Approximation Algorithms for Classification Problems with Pairwise Relationships: Metric Labeling and Markov Random Fields

JON KLEINBERG AND ÉVA TARDOS

Cornell University, Ithaca, New York

Abstract. In a traditional classification problem, we wish to assign one of *k labels* (or classes) to each of *n objects*, in a way that is consistent with some observed data that we have about the problem. An active line of research in this area is concerned with classification when one has information about *pairwise relationships* among the objects to be classified; this issue is one of the principal motivations for the framework of Markov random fields, and it arises in areas such as image processing, biometry, and document analysis. In its most basic form, this style of analysis seeks to find a classification that optimizes a combinatorial function consisting of *assignment costs*—based on the individual choice of label we make for each object—and *separation costs*—based on the *pair* of choices we make for two "related" objects.

We formulate a general classification problem of this type, the *metric labeling problem*; we show that it contains as special cases a number of standard classification frameworks, including several arising from the theory of Markov random fields. From the perspective of combinatorial optimization, our problem can be viewed as a substantial generalization of the multiway cut problem, and equivalent to a type of *uncapacitated quadratic assignment problem*.

We provide the first nontrivial polynomial-time approximation algorithms for a general family of classification problems of this type. Our main result is an $O(\log k \log \log k)$-approximation algorithm for the metric labeling problem, with respect to an arbitrary metric on a set of k labels, and an arbitrary weighted graph of relationships on a set of objects. For the special case in which the labels are endowed with the *uniform metric*—all distances are the same—our methods provide a 2-approximation algorithm.

Categories and Subject Descriptors: F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems

General Terms: Algorithms, Theory

A preliminary version of this article appeared in the *Proceedings of the 40th Annual IEEE Symposium on the Foundations of Computer Science* (Oct.). IEEE Computer Society Press, Los Alamitos, Calif., 1999.

The research of J. Kleinberg was supported in part by a David and Lucile Packard Foundation Fellowship, an ONR Young Investigator Award, and NSF Faculty Early Career Development Award CCR-9701399.

The research of É. Tardos was supported in part by NSF grant CCR-9700163 and ONR grant N00014-98-1-0589.

Authors' address: Department of Computer Science, Cornell University, 4126 Upson Hall, Ithaca, NY 14853, e-mail: {kleinber;eva}@cs.cornell.edu.

Additional Key Words and Phrases: Approximation algorithms, classification, Markov random fields, metric labeling

1. *Introduction*

1.1. Classification Problems with Pairwise Relationships.

A fundamental issue in statistics, pattern recognition, and machine learning is that of *classification*. At a very high level, we can view a classification problem as consisting of a set P of *objects* to be classified, and a set L of *labels* (the classes); the goal is to assign a label to each object in a way that is consistent with some observed data that we have about the problem [Breiman et al. 1984; Dietterich 1997].

The "observed data" in such a problem can take a wide variety of different forms; our interest here is in classification problems for which these data involve pairwise relationships among the objects to be classified. Such problems have a rich history in statistics, where they arise naturally from formalisms such as the theory of *Markov random fields* [Besag 1974; Chellappa and Jain 1993; Kinderman and Snell 1980; Li 1995; Woods 1972]; they have formed some of the core analytical frameworks in areas including image processing [Besag 1986; Cohen 1986; Dubes and Jain 1989; Geman and Geman 1984] and biometric analysis [Besag 1974] and they have found applications in many other fields, including language modeling [Della Pietra et al. 1997] and the categorization of hypertext documents [Chakrabarti et al. 1998].

To motivate the formulation, we consider, as an illustrative example, the well-studied problem of "restoring" an image that has been degraded by noise [Besag 1986; Geman and Geman 1984]. We are given a large grid of pixels; each pixel has a "true" intensity that we are trying to determine, and an "observed" intensity that is the result of corruption by noise. We would like to find the best way to label each pixel with a (true) intensity value, based on the observed intensities. Our determination of the "best" labeling is based on the trade-off between two competing influences: We would like to give each pixel an intensity close to what we have observed; and—since real images are mainly smooth, with occasional boundary regions of sharp discontinuity—we would like spatially neighboring pixels to receive similar intensity values. Below, we give a self-contained formulation of an objective function that naturally captures the trade-off between these issues; we will mention several models that our formulation generalizes, and survey (in Section 2) the connections between our model and the framework of Markov random fields.

Formulations of this type have been applied to a wide range of other image processing tasks as well; for example, in image segmentation and interpretation, the objects are pixels and the labels are logical "regions" into which we want to partition the image (see, e.g., Cohen [1986] and Dubes and Jain [1989]). Labeling problems with pairwise relationships also arise naturally in the other applications discussed above. For example, in biometry, one interprets observations gathered over many points in a geographical region under the assumption that the "true" properties of nearby points will be similar [Besag 1974]; in hypertext categorization, it is possible to enhance the power of heuristics for classifying individual documents using the assumption that pairs of documents with a hyperlink between them are more likely to be about similar topics [Chakrabarti et al. 1998].

1.2. THE METRIC LABELING PROBLEM. One can make these issues precise in the following way. Consider a set P of n objects that we wish to classify, and a set L of k possible labels. A *labeling* of P over L is simply a function $f: P \to L$; we choose a label for each object. The quality of our labeling is based on the contribution of two sets of terms.

—For each object $p \in P$, we have an estimate of its likelihood of having each label $a \in L$. The determination of these likelihoods is assumed to come from some heuristic preprocessing of the data; for our purposes, it results in a nonnegative *cost* $c(p, a)$ for assigning label a to object p. To take a simple example using pixels and intensities, suppose the observed color of pixel (i.e., object) p is `white`; then the cost $c(p, \texttt{black})$ should be high while $c(p, \texttt{white})$ should be low.

—We also have information about pairwise relationships among the objects; if p and q are deemed to be *related*, then we would like them to have the same (or "similar") labels. Thus, we have a graph G over the vertex set P, with edge set E indicating the objects that are related; each edge $e = (p, q)$ will have a nonnegative weight w_e, indicating the strength of the relation. In our example with pixels, the nodes of the graph G are the pixels; in the simplest model, pixels p and q could be connected by an edge $e = (p, q)$ only if they are neighboring, and all edges could have $w_e = 1$.

Moreover, as in our example with pixels and intensities, certain pairs of labels are more similar than others; so we impose a distance $d(\cdot, \cdot)$ on the set L of labels, with larger distance values indicating less similarity. (So $d(\texttt{white}, \texttt{black})$ would presumably be large in our example.) If we assign label a to object p and label b to object q—and $e = (p, q)$ is an edge of G—then we pay a cost of $w_e d(a, b)$.

Thus, the *total cost* of a discrete labeling f is given by

$$Q(f) = \sum_{p \in P} c(p, f(p)) + \sum_{e=(p,q) \in E} w_e d(f(p), f(q)).$$

The *labeling problem* asks for a discrete labeling of minimum total cost. In this article, we assume that the distance $d(\cdot, \cdot)$ is a metric on the set L of labels, that is, it satisfies the triangle inequality, and will refer to the resulting special case as the *metric labeling problem*. Boykov et al. [1999; 2001] independently also introduced and studied this special case.

Despite the fact that such labeling problems are widely used as classification frameworks, there have been no previous algorithms for such labeling problems that provided nontrivial approximation guarantees. Traditionally, standard local search methods have been applied to labeling problems of this general form. Recently, combinatorial flow-based algorithms have been playing an increasingly large role in computer vision (see, e.g., Boykov et al. [1998], Ishikawa and Geiger [1998], and Roy and Cox [1998]). For metric labeling problems arising in vision, Boykov, et al. [1998, 1999, 2001] have developed flow-based local search heuristics with good performance in practice.

In this article, we provide a polynomial-time approximation algorithm for the general metric labeling problem. Our main result is an $O(\log |L| \log \log |L|)$-approximation algorithm for this problem, with respect to an arbitrary weighted graph on the set of objects, and an arbitrary metric on the set of labels. Note that the

approximation guarantee is independent of $|P|$, the size of the object set. For the special case of the *uniform metric*—when all distances are equal to 1—our methods provide a 2-approximation algorithm.

1.3. RELATED PROBLEMS. The reader may notice a connection between the metric labeling problem and the well-studied *multiway cut problem* [Calinescu et al. 1998; Cunningham and Tang 1999; Dahlhaus et al. 1994; Erdös and Székely 1992; Karger et al. 1999], in which we are given a weighted graph with k terminals, and we must find a partition of the graph into k sets so that each terminal is in a separate set, and the total weight of the edges cut is as small as possible. We can view the parts of the partition as representing the k labels of a classification problem; however, the objective function in multiway cut is too simple to capture the issues raised in the applications above. First, we do not wish simply to measure the total edge weight of neighboring objects that receive different labels; the penalty for labeling two objects differently should also depend on the identities of the labels they receive. For example, in image restoration, there should be a smaller penalty for assigning neighboring pixels similar (but unequal) intensities than for assigning them very different intensities. Additionally, we wish to use assignment costs for assigning objects to labels. The multiway cut problem is thus a special case of our problem in which the terminals are "preclassified"—they must receive a certain label—while all other nodes are "unclassified" in the sense that they have no preference among the different labels, or parts of the partition. Using the assignment costs allows us to implicitly define "soft" initial assignments of labels to objects, that is, to make certain assignments expensive and others cheap. Metric labeling is thus a substantial generalization of the multiway cut problem, designed to incorporate these fundamental issues.

In fact, the metric labeling problem can perhaps better be viewed as the *uncapacitated quadratic assignment* problem. In the quadratic assignment problem, a heavily studied topic in discrete optimization, one must find a matching between a set of n given *activities* to n *locations* in a metric space so as to minimize a sum of *assignment costs* for activities, and *flow costs* for activities that "interact." We discuss this problem in more detail in the subsequent section; for now, we simply note that our problem can be obtained from the quadratic assignment by dropping the requirement that at most one activity can be sited at a given location. The activities then correspond to objects, and the locations to labels, in the metric labeling problem.

Finally, how does our formulation of the metric labeling problem relate to the applications discussed above? In the next section, we show how the optimum of a metric labeling problem, as we have described it here, is equivalent to the optimal configuration for a certain general class of Markov random fields. Metric labeling contains, as a special case, the *Generalized Potts model* used by Boykov, Veksler, and Zabih [Boykov et al. 1998; Potts 1952] to obtain strong results for a range of image analysis tasks; it also contains, as special cases, several other basic models proposed in the areas of image processing and biometry [Besag 1974; Cohen 1986; Dubes and Jain 1989; Greig et al. 1989].

1.4. THE PRESENT WORK: SUMMARY OF RESULTS. Our main result is an $O(\log k \log\log k)$-approximation algorithm for the metric labeling problem, with respect to an arbitrary weighted graph on the set of objects, and an arbitrary metric on the set of labels. Recall that k denotes the size of the label set L, and note that our approximation guarantee is independent of n, the size of the set of objects.

As the first main step in obtaining our approximation algorithm for the general problem, we develop a 2-approximation for a special case of metric labeling that we refer to as the *uniform labeling* problem. This is the case in which the metric $d(\cdot, \cdot)$ on the set L of labels is *uniform*: for labels a and b, $d(a, b)$ is equal to 0, if $a = b$, and equal to 1, otherwise. The costs $\{c(p, a)\}$ and the edge weights $\{w_e\}$ remain arbitrary. The uniform labeling problem is of interest in its own right, as it incorporates one of the key issues discussed above: "soft" initial assignments of labels to objects through costs. This special case has been previously studied by Boykov et al. [1998], who termed it the *Generalized Potts models*.

For the special case of uniform labeling, we have learned that Sanjeev Arora has independently obtained a constant-factor approximation algorithm (S. Arora personal communication); and Julia Chuzhoy (personal communication) has been able to improve on our approximation ratio of 2 for the case of $k = 3$ and $k = 4$ labels to $4/3$ and $11/6$ respectively. Concurrently with the present work, Boykov et al. [1999] have shown that for one of their flow-based heuristics for the uniform labeling problem, the objective function value at any local optimum is within a factor of 2 of the global optimum.

The metric labeling problem generalizes the multiway cut problem, and hence it is MAXSNP-hard [Dahlhaus et al. 1994]. In image processing applications [Besag 1986; Cohen 1986; Dubes and Jain 1989; Geman and Geman 1984], the underlying graph G is typically an $n \times n$ grid. We can show the metric labeling problem to be NP-complete even when the underlying graph G is an $n \times n$ grid. (See the proof in Boykov et al. [2001].)

1.5. THE PRESENT WORK: LABELING, CUTS, AND LP RELAXATIONS. We have already indicated connections between metric labeling and the multiway cut problem at an informal level. Indeed, it is easy to see that multiway cut is already a special case of the uniform labeling problem: given an instance of the multiway cut problem with a graph $G = (V, E)$, edge weights w_e for $e \in E$, and *terminals* $t_1, \ldots, t_k \in V$ that we wish to disconnect by deleting edges, we can define the following instance of the uniform labeling problem. The objects will be the nodes V, with the graph of relationships equal to G; the label set is $\{1, 2, \ldots, k\}$. All label assignment costs are set to 0, except that we set $c(t_i, j)$ equal to ∞ (or a very large positive quantity) if $i \neq j$. This forces t_i to receive label i, and we thus assign the remaining labels so as to minimize the total weight of edges with unequal labels at their ends.

The connection between optimization over labeling and optimization over cuts in a graph has been known for some time in work on Markov random fields. Greig et al. [1989] and Besag [1986] showed that the case of $k = 2$ labels is polynomially solvable as a two-terminal minimum cut problem. Boykov et al. [1998] developed the connection between uniform labelings, with $k > 2$ labels, and multiway cuts in a graph, showing a direct reduction from labelings to multiway cuts. However, their reduction required the use of edges of enormously large weight, and so it is not approximation-preserving.

In fact, it is not difficult to see that, once we allow highly variable costs $c(\cdot, \cdot)$ (or infinite costs), direct extension of either the *Isolation Heuristic* of Dahlhaus et al. [1994], or the algorithm of Calinescu et al. [1998] no longer provides a reasonable approximation guarantee. These methods do not deal well with objects that *cannot* receive certain labels, and they can produce results that are arbitrarily far from optimal.

For the case of uniform labeling, we use a linear programming relaxation naturally adapted from the one developed for the multiway cut problem by Calinescu et al. [1998], and also studied in Cunningham and Tang [1999] and Karger et al. [1999]. Although the LP formulation in this special case directly generalizes that of Calinescu et al. [1998], the presence of variable label assignment costs forces us to use a different rounding technique; and indeed, the integrality gap of the formulation now converges to 2, matching our upper bound. (In Calinescu et al. [1998], the analogous formulation has an upper bound of at most 1.5.)

The main difficulties for us arise in moving up to the general metric labeling problem. One problem we face here is that we no longer know of any "natural" linear programming relaxation for the problem. Consequently, we focus first on the case in which the metric on L corresponds to a *hierarchically well-separated tree* [approximations of metric spaces and its algorithmic applications 1996]; here, we formulate a linear programming relaxation and develop a rounding algorithm whose performance guarantee is a constant depending inversely on the "separation parameter" of the underlying tree. Then, using Bartal's [1996, 1998] result that any finite metric space can be probabilistically approximated by hierarchically well-separated trees [approximations of metric spaces and its algorithmic applications 1996; Bartal 1998], we are able to obtain an algorithm for general metric spaces.

We note another way, not considered here, in which the connection between labelings and cuts has been exploited—for the identification of polynomially solvable special cases of the metric labeling problem. This issue has been investigated in recent work of Karzanov [1998]. Boykov et al. [1998] and Ishikawa and Geiger [1998] show a direct reduction to two-terminal minimum cuts—and hence an efficient algorithm—in the special case of metric labeling when the label set is $L = \{0, 1, \ldots, k\}$, and the distance between two labels i and j is $d(i, j) = |i - j|$.

2. *Connections to Other Models*

In this section, we summarize the relationship between the metric labeling problem and various well-studied problems in statistics and combinatorial optimization: in particular, we consider Markov random fields, and the quadratic assignment problem. We begin developing our algorithms in Section 3, and the reader has the option of proceeding directly to that section without loss of continuity.

2.1. Markov Random Fields.

For a given set of objects P and labels L, the theory of *random fields* provides a systematic means for converting the set of all possible labelings of P over L into a probability space. Thus, for every labeling f, the random field assigns it a probability $\Pr[f]$. In parts of the following discussion, one needs the assumption that $\Pr[f] > 0$ for all labelings f, a standard requirement in work on random fields [Besag 1974].

As before, we have a graph $G = (P, E)$ on the set of objects. The random field is *Markovian* if the conditional probability of the label assignment at object p depends only on the label assignments at the neighbors of p in G. That is, a Markov random field (MRF) satisfies

$$\Pr[f(p)|\{f(q) : q \neq p\}] = \Pr[f(p)|\{f(q) : (p, q) \in E\}]. \tag{1}$$

If we consider our initial example of image restoration, we see that the "local" nature of the dependencies in a Markov random field captures our intuitive notion

that the label assigned to a given pixel depends only on the labels of its neighboring pixels.

A theorem of Hammersley and Clifford (see Besag [1974]) allows us to write the distribution of a Markov random field in essentially a closed form, as follows. Let $\mathcal{C}$ be the set of all cliques in G; for a clique $C \in \mathcal{C}$, we say that a *clique potential* associated with C is a function Γ_C that maps labelings of objects in C to real numbers. For a labeling f of all objects and a clique C, we use the notation $f|C$ to indicate the restriction of f to the objects in C. Then, the Hammersley–Clifford theorem states that for a given Markov random field, there exists a set of clique potentials $\{\Gamma_C : C \in \mathcal{C}\}$ so that, for an arbitrary labeling f, $\Pr[f] = Z^{-1} \exp(-\sum_{C \in \mathcal{C}} \Gamma_C(f|C))$. Here Z is a normalizing constant. We can take the negative logarithm of this probability, obtaining a value that we call the *cost*, $\sum_{C \in \mathcal{C}} \Gamma_C(f|C)) + \ln Z$.

In the context of classification, suppose that our data consists of likely labels $f'(p)$ for each object $p \in P$. Assume that these data are noisy, and we wish to decide on the most probable labeling f of P. A standard approach [Li 1995] is the following. We assume that the labeling f is drawn from an underlying MRF that is given. Further, we assume that the labeling f' is obtained from f by the introduction of independent random noise at each object $p \in P$. We wish to decide on the most probable labeling f of P, given the labeling f' of P and the underlying MRF model. The standard objective [Li 1995] is to find the f that maximizes the *a posteriori* probability $\Pr[f \mid f']$.

By Bayes' Law, this probability $\Pr[f \mid f']$ is

$$\Pr[f \mid f'] = \frac{\Pr[f' \mid f] \cdot \Pr[f]}{\Pr[f']}.$$

Using the assumption that f' is obtained from f by the introduction of independent random noise at each object, the term $\Pr[f' \mid f]$ factors into $\prod_p \Pr[f'(p) \mid f(p)]$. The term $\Pr[f]$ is given by the Hammersley–Clifford expansion in terms of cliques. The denominator $\Pr[f']$ is the same for all labelings, as is the constant Z in the Hammersley–Clifford expansion. Taking the negative logarithm of the remaining terms, we get that the most probable labeling is the one that minimizes the following *energy function*,

$$E(f) = \sum_p (-\ln \Pr[f'(p) \mid f(p)]) + \sum_{C \in \mathcal{C}} \Gamma_C(f|C)). \tag{2}$$

We can now see how the metric labeling problem arises from this formalism. In standard applications (see Li [1995]), one very frequently works with a restricted type of MRF satisfying two properties:

(i) *Pairwise interactions.* The MRF is defined purely by pairwise interactions among the objects; so Γ_C is a nonzero function only for cliques of size 2 (i.e., edges of G). We will denote these functions by Γ_e, for $e \in G$; note that each is a function mapping $L \times L$ to $\mathbf{R}$.

(ii) *Homogeneity.* The MRF is spatially homogeneous, so each function Γ_e can be written as $w_e\Gamma$, where w_e is a nonnegative edge weight, and the function Γ mapping $L \times L$ to $\mathbf{R}^+$ is the same for all edges.

We define a *metric Markov random field* to be an MRF that satisfies the above two properties, and for which the function Γ is a metric on the set of labels. Now, for

an arbitrary metric MRF, we can define $c(p, a) = -\ln \Pr[f'(p)|a]$ and rewrite the problem of minimizing the energy function defined in Eq. (2) as follows:

$$\min_f E(f) = \min_f \left[\sum_p c(p, f(p)) + \sum_{e=(p,q)} w_e \Gamma(f(p), f(q)) \right], \tag{3}$$

where Γ is a metric. This is precisely the statement of the metric labeling problem. Note, however, that our transformation is not approximation preserving, as it involved ignoring the constant terms $\Pr[f']$ and Z, and taking a logarithm.

Given that pairwise, homogeneous MRF's are ubiquitous in the application of MRF's to classification problems, we feel that our formulation of metric MRF's is quite general. It contains the Potts model [Ferrari et al. 1995; Potts 1952] as a special case, by setting all $w_e = 1$ and Γ to be the *uniform metric*, and the Generalized Potts model of Boykov et al. [1998] by setting Γ to be the uniform metric and allowing the weights w_e to be arbitrary. Metric MRF's were independently introduced and studied also by Boykov et al. [2001].

2.2. QUADRATIC ASSIGNMENT PROBLEM. The quadratic assignment problem is one of the most well-studied facility location problems in operations research; see Pardalos and Wolkowicz [1994] for a comprehensive recent survey. We are given *n activities* and *n locations* in a metric space; we would like to site each activity at a distinct location in a way that minimizes an objective function consisting of the following types of terms:

—If we site activity i at location ℓ, we incur an operating cost $c(i, \ell)$.

—Moreover, the activities interact, and we wish to keep activities that interact closely near to one another.

Thus, if $w_{ij} \geq 0$ measures the interaction between activities i and j, and $d(\ell, \ell')$ measures the distance between locations ℓ and ℓ', then we pay a cost of $w_{ij}d(\ell, \ell')$ for siting activity i at ℓ and j at ℓ'.

The traditional quadratic assignment is based on minimizing $\sum_i c(i, f(i)) + \sum_{i,j} w_{ij} d(f(i), f(j))$ over all *bijections* f. The problem is known to be hard in the sense that no polynomial-time algorithm can approximate the minimum to any polynomial factor unless P = NP [Queyranne 1986].

The metric labeling problem seeks to minimize precisely the same objective function over all *functions* f (not only over bijections), for the locations can be viewed as the labels, and the set of interaction values $\{w_{ij}\}$ can be viewed as a weighted graph on the set of activities. Thus, metric labeling is equivalent to the *uncapacitated quadratic assignment problem*, in which multiple activities may be sited at the same location. In view of the discussion above, this uncapacitated version—like the traditional version—is NP-complete, due to the fact that activities have different costs for operating at different locations. However, our results show that, in terms of approximability, the metric labeling problem problem is easier than the quadratic assignment problem (unless P = NP).

3. *The Uniform Labeling Problem*

Before considering the general metric labeling problem, we develop an approximation algorithm for the uniform labeling problem. Thus, we consider the case of uniform distances, in which each pair of different labels is at distance 1.

The uniform labeling problem can be expressed as the following integer program, analogous to the integer program used by Calinescu et al. [1998] for the multiway cut problem. We define nonnegative variables x_{pa} for each object p and label a, requiring that $\sum_{a \in L} x_{pa} = 1$; that is, for integer valued x we use $x_{pa} = 1$ to denote that $f(p) = a$—object p is assigned label a. The label cost of an object p is then expressed as $\sum_{a \in L} c(p, a)\, x_{pa}$. Consider two objects p and q. The distance between the assigned labels can be expressed as $d(f(p),\ f(q)) = \frac{1}{2} \sum_{a \in L} |x_{pa} - x_{qa}|$. Hence, the separation cost for an edge $e = (p, q)$ is expressed as $\frac{1}{2} w_e \sum_{a \in L} |x_{pa} - x_{qa}|$. We can write this as an integer linear program by introducing variables z_e for an edge $e = (p, q)$ to express the distance between the labels assigned to objects p and q, and we use z_{ea} to express the absolute value $|x_{qa} - x_{pa}|$ for $a \in L$.

$$
\begin{array}{llll}
\text{(UIP)} & \text{Min} & \sum_{e \in E} w_e z_e + \sum_{p \in P, a \in L} c(p, a)\, x_{pa} & \\
 & \text{subject to} & \sum_{a \in L} x_{pa} = 1 & p \in P \\
 & & z_e = \frac{1}{2} \sum_{a \in L} z_{ea} & e \in E \\
 & & z_{ea} \geq x_{pa} - x_{qa} & e = (p, q), a \in L \\
 & & z_{ea} \geq x_{qa} - x_{pa} & e = (p, q), a \in L \\
 & & x_{pa} \in \{0, 1\} & p \in P, a \in L.
\end{array}
$$

We create the linear programming relaxation of (UIP) by replacing the integrality constraints with $x_{pa} \geq 0$; we call this (ULP). We say that a set of nonnegative variables $\{x_{pa} : p \in P, a \in L\}$ is a *fractional labeling* if $\sum_{a \in L} x_{pa} = 1$ for all objects $p \in P$. The *assignment cost* of the fractional labeling x is $\sum_{p \in P, a \in L} c(p, a)\, x_{pa}$ and the *separation cost* is $\sum_{e=(p,q) \in E} [\frac{1}{2} w_e \sum_{a \in L} |x_{pa} - x_{qa}|]$.

We give a randomized method for rounding the fractional solution to (ULP) to integers, losing at most a factor of 2 in the objective function. The idea is to use the x_{pa} values as probabilities that object p gets assigned to label a. However, the simplest randomized rounding scheme—assigning each object p independently to a label with probabilities equal to its x_{pa} values—works terribly. For if $e = (p, q)$ and p and q have the same set of values $\{x_{pa}\}$, $\{x_{qa}\}$, then $z_e = 0$ in (ULP), and so we must assign p and q the same label if we are to use a term-by-term analysis for the approximation guarantee. Calinescu, et al. [1998] addressed the analogous problem in the setting of the multiway cut problem through a more careful rounding scheme that deals well with separation costs. However, this rounding scheme cannot handle assignment costs, and in our more general setting, can produce results that are arbitrarily far from optimal.

Our randomized rounding scheme can be viewed as a variant of the approach of Calinescu et al. [1998]. We assign labels to vertices in *phases*; initially each vertex has no label assignment. In a single phase, we pick a label a uniformly at random, and a real number α uniformly in $[0, 1]$. For each object p that currently has no label assignment, we give it the label a if $\alpha \leq x_{pa}$. We iterate these phases until each vertex has a label assignment.

The following two lemmas state simple properties of this randomized process:

LEMMA 3.1. *Consider a particular phase, and an object p that has not yet been assigned a label at the start of this phase. The probability that p is assigned label a in this phase is precisely x_{pa}/k, and the probability that object p is assigned any label in this phase is precisely $1/k$. Further, over all phases, the probability that an object p is assigned label a is precisely x_{pa}.*

PROOF. In each phase, while object p has no label assignment, the probability that it is assigned to label a is $(1/k)\,x_{pa}$, which is proportional to x_{pa}, and hence the probability that an object p is assigned label a by the process is precisely x_{pa}.

The probability that an object p is assigned any label in a phase is the sum of the probabilities over all labels a that it is assigned label a, which is exactly $1/k$. □

We say that the two ends of an edge are separated, or "split," by this process if they receive different labels. We say that an edge $e = (p, q)$ is *separated by a single phase* if before the phase both p and q are unassigned, and exactly one of p and q is assigned in this phase. Note that an edge that is separated is separated by some phase, but the reverse is not true, as later the other end of the edge may be assigned the same label.

LEMMA 3.2. *For an edge $e = (p, q)$ the probability that edge e is separated by a given phase, assuming neither end is assigned before the phase, is exactly $2z_e/k = \sum_{a\in L} |x_{pa} - x_{qa}|/k$. The probability that the two objects p and q are separated by the process is at most $2z_e$.*

PROOF. Consider a particular phase, and assume that objects p and q are both unassigned at the beginning of this phase. If label a is selected in this phase, then the probability that exactly one of p, q is assigned to label a is $|x_{pa} - x_{qa}|$. Hence, the probability that p and q are separated by this phase is equal to

$$\frac{1}{k}\sum_{a\in L} |x_{pa} - x_{qa}| = \frac{\sum_{a\in L} z_{ea}}{k} = \frac{2z_e}{k},$$

as claimed.

On the other hand, if p and q are both assigned a label in this phase (recall that they are unassigned at the beginning of the phase), then they clearly are not separated by the process—since all objects assigned a label in this phase receive the same label. The conditional probability that p and q are separated in this phase, given that at least one of p and q is assigned by this phase, is

$$\frac{\Pr[p, q \text{ are separated by this phase}]}{\Pr[\text{at least one of } p, q \text{ is assigned a label by this phase}]}.$$

The numerator is equal to $2z_e/k$. The denominator is at least the probability that object p is assigned, which is $1/k$ by Lemma 3.1, so the conditional probability is at most $2z_e$. This is true in any phase that starts with both p and q unassigned, and hence the probability that edge $e = (p, q)$ is split by the process is at most $2z_e$. □

Now we are ready to give the approximation theorem.

THEOREM 3.3. *Let x be a solution of the linear program (ULP) with assignment cost c_{LP} and separation cost w_{LP}. The randomized rounding procedure described above finds a labeling whose expected assignment cost is c_{LP}, and whose expected separation cost is at most $2w_{LP}$.*

PROOF. By Lemma 3.1, the probability that object p is assigned label a is x_{pa}, and hence the expected contribution of q to the objective function is $\sum_{a\in L} c_{pa}x_{pa}$, precisely its contribution in (ULP). Hence the total expected assignment cost over all objects is c_{LP}.

What is the contribution of edge $e = (p, q)$ to the expected separation cost? This is precisely w_e times the probability that its ends are split, since we are using the uniform metric. For an edge $e = (p, q)$ the probability that the two objects p and q are separated by the process is at most $2z_e$ by Lemma 3.2. Hence, the expected contribution of edge e is at most $2w_e z_e$, which is twice its contribution in (ULP). The total expected assignment cost over all objects is at most $2w_{LP}$. □

Using the optimal fractional labeling x we get a labeling with expected value within a factor of 2 of optimal. In Section 5, we show how to derandomize the procedure using the method of conditional probabilities.

COROLLARY 3.4. *If x is an optimal solution to* (*ULP*)*, then the rounding provides a* 2*-approximation for the uniform labeling problem.*

3.1. A TIGHT EXAMPLE. The following example shows that the integrality gap of our LP formulation is close to 2. Let G be the complete graph on k nodes with all edge weights w_e equal to 1, and let both the object set P and the label set L correspond to the set of nodes. We set all assignment costs equal to 0, except that we set $c(i, i)$ to infinity for all i. The optimum integer solution is to select an arbitrary i, assign all nodes $j \neq i$ label i, and assign node i some other label. The cost of this assignment is $k - 1$ for cutting the $k - 1$ edges out of node i. An optimal LP solution is obtained by setting $x_{ij} = 1/(k - 1)$, for all $i \neq j$. The separation cost of each edge in this assignment is $1/(k - 1)$, and the total over the $k(k - 1)/2$ edges is $k/2$. This is a ratio of $2(k - 1)/k$.

4. *The Metric Labeling Problem*

Next, we consider the metric labeling problem. Interestingly, we will be able to obtain an approximation algorithm despite the fact that we do not know of any "natural" linear programming relaxation for the problem.

We make use of Bartal's [1996, 1998] result that any finite metric space can be probabilistically approximated by hierarchically well-separated tree metrics [approximations of metric spaces and its algorithmic applications 1996; Bartal 1998]. Consequently, we focus first on the case in which the metric on L corresponds to an r-hierarchically well-separated tree [approximations of metric spaces and its algorithmic applications 1996]; here, we formulate a linear programming relaxation and develop a rounding algorithm whose performance guarantee is a constant slightly larger than 2, depending on the separation parameter r of the underlying tree.

An *r-hierarchically well-separated tree metric* (r-HST) is a rooted tree with lengths on its edges satisfying the following properties:

(i) The lengths of all the edges from a node to its children are the same.

(ii) The lengths of the edges along any path from the root to a leaf decrease by at least a factor of r in each step.

(iii) The metric distance between nodes of the tree is the sum of the edge lengths on the unique path between them.

We do not assume that all leaves are at the same level. However, we make the following additional assumptions. First, we require that $r > 2$; and second, we require that the labels L correspond to leaves of the tree. The uniform labeling

problem discussed in the previous section is the special case of this setting in which the tree is a star with the length of each tree edge is equal to $1/2$.

As our first main step, we give a $(2 + 4/(r-2))$-approximation algorithm for the metric labeling problem on an r-HST that repeatedly uses the 2-approximation algorithm of the previous section. We begin with the following lemma, a bound on the diameter that is a consequence of the hierarchical separation property.

By a *subtree* of an r-HST, we mean a tree consisting of all descendants of some node v that is not the root. For such a subtree T, let ℓ_T denote the length of the tree-edge leading upward from the root of T, and let $L(T)$ denote the set of leaves of T, that is, the set of labels that are nodes in the tree T.

LEMMA 4.1. *Let T^i be one of the subtrees rooted at a child of the root. For any two labels a and b, we have $d(a,b) \le 2r/(r-1)\,\ell_{T^i}$.*

PROOF. The distance between the roots of any two of the subtrees T^j and T^ℓ at a child of the root is $2\ell_{T^i}$. Edge lengths decrease by a factor of $r > 2$ as we go down the tree, and hence we get that

$$d(a,b) \le \left(2 + \frac{2}{r} + \frac{2}{r^2} + \cdots\right)\ell_{T^i} = \left(\frac{2r}{r-1}\right)\ell_{T^i}. \qquad \square$$

The metric labeling problem for tree metrics can be expressed as an integer program using the same variables x_{pa} that we used in the uniform case. We have nonnegative variables x_{pa} for each object p and label a such that $\sum_{a\in L} x_{pa} = 1$; that is, for integer-valued x, we use $x_{pa} = 1$ to denote that $f(p) = a$. To express the separation costs, we let $x_p(T)$ denote the fraction of object p assigned by a label in the subtree T, that is, $x_p(T) = \sum_{a\in L(T)} x_{pa}$. When x takes 0-1 values, then $x_p(T)$ is 1 if the object p is assigned to a label $f(p)$ in T and 0 otherwise. For an edge $e = (p,q)$ of our graph G, we need to express the distance between the assigned labels $f(p)$ and $f(q)$ using these variables. Recall that this distance $d(f(p), f(q))$ is the sum of the lengths of the edges of the r-HST on the path from $f(p)$ to $f(q)$. It can be written as $d(f(p), f(q)) = \sum_T \ell_T |x_p(T) - x_q(T)|$, where the sum is over all subtrees T of the r-HST, since $x_p(T)$ and $x_q(T)$ differ by 1 precisely for those subtrees T for which the tree-edge leading upward from the root of T is on the path between the labels $f(p)$ and $f(q)$ in T. The separation cost of an edge $e = (p,q)$ of G is w_e times this distance. We can write this as an integer linear program by introducing variables z_e for the distances, and z_{eT} to express each of the absolute values $|x_p(T) - x_q(T)|$.

$$\begin{array}{lll}
\text{(TIP)} \quad \text{Min} & \sum_{e\in E} w_e z_e + \sum_{p\in P, a\in L} c(p,a)\, x_{pa} & \\
\text{subject to} & \sum_{a\in L} x_{pa} = 1 & p \in P \\
& z_e = \sum_T \ell_T z_{eT} & e \in E. \\
& z_{eT} \ge x_p(T) - x_q(T) & e = (p,q), \text{ and subtree } T. \\
& z_{eT} \ge x_q(T) - x_p(T) & e = (p,q), \text{ and subtree } T. \\
& x_{pa} \in \{0,1\} & p \in P, a \in L.
\end{array}$$

We create the linear programming relaxation of (TIP) by replacing the integrality constraints with $x_{pa} \ge 0$; we call this (TLP). For tree metrics, we use $\sum_e w_e z_e = \sum_e [w_e \sum_T \ell_T |x_p(T) - x_q(T)|]$ as the *separation cost* of the fractional labeling x. As before, we use a randomized rounding scheme using the fractional labeling x as a probability. Note that if x_{pa} is the probability that object p is assigned to

label a then the probability that object p is assigned a label in the subtree T is exactly $x_p(T)$.

We define the randomized rounding algorithm as follows:

(i) We consider first the top level of the tree. Assume that the root has m children, and let $T^1, \ldots, T^m$ denote the subtrees rooted at the children. We use the uniform labeling algorithm to assign each object to one of the subtrees T^i, where object p has probability $x_p(T^i)$ to be assigned to subtree T^i. Thus, at the end of this step, each object has been assigned to one of the subtrees, but not yet placed *within* this subtree.

Intuitively, dealing with the top level of the tree separately makes sense as the distances within each subtree T^i are significantly smaller than the distances between subtrees. Using the uniform labeling algorithm for the assignment of objects to subtrees is appropriate, as the edge lengths from the root to each of its children are the same.

(ii) Once object p has been assigned to subtree T^i, we modify the fractional labeling accordingly by setting $\bar{x}_{pa} = 0$ for all labels a that are not in the subtree T^i, and $\bar{x}_{pa} = x_{pa}/x_p(T^i)$ for all labels $a \in L(T^i)$.

(iii) The objects assigned to each subtree will now be treated as separate subproblems. We use the algorithm recursively on each of the subtrees T^i to obtain the labeling.

To analyze the top level of this algorithm, we use the following property that was proved as Lemmas 3.1 and 3.2 in the previous section.

LEMMA 4.2. *The probability that object p is assigned to the subtree T^i is $x_p(T^i)$. For each edge $e = (p, q)$ of the graphs G, the probability that p and q are assigned to different trees is at most $\sum_{i=1}^{m} |x_p(T^i) - x_q(T^i)|$.*

When we consider a set of objects that have all been assigned to the same subtree, we have the problem that their fractional separation costs have changed. We bound this in Lemma 4.4. The bound relies on the property that edge lengths decrease by a factor of $r > 2$ as we go down the tree, and hence the distance between two labels is dominated by the highest edge on the path between them. We will need an analogous fact for fractional labelings.

For two subtrees T' and T, we write "$T' < T$" to denote the statement that T' is a proper subtree of T consisting of all descendants of a node $v \in T$ that is not the root of T.

LEMMA 4.3. *Let x be a fractional labeling over an r-hierarchically well-separated tree, let p be an object, and let T be a subtree. Then*

$$\sum_{T' < T} \ell_{T'} x_p(T') \leq \frac{1}{r-1} \cdot \ell_T x_p(T).$$

PROOF. If t is the root of T, we say that a subtree is at *level* j if its root is reachable by a path leading j edges downward from t. Note that we have

$$\sum_{\substack{T' < T \\ T' \text{ at level } j}} x_p(T') \leq x_p(T)$$

for all j. Thus,

$$\sum_{T'<T} \ell_{T'}\, x_p(T') = \sum_{j\geq 1} \sum_{\substack{T'<T \\ T' \text{ at level } j}} \ell_{T'}\, x_p(T') \leq \sum_{j\geq 1} r^{-j}\, \ell_T \sum_{\substack{T'<T \\ T' \text{ at level } j}} x_p(T')$$
$$\leq \sum_{j\geq 1} r^{-j}\, \ell_T\, x_p(T) \leq \frac{1}{r-1} \cdot \ell_T\, x_p(T). \qquad \square$$

LEMMA 4.4. *Let $e = (p, q)$ be an edge of the graph G. If both p and q are assigned to subtree T^i, then the new fractional separation cost is at most*

$$\frac{w_e}{\min(x_p(T^i), x_q(T^i))} \left[\sum_{T<T^i} \ell_T |x_p(T) - x_q(T)| + \frac{1}{r-1} \ell_{T^i} \left| x_p(T^i) - x_q(T^i) \right| \right].$$

PROOF. Assume that $x_p(T^i) \leq x_q(T^i)$. The new fractional separation cost for edge e is $w_e \sum_T \ell_T |\bar{x}_p(T) - \bar{x}_q(T)|$, and if p and q are both assigned to T^i then the sum can be written as

$$\begin{aligned}
&\sum_{T<T^i} \ell_T |\bar{x}_p(T) - \bar{x}_q(T)| \\
&= \sum_{T<T^i} \ell_T \left| \frac{x_p(T)}{x_p(T^i)} - \frac{x_q(T)}{x_q(T^i)} \right| \\
&= \sum_{T<T^i} \ell_T \left| \left(\frac{x_p(T)}{x_p(T^i)} - \frac{x_q(T)}{x_p(T^i)} \right) - \left(\frac{x_q(T)}{x_q(T^i)} - \frac{x_q(T)}{x_p(T^i)} \right) \right| \\
&\leq \sum_{T<T^i} \ell_T \frac{1}{x_p(T^i)} |x_p(T) - x_q(T)| + \sum_{T<T^i} \left[\frac{1}{x_p(T^i)} - \frac{1}{x_q(T^i)} \right] \ell_T\, x_q(T) \\
&\leq \frac{1}{x_p(T^i)} \left[\sum_{T<T^i} \ell_T |x_p(T) - x_q(T)| \right] + \frac{1}{r-1} \ell_{T^i} x_q(T^i) \left[\frac{1}{x_p(T^i)} - \frac{1}{x_q(T^i)} \right] \\
&= \frac{1}{x_p(T^i)} \left[\sum_{T<T^i} \ell_T |x_p(T) - x_q(T)| \right] + \frac{1}{r-1} \ell_{T^i} \frac{x_q(T^i) - x_p(T^i)}{x_p(T^i)} \\
&= \frac{1}{x_p(T^i)} \left[\sum_{T<T^i} \ell_T |x_p(T) - x_q(T)| + \frac{1}{r-1} \ell_{T^i} \left(x_q(T^i) - x_p(T^i) \right) \right],
\end{aligned}$$

where the last inequality follows from Lemma 4.3. $\square$

We can now prove the main theorem about labeling with hierarchically well-separated tree metrics. We give a derandomized version of this algorithm in Section 5.

THEOREM 4.5. *Let x be a fractional labeling of objects P for an r-hierarchically well-separated tree metric (with $r > 2$) whose leaves are the labels. Let c_{LP} and w_{LP} be respectively the assignment and separation cost of the fractional assignment x. The randomized rounding procedure described above finds a labeling*

whose expected assignment cost is c_{LP}, and whose expected separation cost is at most $(2 + 4/(r-2))w_{LP}$.

PROOF. We prove both statements about the expected cost by induction on the depth of the tree. If the tree has depth 0 there is nothing to prove. Otherwise, as before, let $T^1, \ldots, T^m$ denote the subtrees rooted at the children of the root.

First, consider the statement about the expected assignment costs. We claim that the probability that an object p is assigned a label a is exactly x_{pa}. This will imply the statement about the expected assignment cost. Consider a top-level subtree. For an object p, the probability that p gets assigned to a subtree T^i is $x_p(T^i)$, and if p is assigned to the subtree T^i, then the fractional assignments are rescaled to be $\bar{x}_{pa} = (1/x_p(T^i))\, x_{pa}$ for each label a in the subtree T^i. By induction, once an object p was assigned to subtree T^i, it gets a label $a \in L(T)$ with probability $\bar{x}_{pa}$, and hence the probability that object p is assigned a label a in $L(T^i)$ is $x_p(T^i)\bar{x}_{pa} = x_{pa}$ as claimed.

Next consider the expected separation cost for an edge $e = (p, q)$ in the graph G. Let $h(p, q)$ be the random variable that is the distance between the labels assigned to objects p and q. The separation cost of edge e is now $w_e h(p, q)$, and the expected separation cost is $w_e \mathrm{E}[h(p, q)]$. The fractional separation cost is $w_e \sum_T \ell_T |x_p(T) - x_q(T)|$. We refer to $\sum_T \ell_T |x_p(T) - x_q(T)|$ as the *fractional distance*, and want to prove that the expected distance at most $2r/(r-2)$ times the fractional distance, which will imply the claim of the lemma.

We compute the expected distance $\mathrm{E}[h(p, q)]$ by considering separately the event $\mathcal{E}'$ that p and q are separated at the top level, and the events $\mathcal{E}_i$ that p and q are both assigned to the same subtree T^i. We can write $\mathrm{E}[h(p, q)]$ as follows:

$$\mathrm{E}[h(p,q)] = \mathrm{E}[h(p,q)|\mathcal{E}'] \cdot \Pr[\mathcal{E}'] + \sum_{i=1}^{m} \mathrm{E}[h(p,q)|\mathcal{E}_i] \cdot \Pr[\mathcal{E}_i].$$

The probability $\Pr[\mathcal{E}']$ that the edge gets separated at the top level is at most $\sum_{i=1}^{m} |x_p(T^i) - x_q(T^i)|$ by Lemma 4.2. The distance in this case is bounded by the diameter of the tree, which is at most $[2r/(r-1)]\,\ell_{T^i}$ for any i by Lemma 4.1. So the first term of the above sum is

$$\mathrm{E}[h(p,q)|\mathcal{E}'] \cdot \Pr[\mathcal{E}'] \le \frac{2r}{r-1} \sum_{i=1}^{m} \ell_{T^i} \left| x_p(T^i) - x_q(T^i) \right|.$$

If both objects are assigned to the same subtree T^i, then by induction the expected distance is at most $2r/(r-2)$ times the fractional distance of the scaled solution. The probabilities that objects p and q are assigned to a subtree T^i are $x_p(T^i)$ and $x_q(T^i)$, respectively, and hence the probability that both objects are assigned to the subtree T^i is at most $\min(x_p(T^i), x_q(T^i))$. Using Lemma 4.4, we can bound $\mathrm{E}[h(p, q)|\mathcal{E}_i] \cdot \Pr[\mathcal{E}_i]$ as follows:

$$\mathrm{E}[h(p,q)|\mathcal{E}_i] \cdot \Pr[\mathcal{E}_i] \le \frac{2r}{r-2} \left[\sum_{T < T^i} \ell_T |x_p(T) - x_q(T)| + \frac{1}{r-1} \ell_{T^i} \left| x_p(T^i) - x_q(T^i) \right| \right].$$

The final bound on $\mathrm{E}[h(p,q)]$ then follows

$$\begin{aligned}\mathrm{E}[h(p,q)] \;&\le\; \frac{2r}{r-1}\sum_{i=1}^{m}\ell_{T^i}\left|x_p(T^i)-x_q(T^i)\right| \\ &\quad+\;\frac{2r}{r-2}\sum_{i=1}^{m}\left[\sum_{T<T^i}\ell_T|x_p(T)-x_q(T)|+\frac{1}{r-1}\ell_{T^i}\left|x_p(T^i)-x_q(T^i)\right|\right] \\ &\quad=\;\frac{2r}{r-2}\sum_{T}\ell_T|x_p(T)-x_q(T)|,\end{aligned}$$

where the last sum is over all subtrees T of the tree that defines the metric on the labels. □

Using the optimal fractional labeling x of (TLP), we get a labeling with expected value within a small factor of optimal.

COROLLARY 4.6. *If x is an optimal solution to (TLP), then the rounding provides a $(2+4/(r-2))$-approximation for the labeling problem for an r-hierarchically well-separated tree metric.*

We now apply Bartal's result on probabilistic approximation of metric spaces to address the labeling problem with general metrics. Bartal's result applied to the metric d on the label set L states that there is a set of r-hierarchically well-separated tree metrics $\{d_T\}$ on a superset L' of L, and probability distribution $\{p_T\}$, so that $d_T(a,b)\ge d(a,b)$ for all $a,b\in L$ and all trees T in the set; and yet for all pairs of labels a,b the expected a-b distance in a randomly selected tree is not much bigger than $d(a,b)$. More precisely, $\sum_T p_T d_T(a,b)\le O(r\log k\log\log k)d(a,b)$. We use this result to obtain an $O(\log k\log\log k)$-approximation algorithm for the labeling problem in general metric spaces.

We first show that at an additional loss of a factor of $2+1/r$ we may assume that the well-separated trees in Bartal's theorem have all labels at the leaves—that is, the internal nodes of the trees are not labels.

LEMMA 4.7. *Let d_T denote an r-HST distance function generated by the tree T. We can modify the tree T to obtain an r-HST distance function generated by a tree T' where all labels are at leaves, and for any two labels a and b we have $d_T(a,b)\le d_{T'}(a,b)\le(2+1/r)d_T(a,b)$.*

PROOF. For each internal node a of T that is a label, we replace this node in the tree by a new node a', and insert the label a as a new child of a' at the same distance as all other children. This change in the tree will only increase the distances between the labels. Next we bound the maximum possible increase. First consider the case when a node a and a child b of node a are both labels. Let $\ell=d_T(a,b)$. Then a is pushed down to be at a distance ℓ from the new node a', and b is pushed down from a new node b'. If the tree is an r-HST, then the new distance $d_{T'}(b,b')$ is at most ℓ/r, and so we have that $d_{T'}(a,b)\le 2\ell+\ell/r=(2+1/r)d_T(a,b)$. It is not hard to see that the above case represents the largest possible change in distance, and so $d_{T'}(a,b)\le(2+1/r)d_T(a,b)$ for all labels a and b. □

It is now easy to show that the following randomized algorithm provides an expected performance guarantee of $O(\log k\log\log k)$: We probabilistically

approximate the metric using 3-HST's, choose a random tree T according to the distribution $\{p_T\}$, and run our algorithm above on the metric induced by this tree. We describe a derandomized version in Section 5.

THEOREM 4.8. *There is a randomized polynomial time $O(\log k \log\log k)$-approximation algorithm for the metric labeling problem where k is the number of labels.*

PROOF. Let d be a metric on the set of labels L. By Bartal's tree-metric approximation, and the above Lemma 4.7, we can find a set of 3-hierarchically well-separated tree-metrics that approximate the metric d. Let $\mathcal{T}$ denote the set of tree metrics, and p_T for $T \in \mathcal{T}$ be a probability distribution on the trees so that for each pair of labels $a, b \in L$ and each tree $T \in \mathcal{T}$ we have $d(a,b) \le d_T(a,b)$ and $\sum_{T\in\mathcal{T}} p_T d_T(a,b) \le O(\log k \log\log k) d(a,b)$.

The algorithm is as follows: We choose a tree metric $T \in \mathcal{T}$ choosing the tree T with probability p_T, and run the algorithm for the tree metric defined by T.

Let $\mathcal{O}$ denote the cost of an optimal labeling for the metric d. Let $\mathcal{O}_T$ for $T \in \mathcal{T}$ denote the cost of the same labeling under the tree metric T. By the definition of the tree-metric approximation, we have that $\sum_{T\in\mathcal{T}} p_T \mathcal{O}_T \le O(\log k \log\log k)\,\mathcal{O}$, as this inequality is true term-by-term for the separation cost of each edge.

Let the random variable $\mathcal{A}$ denote the cost using metric d of the labeling produced by the algorithm, and let $\mathcal{A}_T$ denote the cost of the solution found in the tree metric T. By the definition of the approximation, we have that $\mathcal{A} \le \mathcal{A}_T$ for all $T \in \mathcal{T}$.

We claim that the expected cost of the resulting labeling $E[\mathcal{A}]$ is $O(\log k \log\log k)\,\mathcal{O}$. Let $\mathcal{E}_T$ denote the event that tree T is selected. We have that $\Pr(\mathcal{E}_T) = p_T$ and that $\mathrm{E}[\mathcal{A}_T|\mathcal{E}_T] \le 6\mathcal{O}_T$ by Corollary 4.6 (since we have chosen $r = 3$ for our r-HST). Using this, we get the following:

$$\begin{aligned}\mathrm{E}[\mathcal{A}] &= \sum_{T\in\mathcal{T}} \mathrm{E}[\mathcal{A}|\mathcal{E}_T]\cdot\Pr(\mathcal{E}_T) \le \sum_{T\in\mathcal{T}} p_T \mathrm{E}[\mathcal{A}_T|\mathcal{E}_T] \\ &\le 6\sum_{T\in\mathcal{T}} p_T O_T = O(\log k \log\log k)\,\mathcal{O}.\end{aligned}$$

□

5. *Deterministic Algorithms*

Next we show how to derandomize the randomized algorithms described in the previous two sections to provide deterministic approximation algorithms with the same approximation bound.

5.1. THE UNIFORM LABELING ALGORITHM. To design a deterministic version of our uniform labeling algorithm, we consider a single phase of the experiment. We use the method of conditional probabilities to choose the appropriate step to take in this phase.

Consider the first phase. Let a denote the random label chosen by the phase, let P_a denote the set of objects assigned by this phase, and let $\bar{P}$ denote the set of objects not assigned. We consider the following measures:

— $c_0 = \sum_{p\in P_a} c(p,a)$, the assignment cost of all objects assigned in this phase,

— $w_0 = \sum_{e=(p,q)\in E:|\{p,q\}\cap P_a|=1} w_e$, the separation cost of all pairs of objects separated by this phase,

— $\bar{c}_{LP} = \sum_{p \in \bar{P}, a \in L} c(p, a)\, x_{pa}$, the assignment cost of (ULP) restricted to the unassigned objects in $\bar{P}$,

— $\bar{w}_{LP} = \sum_{e=(p,q) \in E : p,q \in \bar{P}} w_e z_e$, the separation cost of (ULP) restricted to edges between the unassigned objects in $\bar{P}$.

A phase of the derandomized algorithm selects a label a and a value α that leads to a nonempty set P_a and minimizes the quantity $c_0 + w_0 + \bar{c}_{LP} + 2\bar{w}_{LP}$. We repeat the method recursively on the remaining objects to obtain a labeling deterministically. Note that the method uses one solution to the linear program throughout all recursive calls; that is, the definition of $\bar{c}_{LP}$ and $\bar{w}_{LP}$ does not involve solving a new LP. Resolving the linear program at every step might provide a better quality solution, but the given solution is good enough to provide the claimed approximation bound.

Let c_{LP} and w_{LP} denote the assignment and separation cost of (ULP), respectively. We first show that the expected value of quantity $(c_0 + w_0) + (\bar{c}_{LP} + 2\bar{w}_{LP})$ in a phase of the randomized algorithm is at most $c_{LP} + 2w_{LP}$.

LEMMA 5.1. *In one phase of the randomized algorithm,* $\mathrm{E}[c_0 + w_0 + \bar{c}_{LP} + 2\bar{w}_{LP}] \le c_{LP} + 2w_{LP}$.

PROOF. The probability that object p is assigned label a in a single phase is x_{pa}/k by Lemma 3.1, and hence the expected value of the assignment cost c_0 is

$$\mathrm{E}[c_0] = \sum_{p \in P, a \in L} \frac{c(p, a)\, x_{pa}}{k} = \frac{1}{k} c_{LP}.$$

The probability that an edge e is separated by the phase is $2z_e/k$ by Lemma 3.2. Hence, the expected value of the separation cost w_0 is

$$\mathrm{E}[w_0] = \sum_{e \in E} \frac{2 w_e z_e}{k} = \frac{2}{k} w_{LP}.$$

The probability that an object p is *not* assigned is $(k-1)/k$ by Lemma 3.1, and hence the expected value of the remaining LP assignment cost $\bar{c}_{LP}$ is

$$\mathrm{E}[\bar{c}_{LP}] = \frac{k-1}{k} \sum_{p \in P, a \in L} c(p, a)\, x_{pa} = \frac{k-1}{k} c_{LP}.$$

The probability that for an edge $e = (p, q)$ both objects p and q remain unassigned can be bounded by the probability that object p is unassigned, and hence it is at most $(k-1)/k$. Using this bound, we can bound the expected value of the remaining LP separation cost $\bar{w}_{LP}$ as follows:

$$\mathrm{E}[\bar{w}_{LP}] \le \frac{k-1}{k} \sum_{e \in E} w_e z_e = \frac{k-1}{k} w_{LP}.$$

Summing up the expectations, we get the claimed bound. □

THEOREM 5.2. *Let x be a solution of the linear program (ULP) with assignment cost c_{LP} and separation cost w_{LP}. The derandomized rounding procedure described above finds a labeling whose cost is at most $c_{LP} + 2w_{LP}$. If x is an optimal solution to (ULP), then the rounding provides a 2-approximation for the uniform labeling problem.*

PROOF. To prove the first statement, consider a single phase. We use Lemma 5.1 to show that the algorithm chooses a label a and a value α such that the quantity $c_0 + w_0 + \bar{c}_{LP} + 2\bar{w}_{LP}$ is at most $c_{LP} + 2w_{LP}$. Choices of label a and value α that do not assign any object contribute to the expectation exactly $c_{LP} + 2w_{LP}$ times the probability that such a label and value is selected. By Lemma 5.1, the expected value is at most $c_{LP} + 2w_{LP}$, and hence the value of the expectation is at most $c_{LP} + 2w_{LP}$ even if restricted to choices of labels a and values α that assign some objects. Therefore, there have to be possible labels a and values α that assign objects and have $(c_0 + w_0) + (\bar{c}_{LP} + 2\bar{w}_{LP}) \leq c_{LP} + 2w_{LP}$.

We prove by induction on the number of phases that the algorithm results in a solution of cost at most $c_{LP}+2w_{LP}$. We apply the induction hypothesis to the problem on the set P of objects remaining after the first phase. By the induction hypothesis, the cost of the solution obtained for the objects P has cost at most $\bar{c}_{LP} + 2\bar{w}_{LP}$. Hence, the overall cost of the solution is at most $(c_0 + w_0) + (\bar{c}_{LP} + 2\bar{w}_{LP})$, which is at most $c_{LP} + 2w_{LP}$.

The 2-approximation follows from the first statement. □

5.2. HIERARCHICALLY WELL-SEPARATED TREES. The derandomized algorithm for hierarchically well-separated tree metrics is analogous to the derandomized algorithm for the uniform labeling problem. For clarity of exposition, the overall structure of the deterministic algorithm will be somewhat different from that of the randomized algorithm introduced earlier. First, consider the following randomized algorithm:

(i) As before, we consider first the top level of the tree. Assume that the root has m children, and let $T^1, \ldots, T^m$ denote the subtrees rooted at the children. We consider one phase of the uniform labeling algorithm to assign a set of objects to one of the subtrees T^i. More precisely, we select a tree T^i uniformly at random, and a real number α uniformly in $[0, 1]$. We assign object p to the selected subtree T^i if $x_p(T^i) \geq \alpha$.

(ii) As before, the objects assigned to subtree T^i are not yet placed *within* this subtree. At the end of this phase, we will have two smaller subproblems: one subproblem on the subtree T^i consisting of the objects that were assigned to subtree T^i, and another subproblem for the remaining objects on the original tree T. For the subproblem on the original tree T, we use the same fractional labeling, while the subproblem on the subtree T^i uses the modified fractional labeling $\bar{x}_{pa} = x_{pa}/x_p(T^i)$ for all labels $a \in L(T^i)$ and all objects p assigned to subtree T^i. We also use the corresponding separation distance variables $\bar{z}_e$ for edges $e = (p, q)$ between objects p, q that are both assigned to subtree T^i: $\bar{z}_e = \sum_{T<T^i} \ell_T |\bar{x}_p(T) - \bar{x}_q(T)|$.

(iii) The algorithm solves both subproblems recursively to obtain the labeling.

Note the difference in the overall structure of this algorithm and the algorithm introduced earlier: in the earlier algorithm, we first assigned all objects to some subtree, and then continued recursively, for each i, on the set of *all* objects assigned to subtree T^i. In the new algorithm, we immediately create two recursive subproblems after assigning some objects to a subtree T^i; other objects may be assigned to T^i in subsequent, separate subproblems. We adopt this version for the deterministic algorithm simply for clarity of exposition; a deterministic algorithm that more directly follows the control flow of the randomized algorithm can also be obtained by similar means.

Now, consider the first phase of our new algorithm at the top level of the tree: the phase selects a subtree T^i and assigns some objects to this subtree. Let i denote the index of the subtree chosen by the phase, let P_i denote the set of objects assigned to the subproblem, and let $\bar{P}$ denote the set of objects not assigned by this phase. We consider measures analogous to the ones considered in the uniform case. For each of the two subproblems, we consider LP solutions for the subproblems: the restriction of the original LP solution x for the subproblem of the objects $\bar{P}$, and the fractional labeling $\bar{x}$ given above, for the subproblem of the objects P_i. For edges between objects in separate subproblems, we bound the separation cost by the diameter of the tree, which is at most is at most $[2r/(r-1)]\,\ell_{T^i}$ for any i by Lemma 4.1.

— $\bar{c}_s = \sum_{p \in P_i, a \in L(T^i)} \bar{x}_{pa} c(p,a)$, the assignment cost of (TLP) of the fractional labeling $\bar{x}$ for the objects assigned to the subproblem on the tree T^i in this phase,

— $\bar{w}_s = \sum_{e=(p,q) \in E: p,q \in P_i} w_e \bar{z}_e$, the fractional separation cost of (TLP) of the fractional labeling $\bar{x}$ for all pairs of objects assigned to the subproblem on the subtree T^i in this phase,

— $\bar{c} = \sum_{p \in \bar{P}, a \in L} c(p,a) x_{pa}$, the assignment cost of (TLP) of the original fractional labeling x restricted to the unassigned objects $\bar{P}$,

— $\bar{w} = \sum_{e=(p,q) \in E: p,q \in \bar{P}} w_e z_e$, the separation cost of (TLP) of the original fractional labeling x restricted to the edges between the unassigned objects in $\bar{P}$.

— $w_s = [2r/(r-1)]\ell_{T^i} \sum_{e=(p,q) \in E: |\{p,q\} \cap P_i| = 1} w_e$, an upper bound of the separation cost of all pairs of objects separated by this phase.

A phase of the derandomized algorithm selects a top level subtree T^i and a value α that leads to a nonempty set P_i such that the quantity $\bar{c}_s + \bar{c} + w_s + 2r/(r-2)(\bar{w}_s + \bar{w})$ is as small as possible. We repeat the method on both of the subproblems created to obtain a labeling deterministically.

Let c_{LP} and w_{LP} denote the assignment and separation cost of (TLP). We will first prove the analog of Lemma 5.1, that the expected value of the above quantity in a phase of the randomized algorithm is at most $c_{LP} + 2r/(r-2)\, w_{LP}$.

LEMMA 5.3. *In one phase of the above randomized algorithm* $\mathrm{E}[\bar{c}_s + \bar{c} + w_s + 2r/(r-2)(\bar{w}_s + \bar{w})] \le c_{LP} + 2r/(r-2)\, w_{LP}$.

PROOF. The probability that object p is assigned to subtree T^i is $x_p(T^i)/m$, as the probability that the tree T^i is selected is $1/m$, and if the tree T^i is selected, the probability that object p is assigned to this tree is $x_p(T^i)$.

If p is assigned to subtree T^i, its fractional assignment cost is

$$\sum_{a \in L(T^i)} \bar{x}_{pa} c(p,a) = \sum_{a \in L(T^i)} \frac{x_{pa} c(p,a)}{x_p(T^i)}.$$

The contribution of object p to the expected value of $\bar{c}_s$ is the probability that p is assigned to subtree T^i times the above fractional assignment cost, which is

$$\sum_i \left(x_p(T^i)/m\right) \left(\sum_{a \in L(T^i)} \frac{x_{pa} c(p,a)}{x_p(T^i)} \right) = \frac{1}{m} \sum_{a \in L} x_{pa} c(p,a).$$

This is exactly a $1/m$ fraction of its contribution to the LP assignment cost c_{LP}. So the expected value of $\bar{c}_s$ is c_{LP}/m.

By Lemma 3.1, the probability that object p remains unassigned in this phase is $(m-1)/m$, and hence expected value of $\bar{c}$ is $\mathrm{E}[\bar{c}] = [(m-1)/m]\, c_{LP}$.

Similarly, the expected value of $\bar{w}$ is at most $\mathrm{E}[\bar{w}] \le (m-1)\, w_{LP}/m$, as the probability that both ends of an edge $e = (p, q)$ remain unassigned is at most the probability that p remains unassigned, which is $(m-1)/m$.

Now consider the contribution of an edge $e = (p, q)$ of G to the expected value of the fractional separation cost $\bar{w}_s$. Both of the objects p and q are assigned to subtree T^i if the subtree T^i is selected and a value $\alpha \le \min(x_p(T^i), x_q(T^i))$ is selected; hence the probability that both of the objects p, q are assigned to subtree T^i is exactly $\min(x_p(T^i), x_q(T^i))/m$. Lemma 4.4 bounds the fractional separation cost of the edge $e = (p, q)$ assuming that both p and q are assigned to the subtree T^i. To get the expected value of $\bar{w}_s$, we multiply this value with the probability that both p and q are assigned to the subtree T^i, and sum over all edges; we get

$$\mathrm{E}[\bar{w}_s] = \frac{1}{m} \sum_{e=(p,q)\in E} w_e \sum_i \left[\sum_{T<T^i} \ell_T |x_p(T) - x_q(T)| + \frac{1}{r-1} \ell_{T^i} \left| x_p(T^i) - x_q(T^i) \right| \right].$$

Note that the summation in the first term is over proper subtrees $T < T^i$ (as implied by Lemma 4.4), and hence it does *not* include the contribution of the top level subtrees T^i for $i = 1, \ldots, m$ to the LP separation cost w_{LP}.

Finally, by Lemma 3.2, the probability that edge $e = (p, q)$ is separated by this phase is $\sum_i |x_p(T^i) - x_q(T^i)|/m$, that is, depends only on the contribution of the top level subtrees T^i for $i = 1, \ldots, m$ to the LP separation cost w_{LP}. We bound the separation cost by the diameter of the tree, using Lemma 4.1. So the expected value of w_s is

$$\mathrm{E}[w_s] \le \frac{2r}{m(r-1)} \sum_{e=(p,q)\in E} w_e \sum_i \ell_{T^i} \left| x_p(T^i) - x_q(T^i) \right|,$$

because the length ℓ_{T^i} is the same for all top level subtrees. Summing up the expectations, we get

$$\begin{aligned}
&\mathrm{E}\left[\bar{c}_s + \bar{c} + w_s + \frac{2r}{r-2}(\bar{w}_s + \bar{w}) \right] \\
&\quad \le \frac{1}{m} c_{LP} + \frac{m-1}{m} c_{LP} + \frac{2r}{m(r-1)} \sum_{e=(p,q)\in E} w_e \sum_i \ell_{T^i} \left| x_p(T^i) - x_q(T^i) \right| \\
&\qquad + \frac{2r}{r-2} \left[\frac{1}{m} \sum_{e=(p,q)\in E} w_e \sum_i \left(\sum_{T<T^i} \ell_T |x_p(T) - x_q(T)| \right.\right. \\
&\qquad \left.\left. + \frac{1}{r-1} \ell_{T^i} \left| x_p(T^i) - x_q(T^i) \right| \right) + \frac{m-1}{m} w_{LP} \right].
\end{aligned}$$

To see that this is equal to the claimed bound of $c_{LP} + 2r/(r-2)\, w_{LP}$, we consider the assignment cost and the separation cost separately. The assignment cost is included in the sum $1/m + (m-1)/m = 1$ time, as claimed. We will consider the contributions of the separation cost of the top-level trees separately from the other levels; accordingly, we will break up the $2r/(r-2)\, w_{LP}$ term into portions

corresponding to the separation cost for the top-level trees, and for the other levels. The contribution of the separation cost for the top-level trees, $\sum_{e=(p,q)\in E} w_e \sum_i \ell_{T^i} |x_p(T^i) - x_q(T^i)|$, is included in the sum

$$\frac{2r}{m(r-1)} + \frac{2r}{r-2}\left(\frac{1}{m(r-1)} + \frac{m-1}{m}\right)$$

times, which is equal to $2r/(r-2)$ in total. The contributions of the separation cost for other levels of the tree are included

$$\frac{2r}{r-2}\left(\frac{1}{m} + \frac{m-1}{m}\right)$$

times, which is also $2r/(r-2)$ in total. □

As in the proof of Theorem 5.2, we get the following:

THEOREM 5.4. *Let x be a fractional labeling of objects P for an r-hierarchically well-separated tree metric (with $r > 2$) whose leaves are the labels. Let c_{LP} and w_{LP} be, respectively, the assignment and separation cost of the fractional assignment x. The derandomized rounding procedure described above finds a labeling whose cost is at most $c_{LP} + (2 + 4/(r-2))\, w_{LP}$. If x is an optimal solution to (TLP), then the rounding described above provides a $(2 + 4/(r-2))$-approximation for the labeling problem.*

PROOF. As in the proof of Theorem 5.2, we can argue using Lemma 5.3 that there is a choice of subtree i and value α that assigns a nonempty set of objects to the subtree T^i and has $\bar{c}_s + \bar{c} + w_s + 2r/(r-2)(\bar{w}_s + \bar{w}) \le c_{LP} + 2r/(r-2)\, w_{LP}$.

For any such choice, both subproblems are smaller: the subproblem on T has fewer objects, and the subproblem on the subtree T^i is defined over a smaller tree. So we can assume by induction that the theorem holds for the subproblem. Now the overall bound of $c_{LP} + 2r/(r-2) w_{LP}$ for the value of the solution follows by induction: the two subproblems have solutions of value at most $\bar{c}_s + 2r/(r-2)\,\bar{w}_s$, and $\bar{c} + 2r/(r-2)\bar{w}$, respectively. The separation cost of the edges separated in this phase is at most w_s, and hence the overall cost of the solution found is at most

$$\left(\bar{c}_s + \frac{2r}{r-2}\bar{w}_s\right) + \left(\bar{c} + \frac{2r}{r-2}\bar{w}\right) + w_s \le c_{LP} + \frac{2r}{r-2} w_{LP}. \qquad \square$$

5.3. METRIC LABELING ALGORITHM. To make the overall algorithm for general metric spaces deterministic, we need only derandomize the choice of tree in Bartal's tree metric approximation theorem. For this, we use the result of Charikar et al. [1998] that any metric space on k points can be probabilistically approximated by hierarchically well-separated tree metrics using only $O(k \log k)$ trees, and that such a set of trees can be computed deterministically in polynomial time. Thus, we compute these $O(k \log k)$ trees, run the approximation algorithm for each of them, and take the solution on which the objective function with respect to the original metric is best.

THEOREM 5.5. *The above algorithm is a deterministic polynomial time $O(\log k \log\log k)$-approximation algorithm for the metric labeling problem, where k is the number of labels.*

ACKNOWLEDGMENTS. We would like to thank Yuri Boykov, Olga Veksler and Ramin Zabih for pointing out to us the relation of MRF's and multiway cuts, and for numerous discussions on the topic.

REFERENCES

BARTAL Y. 1996. Approximations of metric spaces, and its algorithmic applications. In *Proc. IEEE Symposium on Foundations of Computer Science*.

BARTAL, Y. 1998. On approximating arbitrary metrics by tree metrics. In *Proceedings of the ACM Symposium on Theory of Computing*. ACM, New York.

BESAG, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Stat. Soc. B 36*.

BESAG, J. 1986. On the statistical analysis of dirty pictures. *J. Roy. Stat. Soc. B 48*.

BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 1998. Markov random fields with efficient approximations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE computer Society Press, Los Alamtos, Calif.

BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 1999. Fast approximate energy minimization via graph cuts. In *Proceedings of the International Conference on Computer Vision*.

BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 2001. Fast approximate energy minimization via graph cuts. IEEE *Trans. Patt. Anal. Mach. Intel. 23*, 11 (Nov.), 1222–1239.

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A., AND STONE, C. J. 1984. *Classification and Regression Trees*. Wadsworth & Brooks/Cole.

CALINESCU, G., KARLOFF, H., AND RABANI, Y. 1998. Improved approximation algorithms for multiway cut. In *Proceedings of the ACM Symposium on Theory of Computing.* ACM, New York.

CHAKRABARTI, S., DOM, B., AND INDYK, P. 1998. Enhanced hypertext categorization using hyperlinks. In *Proceedings of ACM SIGMOD.* ACM, New York.

CHARIKAR, M., CHEKURI, C., GOEL, A., GUHA, S., AND PLOTKIN, S. 1998. Approximating arbitrary metrics by $O(n \log n)$ trees. In *Proceedings of the IEEE Symposium on Foundations of Computer Science*. IEEE Computer Society Press, Los Alamitos, Calif.

CHELLAPPA, R., AND JAIN, A. 1993. *Markov Random Fields: Theory and Applications*. Academic Press, Orlando, Fla.

COHEN, F. S. 1986. Markov random fields for image modeling and analysis. In *Modeling and Application of Stochastic Processes*, U. B. Desai, Ed. Kluwer.

CUNNINGHAM, W. H., AND TANG, L. 1994. Optimal 3-terminal cuts and linear programming. In *Proceedings of the MPS Conference on Integer Programming and Combinatorial Optimization*.

DAHLHAUS, E., JOHNSON, D., PAPADIMITRIOU, C., SEYMOUR, P., AND YANNAKAKIS, M. 1994. The complexity of multi-terminal cuts. *SIAM J. Comput. 23*.

DIETTERICH, T. G. 1997. Machine learning research: Four current directions. *AI Mag. 18*.

DUBES, R., AND JAIN, A. 1989. Random field models in image analysis. *J. Appl. Stat. 16*.

ERDÖS, P. L., AND SZÉKELY, L. A. 1992. Evolutionary trees: An integer multicommodity max-flow min-cut theorem. *Adv. Appl. Math 13*.

FERRARI, P., FRIGESSI, A., AND DE SÁ, P. 1995. Fast approximate maximum *a posteriori* restoration of multi-color images. *J. Roy. Stat. Soc. B 57*.

GEMAN, S., AND GEMAN, D. 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intell. 6*.

GREIG, D., PORTEOUS, B., AND SEHEULT, A. 1989. Exact maximum *a posteriori* estimation for binary images. *J. Roy. Stat. Soc. B 48*.

ISHIKAWA, H., AND GEIGER, D. 1998. Segmentation by grouping junctions. In *IEEE Conference on Computer Vision Pattern Recognition*. pp. 125–131.

KARGER, D. R., KLEIN, P., STEIN, C., THORUP, M., AND YOUNG, N. E. 1999. Rounding algorithms for a geometric embedding of multiway cut. In *Proceedings of the ACM Symposium on Theory of Computing*. ACM, New York.

KARZANOV, A. V. 1998. Minimum 0-extension of graph metrics. *Europ. J. Combinat. 19*.

KINDERMAN, R., AND SNELL, J. 1980. *Markov Random Fields and their Applications*. AMS Press.

LI, S. 1995. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag. New York.

PARDALOS, P., AND WOLKOWICZ, H. 1994. Quadratic assignment and related problems. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 16.

POTTS, R. 1952. Some generalized order-disorder transformations. *Proc. Cambridge Phil. Soc. 48*.

QUEYRANNE, M. 1986. Performance ratio of polynomial heuristics for triangle inequality quadratic assignment problems. *Oper. Res. Lett. 4*.

ROY, S., AND COX, I. 1998. A maximum-flow formulation of the *n*-camera stereo correspondence problem. In *Proceedings of the 6th International Conference on Computer Vision*.

WOODS, J. W. 1972. Two-dimensional discrete Markovian fields. *IEEE Trans. Inf. Theory 18*.

RECEIVED JULY 1999; REVISED AUGUST 2002; ACCEPTED AUGUST 2002