Polynomial-Time Metrics for Attributed Trees

Andrea Torsello, Džena Hidović-Rowe, and Marcello Pelillo, *Senior Member*, *IEEE*

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 applied to rooted as well as 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, in general, NP-complete, at least in the unordered case. We experimentally validate the usefulness of our metrics on shape matching tasks and compare them with (an approximation of) edit-distance.

Index Terms—Metrics, tree matching, polynomial-time algorithms, shape recognition.

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], [44], 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. Indeed, the problem of how to measure the similarity (or distance) of pictorial information using graph abstractions has been a widely researched topic for over 20 years.

Early work on the topic included Barrow and Burstall's idea of characterizing the similarity of two graphs using the cardinality of their maximum common subgraphs [2] and the extension of the concept of string edit-distance to graph-matching by Eshera and Fu [9]. Shapiro and Haralick [23] described a relational distance measure between structural descriptions. There have also been attempts to use an information theoretic approach. Here, Wong and You [39] computed the entropy for random graphs, while Boyer and Kak [4] used mutual information. More recently, Christmas et al. [7] and Wilson and Hancock [38] 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 nonsymmetric, negative, or violate the triangular inequality. The lack of metric properties makes 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 a cost with these operations. The edit-distance is found by searching for sequences of edit operations that make the two graphs isomorphic to one another and the distance between the two graphs is then defined 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 [42], 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 introduced a distance measure on unattributed graphs based on the maximum common subgraph and proved that it is a metric. Wallis et al. [35] introduced a variant of this distance based on the size of the minimum common supergraph. Finally, Fernández and Valiente [10] defined a metric based on the difference in size between maximum common subgraph and minimum common supergraph. More recently, in [11], Hidović and Pelillo extended 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], [44], [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

- *A. Torsello and M. Pelillo are with the Dipartimento di Informatica, Università Ca' Foscari di Venezia, Via Torino 155, 30172 Venezia Mestre, Italy. E-mail: {torsello, pelillo}@dsi.unive.it.*
- *D. Hidović-Rowe is with the School of Computer Science, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom. E-mail: D.Hidovic@cs.bham.ac.uk.*

Manuscript received 20 Dec. 2003; revised 24 Feb. 2005; accepted 01 Mar. 2005; published online 12 May 2005.
Recommended for acceptance by M. Basu.
For information on obtaining reprints of this article please send e-mail to: tpami@computer.org and reference IEEECS Log Number TPAMISI-0440-1203.

defined in terms of edit-distance [32], [36]. Zhang and Shasha [41] 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 et al. [22] used a similar algorithm to compare shock trees. Other constrained solutions proven to be solvable in polynomial time have been proposed (see, for example, [40]), yet, in the general case, the problem is NP-complete both for rooted [42] and unrooted trees [43]. Recently, Valiente [33] 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 nonnormalized, 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 makes 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 (unordered) rooted/unrooted and attributed/unattributed trees. They can also be viewed as variants of the metrics developed by Bunke and Shearer [6], Wallis et al. [35], and Fernández 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 only, where our approach provides a clear computational advantage. Note, however, that the ordered case can be dealt with in a straightforward way within our framework by using classic ordered tree isomorphism algorithms [34]. 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. Preliminary versions of this paper were presented in [30], [31].

The outline of the paper is as follows: Section 2 introduces formalisms and concepts required throughout the paper. In Section 3, we define our measures and prove that they satisfy the metric properties. In Section 4, we present a polynomial-time algorithm to calculate the maximum similarity common subtree needed to compute all our metrics. Finally, Section 5 provides experimental validation of the usefulness of the metrics and, in Section 6, 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 $u_0 u_1 \dots u_n$ such that, for all $i = 1 \dots n$, $u_{i-1} \sim u_i$; 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., $\sigma(u, v) = 1$, for all pairs of vertices u and v.

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

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

is symmetric, nonnegative, normalized, and fulfills the property $\sigma(x, y) = 1 \Leftrightarrow x = y$ (identity and uniqueness). Furthermore, it also 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 (2) and the triangular inequality:

$$\begin{aligned}
&\sigma(x, y) + \sigma(y, z) - \sigma(x, z) \leq 1 \Leftrightarrow \\
&(1 - d(x, y)) + (1 - d(y, z)) - (1 - d(x, z)) \leq 1 \Leftrightarrow \\
&- d(x, y) - d(y, z) + d(x, z) \leq 0 \Leftrightarrow d(x, y) + d(y, z) \geq d(x, z).
\end{aligned}$$

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 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 nonnormalized 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 subtree isomorphisms between T_1 and T_2, T_2 and T_3, and T_1 and T_3, respectively. Then, we have: $|T_2| \geq W(\phi_{12}) + W(\phi_{23}) - W(\phi_{13})$ (and also $|T_1| \geq W(\phi_{12}) + W(\phi_{13}) - W(\phi_{23})$ and $|T_3| \geq W(\phi_{13}) + W(\phi_{23}) - W(\phi_{12})$).*

Proof. We only need to prove $|T_2| \geq W(\phi_{12}) + W(\phi_{23}) - W(\phi_{13})$, as the other two inequalities can be reduced to it with a change of variables. Let $V_2^1 = \mathrm{Im}(\phi_{12}) \subseteq V_2$, $V_2^3 = \mathrm{Dom}(\phi_{23}) \subseteq V_2$ 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,

$$\begin{aligned}
&W(\phi_{12}) + W(\phi_{23}) - W(\phi_{13}) \leq W(\phi_{12}) + W(\phi_{23}) - W(\hat{\phi}_{13}) = \\
&\sum_{v \in V_2^1} \sigma(\phi_{12}^{-1}(v), v) + \sum_{v \in V_2^3} \sigma(v, \phi_{23}(v)) - \sum_{v \in \hat{V}_2} \sigma(\phi_{12}^{-1}(v), \phi_{23}(v)) = \\
&\sum_{v \in V_2^1 \setminus V_2^3} \sigma(\phi_{12}^{-1}(v), v) + \sum_{v \in V_2^3 \setminus V_2^1} \sigma(v, \phi_{23}(v)) + \\
&\sum_{v \in \hat{V}_2} \left[\sigma(\phi_{12}^{-1}(v), v) + \sigma(v, \phi_{23}(v)) - \sigma(\phi_{12}^{-1}(v), \phi_{23}(v))\right] \leq \\
&|V_2^1 \setminus V_2^3| + |V_2^3 \setminus V_2^1| + |V_2^1 \cap V_2^3| = |V_2^1 \cup V_2^3| \leq |T_2|,
\end{aligned}$$

where the inequality follows from the normalized property and the triangular inequality for metric-derived similarities (3). □

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

$$d_1(T_1, T_2) = \max(|T_1|, |T_2|) - W(\phi_{12}), \tag{5}$$

$$d_2(T_1, T_2) = |T_1| + |T_2| - 2W(\phi_{12}), \tag{6}$$

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.

where ϕ_{12} is a maximum similarity common subtree isomorphism between T_1 and T_2. Note that 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 of 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 similarity of a subtree isomorphism 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:

   a. $|T_1| \geq |T_2| \geq |T_3|$,
   b. $|T_1| \geq |T_3| \geq |T_2|$,
   c. $|T_2| \geq |T_1| \geq |T_3|$,
   d. $|T_2| \geq |T_3| \geq |T_1|$,
   e. $|T_3| \geq |T_1| \geq |T_2|$, and
   f. $|T_3| \geq |T_2| \geq |T_1|$.

   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: $|T_1| \geq |T_2| \geq |T_3|$, $|T_1| \geq |T_3| \geq |T_2|$, and $|T_2| \geq |T_1| \geq |T_3|$.

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

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

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

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

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

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

   On the other hand, for d_2, we have

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

□

3.2 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 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:

$$d_3(T_1, T_2) = 1 - \frac{W(\phi_{12})}{\max(|T_1|, |T_2|)}, \tag{7}$$

$$d_4(T_1, T_2) = 1 - \frac{W(\phi_{12})}{|T_1| + |T_2| - W(\phi_{12})}, \tag{8}$$

which are the normalized counterparts of the metrics introduced previously.

Theorem 2. *d_3 and d_4 are normalized metrics 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 (nonnegativity, identity and uniqueness, and symmetry) is similar to that of the nonnormalized 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

$$\begin{aligned} &\max(|T_1|, |T_2|) \max(|T_2|, |T_3|) \max(|T_1|, |T_3|) \\ &\geq W(\phi_{12}) \max(|T_2|, |T_3|) \max(|T_1|, |T_3|) \\ &+ W(\phi_{23}) \max(|T_1|, |T_2|) \max(|T_1|, |T_3|) \\ &- W(\phi_{13}) \max(|T_1|, |T_2|) \max(|T_2|, |T_3|). \end{aligned} \tag{9}$$

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: $|T_1| \geq |T_2| \geq |T_3|$, $|T_1| \geq |T_3| \geq |T_2|$, and $|T_2| \geq |T_1| \geq |T_3|$.

1. $|T_1| \geq |T_2| \geq |T_3|$.
   Equation (9) reduces to $|T_1||T_2| \geq W(\phi_{12})|T_2| + W(\phi_{23})|T_1| - W(\phi_{13})|T_2|$.

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

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

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

3. $|T_2| \geq |T_1| \geq |T_3|$.
   The triangular inequality reduces to $|T_1||T_2| \geq W(\phi_{12})|T_1| + W(\phi_{23})|T_1| - W(\phi_{13})|T_2|$. From Lemma 1, we have

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

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 $d_4(T_1, T_3) \leq 1 - \frac{w_{13}}{|T_1|+|T_3|-w_{13}}$. Hence, to prove the inequality, it is sufficient to prove:

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

Let us define the quantities

$$\begin{gathered} x = |T_1| + |T_2| + |T_3| - W(\phi_{12}) - W(\phi_{23}) \geq 0 \\ x_1 = |T_1| - W(\phi_{12}) \geq 0 \\ x_2 = |T_2| - [W(\phi_{12}) + W(\phi_{23}) - w_{13}] \geq 0 \\ x_3 = |T_3| - W(\phi_{23}) \geq 0. \end{gathered} \tag{10}$$

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

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

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

$$\begin{aligned} &x^2[x - x_1 - x_2 - x_3 - W(\phi_{12}) - W(\phi_{23}) + w_{13}] \\ &+ x[W(\phi_{12})(x_1 + x_2) + W(\phi_{23})(x_2 + x_3) \\ &- w_{13}(x_1 + x_3) + x_1x_3] \\ &+ x_1x_2[x - W(\phi_{12}) - x_3] + x_2x_3[x - W(\phi_{23})] + x_1x_3w_{13}. \end{aligned}$$

This polynomial is a sum of nonnegative terms and, hence, it will be greater than or equal to 0. In fact, by expanding the definition, we have:

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

Finally, to prove that

$$W(\phi_{12})(x_1 + x_2) + W(\phi_{23})(x_2 + x_3) - w_{13}(x_1 + x_3) + x_1x_3 \geq 0,$$

we distinguish between the cases where $w_{13} \geq W(\phi_{12})$ and $w_{13} < W(\phi_{12})$. In the former case, the term can be expanded to:

$$\begin{aligned} &[|T_1| - (W(\phi_{12}) + w_{13} - W(\phi_{23}))][|T_3| - (w_{13} + W(\phi_{23}) - W(\phi_{12}))] \\ &+ (W(\phi_{12}) + W(\phi_{23}) - w_{13})[|T_2| - (W(\phi_{12}) + W(\phi_{23}) - w_{13})] \\ &+ (|T_2|w_{13} - W(\phi_{12})W(\phi_{23})). \end{aligned}$$

This expression is nonnegative since it is a sum of nonnegative terms. In fact, for Lemma 1, we have:

$$\begin{aligned} &|T_1| - (W(\phi_{12}) + w_{13} - W(\phi_{23})) \geq \\ &|T_1| - (W(\phi_{12}) + W(\phi_{13}) - W(\phi_{23})) \geq 0, \\ &|T_3| - (w_{13} + W(\phi_{23}) - W(\phi_{12})) \geq \\ &|T_3| - (W(\phi_{13}) + W(\phi_{23}) - W(\phi_{12})) \geq 0, \text{and} \\ &|T_2| - (W(\phi_{12}) + W(\phi_{23}) - w_{13}) \geq 0, \end{aligned}$$

while, by construction, we have: $W(\phi_{12}) + W(\phi_{23}) - w_{13} \geq 0$. Furthermore, since we assume $w_{13} \geq W(\phi_{12})$, we have

$$|T_2|w_{13} - W(\phi_{12})W(\phi_{23}) \geq |T_2|W(\phi_{12}) - W(\phi_{12})W(\phi_{23}) \geq 0.$$

On the other hand, if $w_{13} < W(\phi_{12})$, we can write the term as:

$$\begin{aligned} &[|T_1| - (W(\phi_{12}) + w_{13} - W(\phi_{23}))](|T_3| - W(\phi_{23})) + \\ &(W(\phi_{12}) + W(\phi_{23}))[|T_2| - (W(\phi_{12}) + W(\phi_{23}) - w_{13})] + \\ &(|T_1| - W(\phi_{12}))(W(\phi_{12}) - w_{13}) \geq 0, \end{aligned}$$

where the inequality holds because the expression is the sum of nonnegative terms. Hence, the triangular inequality holds. □

The four metrics involve two independent dichotomies, one deriving from the decision as to whether or not to use a normalized measure, the other involving a choice between the maximum and the average tree-size as a balancing factor. The decision concerning normalization depends mainly on the specific application at hand and on whether there is a need for an upper bound on the measure. On the other hand, using the average size as a balancing factor allows us to take into account the size of the smaller tree, information which is completely discarded when using the maximum size as a balancing term.

A natural question arises as to when the proposed metrics are to be preferred over edit-distance measures. Typically,

```
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)
    Cw=children(w)
    for each ui in Cu
        for each wj in Cw
            wij=AnchoredSimilarity(ui,wj)
return σ(u,w) + Assign({wij})
```

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

real-world problems approached via graph-theoretic techniques involve two complementary types of noise: one which acts on the attribute space by altering the attributes separately on each vertex and one which instead affects the topology of the whole structure by modifying the vertex relations. Our metrics, being based on the notion of subtree isomorphism, are designed to be robust under attribute perturbations as well as peripheral structural noise, whereas edit-distance measures cope well with more severe noise affecting the inner part of the structures being matched. Clearly, in general, it is not easy to understand what the typical noise for the problem at hand is and, hence, to make an informed decision as to the metric to use.

4 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 isomorphism algorithm presented by Matula [15] (see also [34]), extending it to deal with attributed trees. We give an algorithm to solve 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 as well.

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 fact, 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 $j = 1, \cdots, m$,

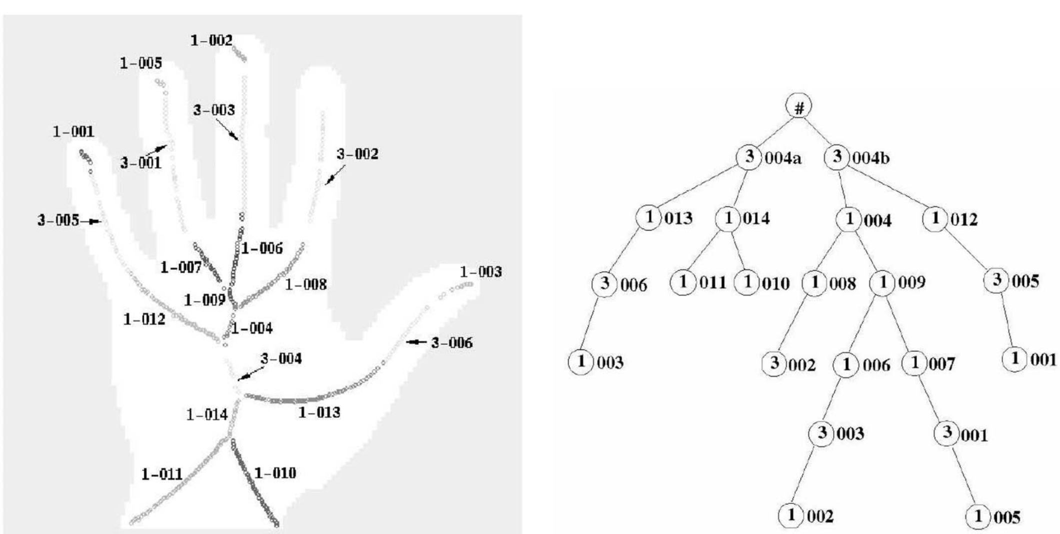

Fig. 2. An example of a shape, its skeleton, and the corresponding shock tree.

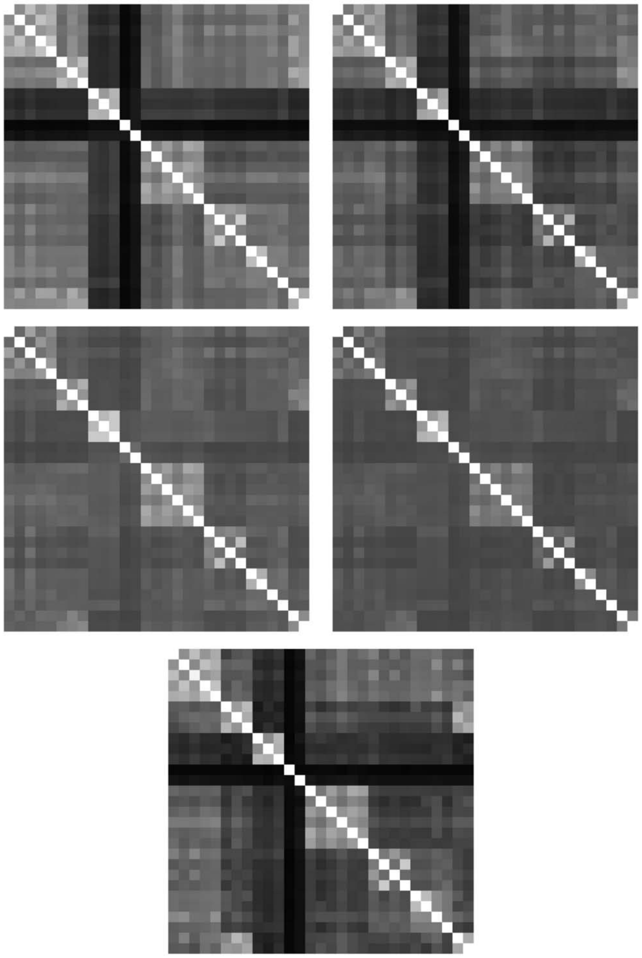

Fig. 3. Distance matrices from the first experiment. Top to bottom, left to right: d_1, d_2, d_3, d_4, and edit-distance.

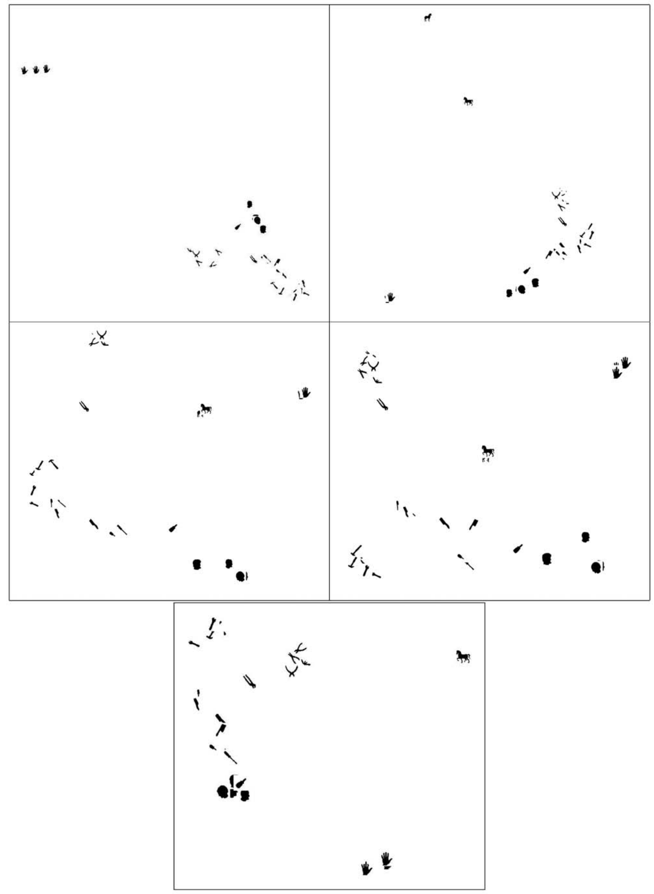

Fig. 4. Multidimensional scaling from the first experiment. Top to bottom, left to right: d_1, d_2, d_3, d_4, and edit-distance.

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

$$W(\phi^{(u,w)}) = \sigma(u,w) + \max_{\pi \in \Sigma_n^m} \sum_{i=1}^{n} W_{i\pi(i)}, \tag{11}$$

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:

$$\phi^{(u,w)}(x) = \begin{cases} w & \text{if } x = u \\ \widehat{\phi}^{(u_i,w_{\pi(i)})}(x) & \text{if } x \in \mathrm{Dom}(\widehat{\phi}^{(u_i,w_{\pi(i)})}), \end{cases} \tag{12}$$

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

Fig. 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 (11). To this end, it needs a procedure for solving an assignment (or, equivalently, a bipartite matching) problem, which the algorithms literature abound in (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 assignment 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 (11), 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].

Finally, if we have two unrooted trees $T_1 = (V_1, E_1)$ and $T_2 = (V_2, E_2)$, we can 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

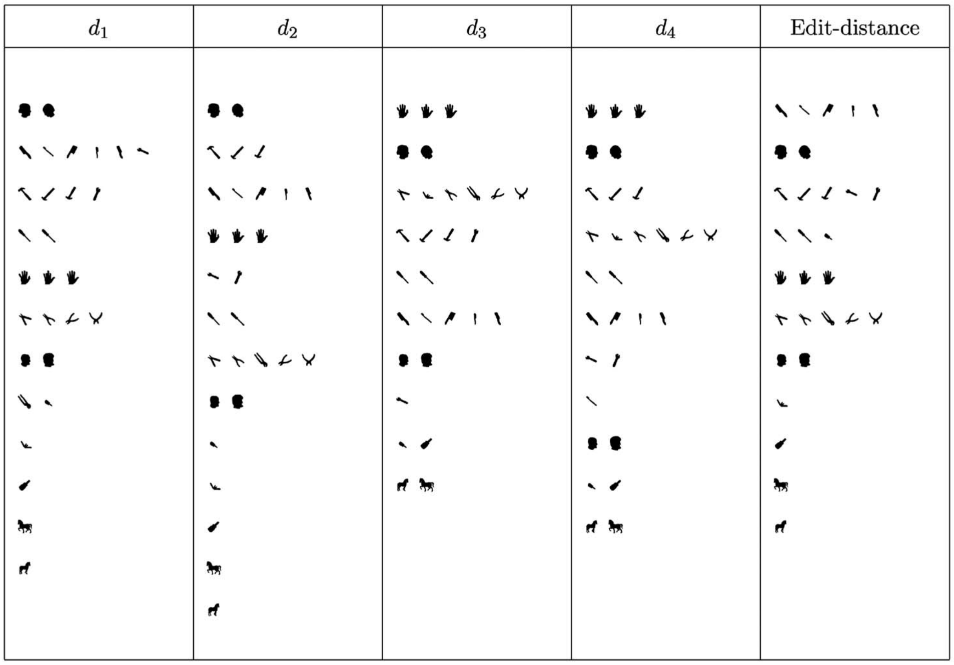

Fig. 5. Clusters obtained with Normalized Cut in the first experiment.

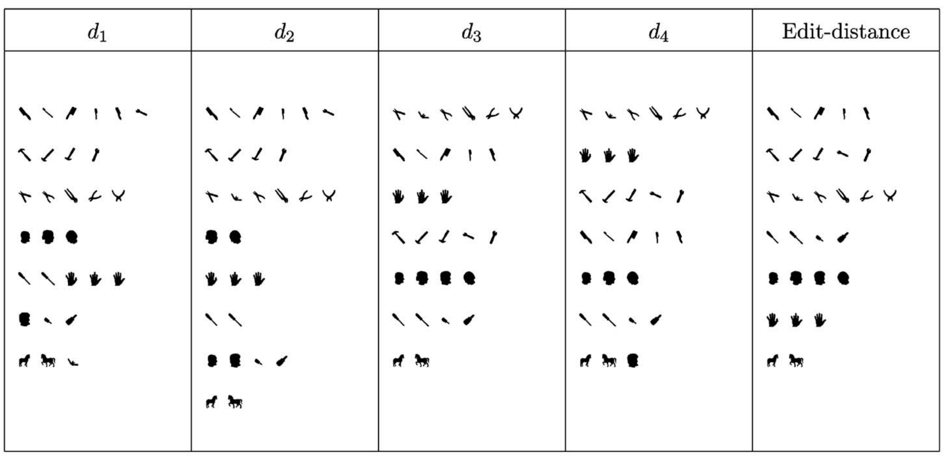

Fig. 6. Clusters obtained with Dominant Sets in the first experiment.

similarity. This yields a straightforward $O(bN^3M)$ algorithm for unrooted trees, which consists of iteratively calling $\texttt{MaxSimilarity}(T_1^u, T_2^w)$ for all $u \in V_1$ and $w \in V_2$ and taking the maximum. However, it is a well-known result that we do not actually need to try all possible pairs of roots since, by simply fixing the root in one tree and leting the other vary among all possible vertices in the other tree, the algorithm is still guaranteed to achieve the maximum similarity [34]. This follows easily from the observation made above that there 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. A similar approach has been recently used in [37].

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 et al. 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 an unordered 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. Note that, here, order need not be preserved since articulation or pose variation might change the relative position of parts. Fig. 2 shows an example silhouette, its morphological skeleton divided into its constituent shock groups, and the extracted shock tree. 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.

Due to computational complexity, we could not compare our metrics to exact edit-distance, but we had to resort to an approximation. To this end, we used the relaxation labeling algorithm presented in [29]. While this algorithm provides a reasonable approximation of edit-distance, it is clear that a comparison with exact edit-distance might provide slightly different results. Nonetheless, since the approximation error was empirically shown to be small in this application domain and any practical application would have to resort to some approximation, we believe the comparison to still be meaningful. In this experiment, the edit costs were defined as follows: The cost of matching node u to node w was set to be equal to the distance between their attributes, while the cost of removing any node was set to be equal to 1. Note that, with these costs, edit-distance is not normalized.

Our shape database contained 29 shapes from eight different classes. The size of the trees in this set ranged from 8 to 40 vertices, with an average size of 14.2 vertices. Fig. 3 shows the distance matrices obtained using our metrics and edit-distance. The first row contains our nonnormalized 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 nonnormalized matrices. To better visualize the distances, we performed 2D multidimensional scaling (MDS) on the five matrices. The results can be observed in Fig. 4.

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: Shi and Malik's Normalized Cut [24], which has become a standard benchmark among pairwise clustering algorithms and is commonly used in the vision community, and Dominant Sets, a powerful and simple pairwise clustering approach recently introduced by Pavan and Pelillo [18]. Fig. 5 shows the clusters obtained with Normalized Cut, displayed in order of extraction, and Fig. 6 presents the clusters obtained with the Dominant Sets approach. While the performance of the clustering algorithms on this shape recognition task varied significantly, the dependency on the choice of the distance measure was less pronounced. Nonetheless, some differences can be observed. In particular, we notice how Normalized Cut exhibits a well-known tendency to oversegment the data, a behavior particularly visible on the nonnormalized 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 has the largest 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 nonnormalized 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.

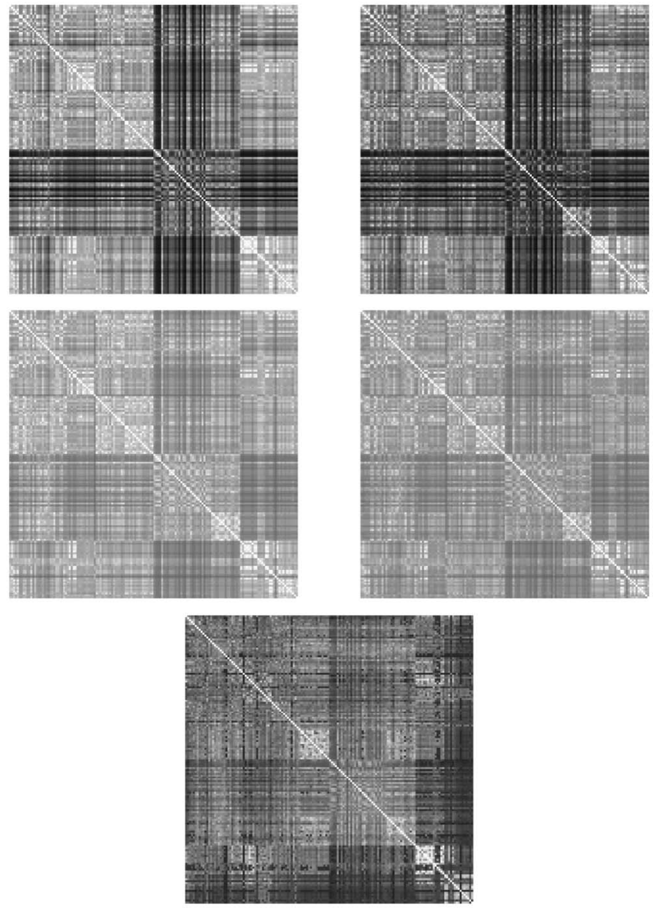

Fig. 7. Distance matrices from the second experiment. Top to bottom, left to right: d_1, d_2, d_3, d_4, and edit-distance.

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

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

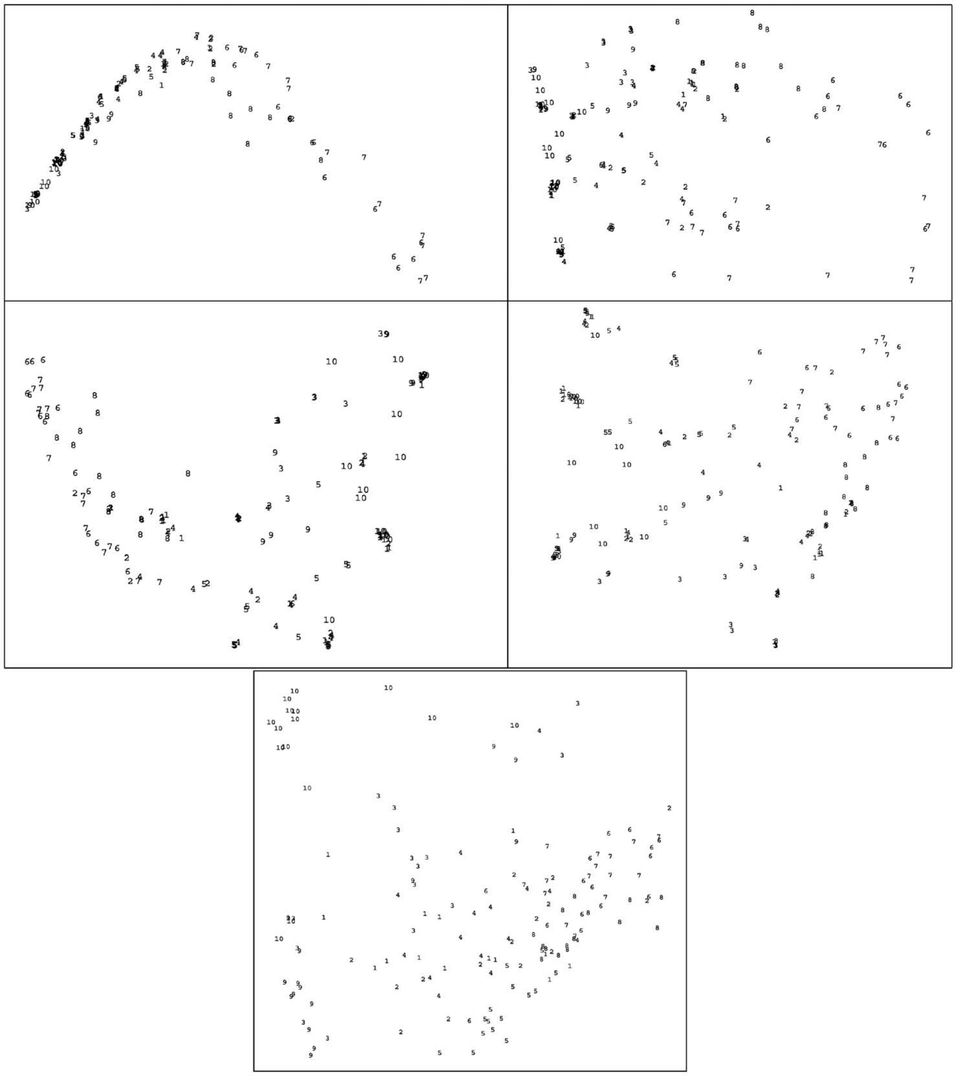

Fig. 8. Multidimensional scaling from the second experiment. Top to bottom, left to right: d_1, d_2, d_3, d_4, and edit-distance. The numbers in each plot represent class labels.

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. The tree-size ranged from 4 to 33 vertices, with an average of 13.2 vertices per tree. 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.

Fig. 7 shows the distance matrices obtained using our metrics and edit-distance and Fig. 8 shows the results of MDS applied to them. Note that all measures extract the same block structure, with nonnormalized 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, as previously discussed, the metrics d_1 and d_3 take the tree-similarity, which is smaller than the cardinality of the smallest tree, and balance it against the cardinality of the maximum tree. The other two proposed 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., the Rand index. We count the number of pairs of shapes that belong to the same class 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

TABLE 1
Validation Measures of Clusters Obtained with Normalized Cut in the Second Experiment

	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 2
Validation Measures of Clusters Obtained with Dominant Sets in the Second Experiment

	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%

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 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 1,440 shapes abstracted as trees having from 4 to 131 vertices with an average of 30.7 vertices. Using our metrics, we were able to extract the full distance matrices within a few hours, but it was infeasible to compute an approximation of the 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 an approximation of edit-distance, even on this reduced database, took a full weekend.

Fig. 9 displays the results of applying MDS to the distance matrices obtained with our measures. Here, the gray level of the point varies uniformly from black on the first shape to light gray 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} be 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.

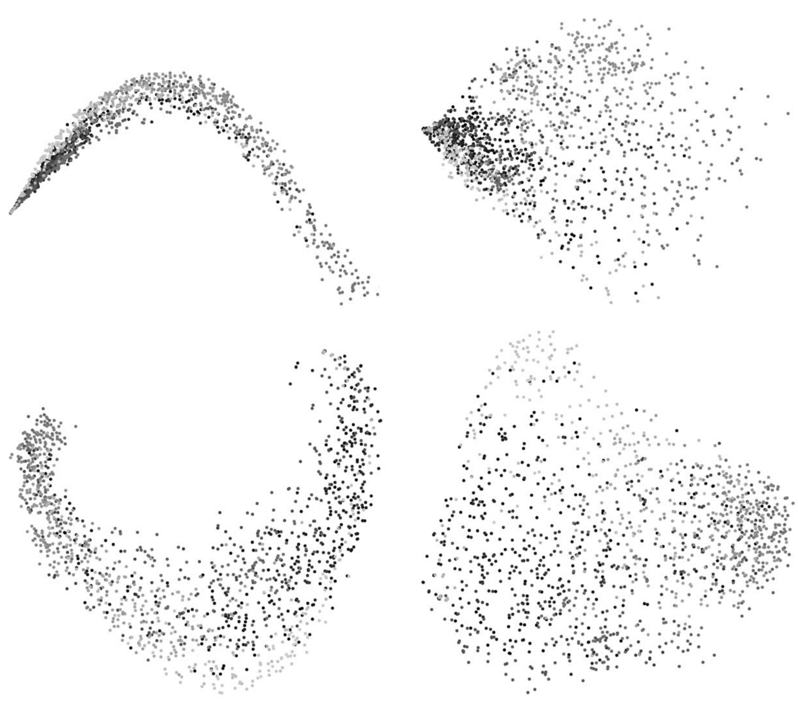

Fig. 9. Multidimensional scaling of the distances obtained with our metrics from the third experiment. Top to bottom, left to right: d_1, d_2, d_3, and d_4.

TABLE 3
Davies-Bouldin Index of Clusters Obtained with Normalized Cut in the Third Experiment

	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 4
Davies-Bouldin Index of Clusters Obtained with Dominant Sets in the Third Experiment

	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 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 algorithm. As was the case with the previous experiments, all five metrics produce comparable results.

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 have provided a polynomial-time algorithm to calculate them. We have proven that these measures satisfy the metric properties and have experimentally validated their usefulness by comparing them with an approximation of 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, however, their computation is orders of magnitude faster. As for the choice among our four metrics, the experimental results, while not pointing to a clear winner, show that the metrics that balance the similarity against the average size (i.e., d_2 and d_4) distribute the structures better in the embedding space, thereby confirming our intuition.

In addition to computational complexity issues, the choice between the proposed metrics and edit distance depends, as previously noted, on the nature of the noise in the problem at hand. As far as our shape categorization application is concerned, given the intrinsic properties of shock trees, intraclass variation are expected to lead to attribute and peripheral noise only, thereby making our metrics a natural choice over edit-distance. Indeed, our experimental results do confirm our expectations. As a side note, we mention that Bunke [5] has recently shown that, under certain assumptions on the edit-costs, the graph edit-distance problem can be reduced to the maximum common subgraph problem. It would be interesting to see whether analogous results hold for subtree isomorphism as well, which is different to the maximum common subgraph problem due to the connectivity constraints.

Acknowledgments

The authors would like to thank the reviewers for their helpful comments which allowed us to improve the presentation of the paper.

References

[1] R.K. Ahuja, T.L. Magnanti, and J.B. Orl, *Network Flows.* Upper Saddle River, N.J.: Prentice-Hall, 1993.
[2] H.G. Barrow and R.M. Burstall, "Subgraph Isomorphism, Matching Relational Structures and Maximal Cliques," *Information Processing Letters,* vol. 4, no. 4, pp. 83-84, 1976.
[3] H. Blum and R.N. Nagel, "Shape Description Using Weighted Symmetric Axis Features," *Pattern Recognition,* vol. 10, no. 3, pp. 167-180, 1978.
[4] K.L. Boyer and A.C. Kak, "Structural Stereopsis for 3-D Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 10, no. 2, pp. 144-166, Feb. 1988.
[5] H. Bunke, "On a Relation between Graph Edit Distance and Maximum Common Subgraph," *Pattern Recognition Letters,* vol. 18, no. 8, pp. 689-694, 1997.
[6] H. Bunke and K. Shearer, "A Graph Distance Metric Based on the Maximal Common Subgraph," *Pattern Recognition Letters,* vol. 19, nos. 3-4, pp. 255-259, 1998.
[7] W.J. Christmas, J. Kittler, and M. Petrou, "Structural Matching in Computer Vision Using Probabilistic Relaxation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 8, pp. 749-764, Aug. 1995.
[8] S.J. Dickinson, A.P. Pentland, and A. Rosenfeld, "3-D Shape Recovery Using Distributed Aspect Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 2, pp. 174-198, Feb. 1992.
[9] M.A. Eshera and K.-S. Fu, "An Image Understanding System Using Attributed Symbolic Representation and Inexact Graph-Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 8, no. 5, pp. 604-618, May 1986.
[10] M.L. Fernández and G. Valiente, "A Graph Distance Metric Combining Maximum Common Subgraph and Minimum Common Supergraph," *Pattern Recognition Letters,* vol. 22, nos. 6-7, pp. 753-758, 2001.
[11] D. Hidović and M. Pelillo, "Metrics for Attributed Graphs Based on the Maximal Similarity Common Subgraph," *Int'l J. Pattern Recognition Artificial Intelligence,* vol. 18, no. 3, pp. 299-313, 2004.
[12] S. Ioffe and D.A. Forsyth, "Human Tracking with Mixtures of Trees," *Proc. Int'l Conf. Computer Vision,* vol. 1, pp. 690-695, 2001.
[13] A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data.* Englewood Cliffs, N.J.: Prentice Hall, 1988.
[14] B.B. Kimia, A.R. Tannenbaum, and S.W. Zucker, "Shapes, Shocks, and Deformations I: The Components of Shape and the Reaction-Diffusion Space," *Int'l J. Computer Vision,* vol. 15, no. 3, pp. 189-224, 1995.
[15] D.W. Matula, "Subtree Isomorphism in $O(n^{5/2})$," *Annals of Discrete Math.,* vol. 2, no. 1, pp. 91-106, 1978.
[16] B. Moayer and K.-S. Fu, "A Tree System Approach for Fingerprint Pattern Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 8, no. 3, pp. 376-387, 1986.
[17] C.H. Papadimitriou and K.J. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity.* Englewood Cliffs, N.J.: Prentice-Hall, 1982.
[18] M. Pavan and M. Pelillo, "A New Graph-Theoretic Approach to Clustering and Segmentation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 145-152, 2003.
[19] M. Pelillo, K. Sidiqi, and S.W. Zucker, "Matching Hierarchical Structures Using Association Graphs," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 11, pp. 1105-1120, Nov. 1999.

[20] M. Peura, "Attribute Trees in Image Analysis: Heuristic Matching and Learning Techniques," *Proc. Int'l Conf. Image Analysis Processing,* pp. 1160-1165, 1999.
[21] H. Samet, "Distance Transform for Images Represented by Quadtrees," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 4, no. 3, pp. 298-303, 1982.
[22] T.B. Sebastian, P.N. Klein, and B.B. Kimia, "Recognition of Shapes by Editing Their Shock Graphs," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 5, pp. 550-571, May 2004.
[23] L.G. Shapiro and R.M. Haralick, "Relational Models for Scene Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 4, no. 6, pp. 595-602, 1982.
[24] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 888-905, Aug. 2000.
[25] K. Siddiqi, A. Shokoufandeh, S.J. Dickinson, and S.W. Zucker, "Shock Graphs and Shape Matching," *Int'l 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," *Proc. IEEE Conf. Computer Vision and 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 Analysis and Machine Intelligence,* vol. 16, no. 12, pp. 1233-1238, Dec. 1994.
[28] A. Torsello and E.R. Hancock, "A Skeletal Measure of 2D Shape Similarity," *Proc. Int'l Workshop Visual Form,* pp. 260-271, 2001.
[29] A. Torsello and E.R. Hancock, "Efficiently Computing Weighted Tree Edit-Distance Using Relaxation Labeling," *Proc. Int'l Workshop Energy Minimization Methods in Computer Vision and Pattern Recognition,* pp. 438-453, 2001.
[30] A. Torsello, D. Hidović, and M. Pelillo, "A New Polynomial-Time Metric on Attributed Trees," *Proc. Eighth European Conf. Computer Vision,* vol. 4, pp. 414-427, 2004.
[31] A. Torsello, D. Hidović, and M. Pelillo, "Four Metrics for Efficiently Comparing Attributed Trees," *Proc. 17th Int'l Conf. Pattern Recognition,* vol. 2, pp. 467-470, 2004.
[32] W.H. Tsai and K.-S. Fu, "Error-Correcting Isomorphism of Attributed Relational Graphs for Pattern Analysis," *IEEE Trans Systems, Man, and Cybernetics,* vol. 9, no. 12, pp. 757-768, 1979.
[33] G. Valiente, "An Efficient Bottom-Up Distance between Trees," *Proc. Eighth Int'l Symp. String Processing Information Retrieval,* pp. 212-219, 2001.
[34] G. Valiente, *Algorithms on Trees and Graphs.* Berlin: Springer, 2002.
[35] W.D. Wallis, P. Shoubridge, M. Kraetz, and D. Ray, "Graph Distances Using Graph Union," *Pattern Recognition Letters,* vol. 22, no. 6-7, pp. 701-704, 2001.
[36] J.T.-L. Wang and K. Zhang, "Finding Similar Consensus between Trees: An Algorithm and a Distance Hierarchy," *Pattern Recognition,* vol. 34, no. 1, pp. 127-137, 2001.
[37] J.T.L. Wang, K. Zhang, G. Chang, and D. Shasha, "Finding Approximate Patterns in Undirected Acyclic Graphs," *Pattern Recognition,* vol. 35, pp. 473-483, 2002.
[38] R.C. Wilson and E.R. Hancock, "Structural Matching by Discrete Relaxation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 6, pp. 634-648, June 1997.
[39] A.K.C. Wong and M. You, "Entropy and Distance of Random Graphs with Application to Structural Pattern Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 7, no. 5, pp. 599-609, May 1985.
[40] K. Zhang, "A Constrained Edit-Distance between Unordered Labeled Trees," *Algorithmica,* vol. 15, no. 3, pp. 205-222, 1996.
[41] K. Zhang and D. Shasha, "Simple Fast Algorithms for the Editing Distance between Trees and Related Problems," *SIAM J. Computing,* vol. 18, no. 6, pp. 1245-1262, 1989.
[42] K. Zhang, R. Statman, and D. Shasha, "On the Editing Distance between Unordered Labeled Trees," *Informtion Processing Letters,* vol. 42, no. 3, pp. 133-139, 1992.
[43] K. Zhang, J.T.L. Wang, and D. Shasha, "On the Editing Distance between Undirected Acyclic Graphs," *Int'l J. Foundations of Computer Science,* vol. 7, no. 1, pp. 43-57, 1996.
[44] S.C. Zhu and A.L. Yuille, "FORMS: A Flexible Object Recognition and Modelling System," *Int'l J. Computer Vision,* vol. 20, no. 3, pp. 187-212, 1996.

Andrea Torsello received the Laurea degree with honors in computer science from Ca' Foscari University of Venice, Italy, in 1997. In 2004, he received the PhD degree in computer science from the University of York, United Kingdom. Currently, he has a postdoctoral position at Ca' Foscari University of Venice, Italy. His research interests are in the area of computer vision and pattern recognition, in particular, the application of stochastic and structural approaches to shape analysis. He is currently coediting a special issue of *Pattern Recognition* on similarity-based pattern recognition.

Džena Hidović-Rowe received the Laurea degree with honors in computer science from Ca' Foscari University of Venice, Italy, in 2002. Currently, she is pursuing the PhD degree in computer science at the University of Birmingham, United Kingdom. Her research interests include tree and graph metrics for image analysis and pattern recognition, as well as medical image understanding and physics-based models of tissue optics.

Marcello Pelillo received the Laurea degree with honors in computer science from the University of Bari, Italy, in 1989. From 1988 to 1989, he was at the IBM Scientific Center in Rome, where he was involved in studies on natural language and speech processing. In 1991, he joined the faculty of the University of Bari, Italy, as an assistant professor of computer science. Since 1995, he has been with the University of Venice, Italy, where he is currently an associate professor of computer science and the chair of the board of study of the computer science school. He held visiting research positions at Yale University, University College London, McGill University, the University of Vienna, and York University (England). His research interests are in the areas of computer vision, pattern recognition, and neural computation, where he has published more than 100 papers in refereed journals, handbooks, and conference proceedings. He has organized a number of scientific events, including the Neural Information Processing Systems (NIPS) 1999 Workshop on "Complexity and Neural Computation: The Average and the Worst Case." In 1997, he established a new series of international workshops devoted to energy minimization methods in computer vision and pattern recognition (EMMCVPR), which has now reached the fifth edition. He also coedited two journal special issues on this theme, one in 2000 for *Pattern Recognition* and the other in 2003/2004 for the *IEEE Transactions on Pattern Analysis and Machine Intelligence*. In 2001, he was a guest coeditor of a special issue of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* devoted to graph algorithms in computer vision. He is currently coediting a special issue of *Pattern Recognition* on similarity-based pattern recognition. He has been on the program committees of several international conferences and workshops and serves on the editorial board for the journal, *Pattern Recognition*. Professor Pelillo is a senior member of the IEEE and a member of the IAPR and the Pattern Recognition Society.