edit-distance on the entire database. For this reason, in order to be able to compare the results with edit-distance, we also performed experiments using a smaller database consisting of 50 shapes. The calculation of edit-distance, even on this reduced database, took a full weekend.

Figure 8 displays the results of applying MDS to the distance matrices obtained with our measures. Here the hue of the point varies uniformly from red on the first shape to purple on the last. While there is no clear separation, there is a clear locality in shape-space of trees with similar indices.

In this case, we did not have the ground truth for the class memberships, so we needed a different cluster-validation measure. We opted for a standard measure that favors compact and well-separated clusters: the Davies-Bouldin index [13]. Let e_i be the average distance between elements in class i, and d_{ij} the average distance between elements in cluster i and elements in cluster j The Davies-Bouldin index is

$$DB = \frac{1}{c} \sum_{i=1}^{c} \max_j R_{ij}$$

where c is the number of clusters and $R_{ij} = \frac{e_i + e_j}{d_{ij}}$ is the cluster separation measure. Clearly, lower values correspond to better separated and more compact clusters.

Table 3 provides the values of the Davies-Bouldin index on the clusters extracted using Normalized Cut, while Table 4 shows the value obtained using the Dominant Sets

	50 trees	1440 trees
d_1	0.0695	0.0057
d_2	0.0670	0.0055
d_3	0.0723	0.0074
d_4	0.0670	0.0068
edit	0.0635	—

Table 4: Davies-Bouldin index of clusters obtained with Dominant Sets in the third experiment.

algorithm.

6   Conclusions

In this paper we have presented four novel distance measures for attributed trees based on the notion of a maximum similarity subtree isomorphism, and provided a polynomial-time algorithm to calculate them. We have proven that this measures satisfy the metric properties and have experimentally validated their usefulness by comparing them with edit-distance on three different shape recognition tasks. Our experimental results show that, in terms of quality, the proposed metrics compare well with edit-distance, their computation being, however, orders of magnitude faster.

References

[1] R. K. Ahuja, T. L. Magnanti, J. B. Orlin. *Network Flows*. Prentice-Hall, Upper Saddle River, NJ, 1993.

[2] H. G. Barrow and R. M. Burstall. Subgraph isomorphism, matching relational structures and maximal cliques. *Inform. Process. Letters*, vol. 4, pp. 83–84, 1976.

[3] H. Blum and R. N. Nagel. Shape description using weighted symmetric axis features. *Pattern Recognition*, vol. 10, pp. 167–180, 1978.

[4] K. L. Boyer and A. C. Kak. Structural stereopsis for 3-D vision. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 144–166, 1988.

[5] H. Bunke and A. Kandel. Mean and maximum common subgraph of two graphs. *Pattern Recognition Letters*, vol. 21, pp. 163–168, 2000.

[6] H. Bunke and K. Shearer. A graph distance metric based on the maximal common subgraph. *Pattern Recognition Letters*, vol. 19, pp. 255–259, 1998.

[7] W. J. Christmas and J. Kittler. Structural matching in computer vision using probabilistic relaxation. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 749–764, 1995.

[8] S. J. Dickinson, A. P. Pentland, and A. Rosenfeld. 3-D shape recovery using distributed aspect matching. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 2, pp. 174–198, 1992.

[9] M. A. Eshera and K.-S. Fu. An image understanding system using attributed symbolic representation and inexact graph-matching. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 604–618, 1986.

[10] M. L. Fernandez and G. Valiente. A graph distacne metric combining maximum common subgraph and minimum common supergraph. *Pattern Recognition Letters*, vol. 22, pp. 753–758, 2001.

[11] D. Hidović and M. Pelillo. Metrics for attributed graphs based on the maximal similarity common subgraph. *Int. J. Pattern Recognition Artif. Intell.*, 2004 (in press).

[12] S. Ioffe and D. A. Forsyth. Human tracking with mixtures of trees. in *Proc. Int. Conf. Computer Vision*, Vol. I, pp. 690–695, 2001.

[13] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data.* Prentice Hall, Englewood Cliffs, NJ, 1988.

[14] B. B. Kimia, A. R. Tannenbaum, and S. W. Zucker. Shapes, shocks, and deforamtions I. *Int. J. Computer Vision*, vol. 15, pp. 189-224, 1995.

[15] D. W. Matula. An algorithm for subtree identification, *SIAM Review*, vol. 10, pp. 273–274, 1968.

[16] B. Moayer and K.-S. Fu. A tree system approach for fingerprint pattern recognition, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 3, pp. 376–387, 1986.

[17] C. H. Papadimitriou and K. J. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity.* Englewood Cliffs, NJ: Prentice-Hall, 1982.

[18] M. Pavan and M. Pelillo. A new graph-theoretic approach to clustering and segmentation. *Proc. CVPR 2003 - IEEE Conf. on Computer Vision and Pattern Recognition.* IEEE Computer Society Press, 2003, Vol. I, pp. 145-152.

[19] M. Pelillo, K. Sidiqi, and S. W. Zucker. Matching hierarchical structures using association graphs. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 11, pp. 1105–1120, 1999.

[20] M. Peura. Attribute trees in image analysis: Heuristic matching and learning techniques. In *Proc. Int. Conf. Image Anal. Processing*, pp. 1160–1165, 1999.

[21] H. Samet. Distance transform for images represented by quadtrees. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 4, no. 3, pp. 298–303, 1982.

[22] T. B. Sebastian, P. N. Klein, and B. B. Kimia. Recognition of shpes by editing their shock graphs. *IEEE Trans. Pattern Anal. Machine Intell.*, to appear, 2004.

[23] L. G. Shapiro and R. M. Haralick. Relational models for scene analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 4, pp. 595–602, 1982.

[24] J. Shi and J. Malik, Normalized cuts and image segmentation, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 8, pp. 888–905, 2000.

[25] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker. Shock graphs and shape matching. *Int. J. Computer Vision*, vol. 35, no. 1, pp. 13–32, 1999.

[26] A. Shokoufandeh, S. J. Dickinson, K. Siddiqi, and S. W. Zucker. Indexing using a spectral encoding of topological structure. In *Proc. IEEE Conf. Computer Vision Pattern Recognition*, pp. 491–497, 1999.

[27] E. Tanaka. A metric between unrooted and unordered trees and its bottom-up computing method. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 12, pp. 1233–1238, 1994.

[28] A. Torsello and E. R. Hancock. A skeletal measure of 2D shape similarity. In C. Arcelli, L. P. Cordella, G. Sanniti di Baja (Eds.), *Visual Form 2001*. Springer-Verlag, Heidelberg (LNCS 2059), pages 260–271, 2001.

[29] A. Torsello and E. R. Hancock. Efficiently computing weighted tree edit-distance using relaxation labeling. In M. Figueiredo, J. Zerubia, A.K. Jain (Eds.), *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Springer-Verlag, Heidelberg (LNCS 2134), pp. 438–453, 2001.

[30] W. H. Tsai and K.-S. Fu. Error-correcting isomorphism of attributed relational graphs for pattern analysis. *IEEE Trans Syst. Man Cybern.*, vol. 9, pp. 757–768, 1979.

[31] G. Valiente. An efficient bottom-up distance between trees, in *Proc. Int. Symp. String Processing Information Retrieval*, pp. 212–219, 2001.

[32] W. D. Wallis, P. Shoubridge, M. Kraetz, and D. Ray. Graph distances using graph union. *Pattern Recognition Letters*, vol. 22, pp. 701–704, 2001.

[33] J. T.-L. Wang and K. Zhang. Finding similar consesnus between trees: An algorithm and a distance hierarchy. *Pattern Recognition*, vol. 34, no. 1, pp. 127–137, 2001.

[34] R. C. Wilson and E. R. Hancock. Structural matching by discrete relaxation. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 6, pp. 634–648, 1997.

[35] A. K. C. Wong and M. You. Entropy and distance of random graphs with application to structural pattern recognition. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 7, pp. 599–609, 1985.

[36] K. Zhang. A constrained edit-distance between unordered labeled trees. *Algorithmica*, vol. 15, pp. 205-222, 1996.

[37] K. Zhang and D. Shasha. Simple fast algorithms for the editing distance between trees and related problems. *SIAM J. Comput.*, vol. 18, pp. 1245–1262, 1989.

[38] K. Zhang, R. Statman, and D. Shasha. On the editing distance between unordered labeled trees. *Inform. Process. Letters*, vol. 42, pp. 133–139, 1992.

[39] K. Zhang, J. T. L. Wang, and D. Shasha. On the editing distance between undirected acyclic graphs. *Int. J. Found. Computer Sci.*, 7(1):43–57, 1996.

[40] S. C. Zhu and A. L. Yuille. FORMS: A flexible object recognition and modelling system. *Int. J. Computer Vision*, vol. 20, no. 3, pp. 187–212, 1996.