THEORY OF COMPUTING, Volume 7 (2011), pp. 49–74
www.theoryofcomputing.org

# Metric Clustering via Consistent Labeling*

Robert Krauthgamer† Tim Roughgarden‡

Received: November 19, 2009; published: March 26, 2011.

**Abstract:** We design approximation algorithms for a number of fundamental optimization problems in metric spaces, namely computing separating and padded decompositions, sparse covers, and metric triangulations. Our work is the first to emphasize *relative guarantees* that compare the produced solution to the optimal one for the input at hand. By contrast, the extensive previous work on these topics has sought *absolute* bounds that hold for every possible metric space (or for a family of metrics). While absolute bounds typically translate to relative ones, our algorithms provide significantly better relative guarantees, using a rather different algorithm.

Our technical approach is to cast a number of metric clustering problems that have been well studied—but almost always as disparate problems—into a common modeling and algorithmic framework, which we call the *consistent labeling* problem. Having identified the common features of all of these problems, we provide a family of linear programming relaxations and simple randomized rounding procedures that achieve provably good approximation guarantees.

**ACM Classification:** F.2.2

**AMS Classification:** 68Q17, 68W25, 90C59

**Key words and phrases:** approximation algorithms, metric spaces, decompositions

*A preliminary version appeared in SODA 2008 [24]. The current version includes some previously omitted proofs, some additional hardness results (Section 5), and various minor corrections.

†Supported in part by The Israel Science Foundation grant #452/08 and by a Minerva grant. This research was performed in part while at IBM Almaden.

‡Supported in part by an NSF CAREER Award CCF-0448664, an Alfred P. Sloan Fellowship, and an ONR Young Investigator Award. This research was performed in part while the author was visiting IBM Almaden.

 DOI: 10.4086/toc.2011.v007a005

# 1 Introduction

Metric spaces[1] arise naturally in a variety of computational settings, and are commonly used to model diverse data sets such as latencies between nodes in the Internet, dissimilarity between objects such as documents and images, and the cost of traveling between physical locations. Additionally, metric spaces are a useful technical tool, for example when analyzing algorithms based on a linear or semidefinite programming relaxation of SPARSEST CUT and other NP-hard problems.

Many useful computational tasks in metric spaces revolve around different types of clustering problems. In these problems, the goal is to produce, for a given metric space $(X,d)$, a collection $\mathcal{S}$ of subsets of $X$ such that, vaguely speaking, nearby points in $X$ tend to appear in the same subset.

This paper makes two broad contributions to the study of algorithms for metric clustering problems. First, we study a number of basic metric clustering problems from an optimization perspective, and design polynomial-time algorithms that provably achieve a near-optimal clustering for every metric space. The large literature on these metric clustering problems has focused exclusively on *absolute* (worst-case) bounds, seeking guarantees that hold for *every possible metric space* (or for every metric in a certain family). By contrast, we emphasize *relative guarantees*, where the objective is to compute a clustering that is close to optimal *for the given input*. Most absolute bounds translate easily to relative ones (in particular, they are efficiently computable), but our algorithms provide significantly better relative guarantees than those implied by the known absolute results. At a high level, our work can be viewed as a parallel to computing an optimal embedding of an input metric space into Euclidean space using semidefinite programming [30], or the recent line of research on computing embeddings with (approximately) minimum distortion, initiated by Kenyon, Rabani, and Sinclair [17]; for a recent account, see also [6, 33].

Why study relative guarantees? The quest for absolute bounds has obviously been very fruitful, but these bounds may not be very strong for a particular instance at hand, which may admit a much better solution than the worst-possible metric. A popular approach for eluding worst-case absolute bounds is to impose additional structure on the input metric, such as planarity or low-dimensionality, and then prove improved absolute bounds for that restricted class of metrics. But given an arbitrary distance matrix representing, say, latencies in the Internet, it may be highly non-trivial to ascertain whether the corresponding metric is close to one of these families. In contrast, an approximation algorithm guarantees a good solution provided only that one exists. Technically, this requires one to design a "unified" algorithm that works regardless of the precise reason the input admits an improved bound.

An approximation algorithm is also useful for inputs where the known absolute bounds are non-constructive. In this case, the approximation algorithm recovers, from the existential proof, an efficient algorithm that achieves nearly the same absolute guarantees. In a sense, this is true for planar metrics,[2] where, to date, no algorithm is known to efficiently determine whether an input metric is planar (or close to being planar). Consequently, the decomposition algorithm for planar metrics by Klein, Plotkin, and Rao [18] (see also [35, 11]) can only be applied if the planar metric is accompanied by a planar graph that realizes the metric. One immediate outcome of our approximation algorithms is that several results

[1] A metric space $(X,d)$ comprises a set $X$ of points and a distance function $d : X \times X \to R$ that is nonnegative and symmetric, and that satisfies the triangle inequality and the property that $d(x,y) = 0$ if and only if $x = y$.

[2] We call a metric *planar* if it can be derived from the shortest-path distances in a planar graph with nonnegative edge weights.

| Problem | Approximation factor | | Absolute guarantee |
|---|---|---|---|
| Separating Decomposition | 2 | [Theorem 3.4] | $O(\log n)$ [4] |
| Padded Decomposition | $O(1)$ bicriteria | [Theorem 3.9] | $O(\log n)$ [4] |
| Sparse Cover (stretch $k$) | $O(\log n)$ | [Corollary 4.2] | $2kn^{1/k}$ [3] |
| $(\varepsilon,\rho)$-Triangulation | $O(\ln \frac{1}{\varepsilon})$ bicriteria | [Corollary 4.5] | $n$ (trivial) |

Table 1: Our approximation factors and those implied by previous work on absolute bounds.

that rely on this decomposition, such as the low-distortion embedding into normed spaces of [35, 23], do not require a planar realization of the input metric and hold under the weaker assumption that a planar realization exists (or even that the input metric is close, by means of distortion, to a planar metric).[3]

Moreover, our algorithms are based on linear programming (LP) relaxations, and thus automatically generate a "certificate" of near-optimality (namely, the optimal fractional solution). These simple certificates could possibly be used to prove that a good solution does not exist (e. g., by bounding the optimal fractional solution using duality). Our relative guarantees prove that this lower bound approach is universal, in the sense that a near-optimal certificate always exists.

The second contribution of the paper is to cast a number of metric clustering problems that have been well studied—but almost always as disparate problems—into a common modeling and algorithmic framework, which we call the *Consistent Labeling* problem. At a high level, an instance of CONSISTENT LABELING is described by a set $A$ of objects, a list $L_a$ of allowable labels for each object $a \in A$, and a collection $\mathcal{C}$ of subsets of $A$. The goal is to assign each object few labels so that subsets are *consistent*, in the sense that the objects of a subset are all assigned a common label. The objects possessing a given label can be viewed as a "cluster" of objects (where clusters can overlap when we allow multiple labels per object), and the consistency constraint for a set $S \in \mathcal{C}$ requires that at least one cluster contains all of the objects of $S$ (i. e., there is at least one label common to all objects in $S$). In this paper, we show that many metric clustering problems are special cases of different variants of CONSISTENT LABELING. We then provide a family of LP relaxations for all of these problems, and design simple randomized rounding procedures that achieve provably good (relative) approximation guarantees.

We now detail the optimization problems for which we design approximation algorithms: these are two decomposition problems and two covering problems, as described in the sequel. Table 1 displays highlights of our results.

## 1.1 Metric decompositions

Let $(X,d)$ be a finite metric space on $n = |X|$ points. A *cluster* is a subset of the points $S \subseteq X$. The *ball* (in $X$) of radius $r \geq 0$ centered at $x \in X$ is $B(x,r) = \{y \in X : d(x,y) \leq r\}$. The *diameter* of a cluster $C$ is $\mathrm{diam}(C) = \max_{x,y \in C} d(x,y)$, and its *radius* is $\mathrm{rad}(C) = \min_{x_0 \in X} \max_{z \in C} d(x_0,z)$; a point $x_0$ attaining the radius is called a *center* of $C$.

[3] This argument applies more generally to excluded-minor graphs. The situation is similar also regarding the absolute guarantees shown in [28] for low-genus graphs and those shown in [8] for metrics that admit a low-distortion embedding into a low-dimensional Euclidean space.

Perhaps the simplest genre of metric clustering problems asks for a partition of $X$ into clusters of bounded radius while separating "few" points. We address the two fundamental variants of this notion: computing *separating decompositions* and *padded decompositions*. Both of these are central tools in metric embeddings (e. g., for designing probabilistic embeddings into trees [4, 10] and embeddings into $\ell_2$ [35, 23], respectively) and useful in algorithmic applications. Earlier incarnations of these concepts appeared, e. g., in [3, 29, 31, 13].

**Separating decomposition** Formally, a *decomposition* of $X$ is a probability distribution $\mu$ over partitions of $X$. Let $P$ be a partition of $X$; as mentioned above, we shall refer to the elements of $P$ as clusters. For $x \in X$, let $P(x)$ denote the cluster $S \in P$ that contains $x$, so $x \in S \in P$. We say that a partition *separates* two points $x, y \in X$ if it assigns them to distinct clusters. A partition is $\Delta$-*bounded* if each of its clusters has radius at most $\Delta$, and a decomposition is $\Delta$-*bounded* if every partition in its support is $\Delta$-bounded.[4] A $\Delta$-bounded decomposition $\mu$ is called $\alpha$-*separating* for $\alpha \geq 0$ if for all $x, y \in X$,

$$\Pr_{P \in \mu}[P(x) \neq P(y)] \leq \frac{\alpha \cdot d(x,y)}{\Delta} \tag{1.1}$$

or, equivalently,

$$\Pr_{P \in \mu}[P(x) = P(y)] \geq 1 - \frac{\alpha \cdot d(x,y)}{\Delta}\,. \tag{1.2}$$

We denote the minimum value $\alpha \geq 0$ satisfying (1.1) by $\alpha^*(X, \Delta)$. Bartal [4] designed an algorithm that achieves $\alpha = O(\log n)$ for every $n$-point input metric $X$, and showed that this bound is tight (i. e., the best possible in the worst case). Constant absolute bounds are known for planar metrics [18, 35, 11] and other restricted classes of metrics [8, 14, 22, 28].

Our first result is a 2-approximation algorithm for the problem of computing $\alpha^*(X, \Delta)$ (and constructing a corresponding decomposition). To see how this problem relates to the CONSISTENT LABELING problem, take both the object set and the label set to be the points $X$. The label set $L_x$ for a point $x \in X$ is defined to be the points in the ball $B(x, \Delta)$. We also impose the restriction that each point receives only one label. We can then interpret the set of vertices with a given label $z \in X$ as a cluster of radius at most $\Delta$ (centered at $z$), and these clusters form a partition of $X$. There is one consistency constraint for each pair of points; the constraint is satisfied if and only if the points are given the same label (i. e., assigned to the same cluster). The goal is to produce a distribution over feasible labelings such that the maximum probability of a set being labeled inconsistently (i. e., a pair $x, y \in X$ being separated by the partition), with suitable weighting by $1/d(x, y)$, is minimized.

**Remark 1.1.** Another application of the above approximation algorithm for computing separating decompositions was found in [5]: a constant-factor approximation algorithm for the problem of computing the least distortion embedding of an input metric into a distribution of dominating ultrametrics. This problem falls into the aforementioned category of computing an embedding with approximately minimum distortion [17, 6].

[4] Previous literature sometimes uses diameter instead of radius. Obviously the two quantities are within a factor of 2 of each other, and for us the radius is more convenient.

**Padded decomposition** Using the definitions above, a $\Delta$-bounded decomposition $\mu$ is $(\beta,q)$-*padded* for $\beta,q>0$ if for all $x \in X$,

$$\Pr_{P\in\mu}[B(x,\Delta/\beta)\subseteq P(x)] \geq q\,. \tag{1.3}$$

For a given $q$, we denote the smallest $\beta>0$ satisfying (1.3) by $\beta^*(X,\Delta,q)$. We can model computing a padded decomposition as a CONSISTENT LABELING problem in the same way as for a separating decomposition, except that now the collection $\mathcal{C}$ of consistency sets is not all pairs of points, but rather all balls of radius $\Delta/\beta$.

Computing near-optimal padded decompositions appears to be technically harder than separating decompositions, but using a more sophisticated rounding algorithm we can compute a $\Delta$-bounded decomposition that is $(2\beta^*,q/12)$-padded, where $\beta^*=\beta^*(X,\Delta,q)$. This bicriteria guarantee is often as useful for applications as a true approximation; in fact, in the aforementioned applications of padded decompositions, the parameter $q$ is fixed to an arbitrary constant such as $1/2$, and relaxing it to $q/12$ is as good as any other positive constant.

The problem of computing a near-optimal padded decomposition has not been studied previously, and the absolute guarantees yield, at best, an $O(\log n)$-approximation (recall $n=|X|$). Also, while there is a relationship between padded and separating decompositions of the form $\alpha^*(X,\Delta)\leq 4\beta^*(X,\Delta/2,1/2)$ [27], in general the two quantities can be very different; e. g., in $m$-dimensional Euclidean space, $\alpha^*=\Theta(\sqrt{m})$ [8] and $\beta^*=\Theta(m)$ [27, Section 2.1].

## 1.2 Covering problems

*Covering problems* form a second genre of metric clustering problems, where the goal is to minimize the overlap between clusters subject to some type of covering constraint. We focus on the following two such problems.

**Sparse cover** Consider an undirected graph $G=(V,E)$ with positive edge lengths and a list $C_1,\dots,C_p$ of subsets of nodes. The graph vertices naturally represent points in a metric space, namely, $n=|V|$ points with distances corresponding to shortest-path lengths in $G$, and thus the terminology from Section 1.1 extends to the current scenario (e. g., a cluster is a subset of $V$). We restrict our discussion to the case $p=n$, which includes the typical case where the subsets $C_i$ correspond to balls around the vertices. A *sparse cover* [3] is a $\Delta$-bounded collection $\mathcal{S}$ of clusters, such that every subset $C_i$ is contained in some cluster of $\mathcal{S}$. The *degree* of a vertex $v$ in $\mathcal{S}$ is the number of clusters of $\mathcal{S}$ that contain $v$. Awerbuch and Peleg [3] use sparse covers as a building block for a number of distributed network algorithms, including a routing scheme with low stretch[5] and small storage at every node. Specifically, the stretch of the routing scheme in [3] is proportional to $\Delta/\max_i \operatorname{rad}(C_i)$, and the maximum degree in the sparse cover determines the nodes' storage requirements. Awerbuch and Peleg [3] give absolute bounds for computing a sparse cover: for each integer $k\geq 1$, they show how to construct a sparse cover with $\Delta/\max_i \operatorname{rad}(C_i)\leq k$ and maximum degree at most $2kn^{1/k}$. This immediately implies a similar relative guarantee of $2kn^{1/k}$ on the maximum degree.

[5] The *stretch* of a routing scheme is the largest factor by which the length of an employed routing path exceeds that of a shortest path between the same source and destination.

We study the metric variant of sparse covers, when $G$ is a complete graph representing a metric space. This variant is essentially the same as the *Nagata dimension* of a metric space (see [2, 25]).[6]

We model the problem of computing a sparse cover with minimum maximum degree (where $\Delta > 0$ and the subsets $C_i$ are given as part of the input) as a CONSISTENT LABELING problem and give an $O(\log n)$-approximation algorithm. In the CONSISTENT LABELING formulation, both objects and labels correspond to the vertices $V$, and a label can only be assigned to an object if they correspond to two vertices at distance at most $\Delta$ in $G$. A $\Delta$-bounded collection of clusters induces a feasible labeling, and the degree of a vertex in the clustering is precisely the number of labels the vertex is assigned. Finally, the constraint of containing a set $C_i$ in at least one cluster naturally translates to a consistency constraint for the subset $C_i$, and conversely a feasible labeling induces a sparse cover. Computing a sparse cover with minimum maximum degree thus translates to computing a feasible labeling that labels all the sets consistently while minimizing the maximum number of labels allowed at an object.

**Metric triangulation** Finally, we consider computing metric triangulations of small order [15, 19]. Network triangulation is a heuristic for estimating distances in a network, initially suggested by Guyton and Schwartz [15]. Motivated by the practical success of this heuristic, Kleinberg, Slivkins, and Wexler [19] initiated a theoretical study of triangulation in metric spaces, formally defined as follows. A *triangulation* of a metric $(X,d)$ assigns to every $x \in X$ a collection of *beacons* $S_x \subseteq X$. The triangulation has *order* $k$ if $\max\{|S_x| : x \in X\} \le k$. We are interested in low-order triangulations in which the distance between every $x,y \in X$ can be estimated from their distances to $S_x \cap S_y$ using the triangle inequality. Formally, define

$$D^+(x,y) = \min_{b \in S_x \cap S_y} [d(x,b) + d(b,y)]$$
$$D^-(x,y) = \max_{b \in S_x \cap S_y} |d(x,b) - d(b,y)|\,.$$

The triangulation is called an $(\varepsilon,\rho)$*-triangulation* (for $0 \le \varepsilon \le 1$ and $\rho \ge 1$) if for all but an $\varepsilon$-fraction of the pairs $x,y \in X$ we have $D^+(x,y) \le \rho \cdot d(x,y)$ and $D^-(x,y) \ge d(x,y)/\rho$. Let $k_{\mathrm{opt}}(X,\varepsilon,\rho)$ denote the smallest $k > 0$ such that $(X,d)$ admits an $(\varepsilon,\rho)$-triangulation of order $k$.

The problem of computing a near-optimal metric triangulation—that is, computing $k_{\mathrm{opt}}(X,\varepsilon,\rho)$—has not been studied before, although several absolute guarantees are known. In [19], it is shown that doubling metrics admit an $(\varepsilon,\rho)$-triangulation of constant order (the upper bound depends only on $\varepsilon,\rho$ and the doubling constant), and additional bounds are proved in [37, 38]. However, in some metrics triangulation requires a very high order (e. g., $\Omega(n)$ in uniform metrics and $n^{\Omega(1)}$ in tree metrics [21], for fixed $\varepsilon,\rho$), and thus absolute bounds cannot yield any nontrivial approximation ratio. While this problem is quite different from the SPARSE COVER application discussed above, we formulate and approximate both in a common way. In particular, our techniques immediately yield good bicriteria approximation algorithms for minimizing the order of a triangulation subject to being able to estimate almost all pairwise distances.

[6] Our variant measures all distances in the metric space and corresponds to the so-called *weak diameter* bound. In the context of a graph with shortest-path distances, the construction of [3] satisfies the more stringent *strong diameter* bound, where distances inside a cluster are determined by shortest paths in the induced subgraph.

### 1.3 Overview and techniques

At a high level, our algorithms follow the well-known paradigm of solving a linear programming relaxation of the problem and applying randomized rounding. We thus start, in Section 2, by formulating LP relaxations for several variants of the CONSISTENT LABELING problem.

Section 3 gives approximation algorithms for the problems of computing separating and padded decompositions. We model them as special cases of a maximization version of the CONSISTENT LABELING problem, where the goal is to maximize the fraction of consistent sets while obeying an upper bound on the number of labels assigned to every object. To round our linear programming relaxations (given in Section 2) in a "coordinated" way that encourages consistently labeled sets, we build on a rounding procedure of Kleinberg and Tardos [20]. This procedure was designed for the metric labeling problem with the uniform label-metric (which, in turn, is a modification of the multiway cut algorithm of Calinescu, Karloff, and Rabani [7]). The differences between our intended applications and the metric labeling problem necessitate extensions to their analysis; for example, we require guarantees for maximization rather than minimization problems, and for general set systems rather than for pairs of points (i. e., hypergraphs instead of graphs). Our extensions to the Kleinberg-Tardos rounding algorithm and analysis lead, for example, to a 2-approximation algorithm for the separating decomposition problem. The padded decomposition problem is significantly more challenging, and requires us to enhance this basic rounding algorithm in two ways: first, we limit the number of rounding phases to control the proliferation of different labels; second, we add two postprocessing steps that first weed out some problematic labels and then expand the residual clusters to ensure the padding properties.

Section 4 gives a family of approximation algorithms that approximate, in particular, the sparse cover and metric triangulation problems. Our algorithm and analysis techniques are essentially "dualized" versions of those used earlier for the maximization versions of CONSISTENT LABELING.

**Remark 1.2.** In some of the problems we study, the goal is to produce a probability distribution over labelings (or partitions). We permit a solution in the form of an algorithm that is randomized and reports one (random) labeling; the distribution over labelings is the algorithm's output. If an explicit probability distribution is desired, it can be obtained (with a minor loss) by sampling the randomized algorithm sufficiently many times and applying standard concentration arguments.

## 2 Linear programming relaxations for CONSISTENT LABELING

Motivated by the breadth of applications in the Introduction, we examine several variants of the CONSISTENT LABELING problem. This section formally defines these variants and gives a family of linear programming relaxations for them. We often omit the straightforward proofs that they are in fact relaxations.

### 2.1 Common ingredients

In all cases, the input includes a set $A$ of objects, a set $L_a$ of allowable labels for each object $a$ (drawn from a ground set $L$), and a collection $\mathcal{C}$ of subsets of $A$. In some applications, we also allow each set $S \in \mathcal{C}$ to have a nonnegative weight $w_S$. A *feasible labeling* assigns to every object $a$ some subset of $L_a$.

Our two main objectives are to minimize the number of labels assigned to each object, and to maximize the number (or total weight) of sets that are consistently labeled, meaning that a common label is assigned to all of the objects in the set.

The following variables and constraints are common to all our relaxations. The variable $x_{ai}$ represents the assignment of label $i \in L$ to object $a \in A$; intuitively, it is an indicator (taking values in $\{0,1\}$), but in some of our problems a fractional value is also admissible. Constraint (2.1) below then controls the number of (fractional) labels assigned to each object. In some applications, $k$ will be a decision variable; in others, it will be part of the problem input. The variable $y_{iS}$ encodes the extent to which set $S$ is consistently labeled with the label $i$, giving rise to the constraint (2.2) below. The variable $z_S$ encodes the extent to which set $S$ is (fractionally) consistently labeled, giving rise to constraints (2.3) and (2.4) below. The fifth constraint below enforces the restriction that objects are assigned only to allowed labels.

$$1 \le \sum_{i \in L} x_{ai} \le k \quad \text{for every object } a \in A\,; \tag{2.1}$$

$$y_{iS} \le x_{ai} \quad \text{for every set } S \in \mathcal{C}\text{, label } i \in L\text{, and object } a \in S\,; \tag{2.2}$$

$$z_S \le \sum_{i \in L} y_{iS} \quad \text{for every set } S \in \mathcal{C}\,; \tag{2.3}$$

$$z_S \le 1 \quad \text{for every set } S \in \mathcal{C}\,; \tag{2.4}$$

$$x_{ai} = 0 \quad \text{for every object } a \in A \text{ and label } i \notin L_a\,. \tag{2.5}$$

We always assume that all LP variables are nonnegative; this applies in particular to each variable of the form $x_{ai}$, $y_{iS}$, and $z_S$.

## 2.2 Maximization version

In the *Maximum Consistent Labeling (*MAX CL*)* problem, the objective is to compute a feasible labeling that assigns at most $k$ labels to every object ($k$ is part of the input) and maximizes the total weight of the consistently labeled sets. Our LP relaxation for MAX CL is to optimize

$$\max \sum_{S \in \mathcal{C}} w_S z_S \tag{2.6}$$

subject to (2.1)–(2.5).

Padded and separating decompositions motivate the *Maximum Fair Consistent Labeling (*MAX FAIR CL*)* problem, where given an input as in MAX CL, the goal is to compute a *distribution* over feasible labelings that assign at most $k$ labels to every object (with probability 1) and maximizes the minimum weighted probability (over $S \in \mathcal{C}$) that a set $S$ is labeled consistently. Computing both separating and padded decompositions are special cases of MAX FAIR CL with $k = 1$, where the sets correspond to pairs of points, and to balls of radius $\Delta/\beta$ around each point in the given metric space, respectively. Our LP relaxation for this problem maximizes a decision variable $\alpha$ subject to (2.1)–(2.5) and

$$w_S z_S \ge \alpha \qquad \text{for every set } S \in \mathcal{C}\,. \tag{2.7}$$

### 2.3 Minimization version

In the minimization version of consistent labeling, we constrain (from below) the fraction of consistently labeled sets and seek a labeling that uses as few labels per object as possible. (We could also include set weights, but our applications do not require them.) We call this problem the *Minimum Consistent Labeling (*MIN CL*)* problem.

In the *complete* special case, we demand that all sets are consistently labeled. Formally, the *Minimum Complete Consistent Labeling (*MIN CCL*)* problem is, given the usual data, to compute a feasible labeling that consistently labels all sets and minimizes the maximum number of labels assigned to an object. In our LP relaxation for MIN CCL, we minimize the decision variable $k$ subject to (2.1)–(2.5) and the additional constraint that (2.4) holds with equality for every set $S \in \mathcal{C}$.

As noted in the Introduction, computing a sparse cover of a network is a special case of MIN CCL. Several extensions to the MIN CCL problem are easily accommodated; we use METRIC TRIANGULATION as a case study in Section 4.

Before proceeding to our approximation algorithms, we note that the MAX CL, MAX FAIR CL, and MIN CCL problems are all APX-hard (see Section 5 for details).

## 3 Maximum consistent labeling

This section gives a generic approximation algorithm for the MAX CL and MAX FAIR CL problems. We then refine the algorithm and its analysis to give an approximation algorithm for computing a separating decomposition (Theorem 3.4). Subsequently, we enhance the algorithm to handle the more difficult task of approximating an optimal padded decomposition (Theorem 3.9). We remark that the authors of [26] study approximation algorithms for a different problem of maximizing consistencies, which is closer in spirit to MAX $k$-CUT.

We remark that a polynomial-time algorithm for MAX CL implies, by duality and the ellipsoid method, a polynomial-time algorithm for MAX FAIR CL. We do not use this approach here, since a feature of our algorithm is that it handles both problems directly.

### 3.1 Approximation algorithm for MAX CL and MAX FAIR CL

We first give a $\Theta(1/f_{\max})$-approximation algorithm for weighted MAX CL and MAX FAIR CL, where $f_{\max} = \max_{S\in\mathcal{C}} |S|$ denotes the largest cardinality of a set of $\mathcal{C}$. We build on a rounding procedure that was designed by Kleinberg and Tardos [20] for the metric labeling problem with the uniform metric, even though our context is quite different. First, we wish to maximize the probability of consistency, as in (1.2), rather than minimize the probability of inconsistency, as in (1.1). Second, an object may get multiple labels ($k$) rather than one label ($k = 1$). Third, the notion of consistency is not as simple, as it involves a subset $S$ (whose size may be bigger than 2) and each object in $S$ has $k$ labels (where $k$ may be bigger than 1). Fourth, we may want to produce a distribution (in MAX FAIR CL) rather than only one solution. It is thus a pleasant surprise that the algorithm in [20] lends itself to our setting; in fact, our algorithm can be easily seen to generalize theirs from $k = 1$ labels to general $k$.

Our randomized approximation algorithm is shown in Figure 1. After solving the appropriate LP relaxation, the rounding algorithm is the same for both MAX CL and MAX FAIR CL: we repeatedly

**Input: an instance of MAX CL or MAX FAIR CL.**
1. Solve the appropriate LP relaxation: for MAX CL, maximize (2.6) subject to (2.1)–(2.5); for MAX FAIR CL, maximize $\alpha$ subject to (2.1)–(2.5) and (2.7). Let $(x^*, y^*, z^*)$ denote the optimal LP solution.
2. Repeat until every object has been assigned at least $k$ labels (counting multiplicities):
3. Choose a label $i \in L$ and a threshold $t \in [0,1]$ uniformly at random.
4. For each object $a \in A$, if $x^*_{ai} > t$, then add $i$ to the set of labels assigned to $a$.
5. Output for each object the first $k$ labels it received.

Figure 1: The MAX CL and MAX FAIR CL algorithms.

choose a label $i \in L$ and a threshold $t \in [0,1]$ independently and uniformly at random, and for all objects $a$ with $x^*_{ai}$ larger than the threshold $t$, we add $i$ to the set of labels assigned to $a$. (If $i$ is already assigned to $a$, then this assignment is redundant.) The algorithm terminates when every object has been assigned a label in at least $k$ iterations (not necessarily distinct labels). To respect the constraint on the number of labels, each object retains only the first $k$ labels that it was assigned. This final step, together with the LP constraint (2.5), ensures that the output of the algorithm is a feasible labeling.

Our analysis hinges on the following lemma, which lower bounds the probability that a set is consistently labeled by our rounding algorithm. We also use the lemma in Section 4 for minimization versions of CONSISTENT LABELING.

**Lemma 3.1.** *Consider an execution of the algorithm of Figure 1. For every set $S \in \mathcal{C}$,*

$$\Pr[S \textit{ consistently labeled}] \geq 1 - \left(1 - \frac{z^*_S}{k|S|}\right)^k \geq \frac{z^*_S}{2|S|}\,.$$

*Proof.* Fix a set $S \in \mathcal{C}$. For each label $i \in L$, let $x_i^{\max}$ and $x_i^{\min}$ denote $\max_{a\in S} x^*_{ai}$ and $\min_{a\in S} x^*_{ai}$, respectively. Let $F \subseteq L$ denote the set of labels $i$ for which $x_i^{\max} > 0$.

Now fix an iteration. Call the iteration *active* if at least one object of $S$ receives a (possibly redundant) label. In an active iteration, the conditional probability that the label $i$ was chosen is $x_i^{\max}/\sum_{j\in L} x_j^{\max}$. Let $\mathcal{E}_S$ denote the event that $S$ is consistently labeled (not necessarily for the first time) in this iteration. Thus,

$$\begin{aligned}
\Pr[\mathcal{E}_S \mid \text{active}] &= \sum_{i\in F} \Pr[\mathcal{E}_S \mid \text{active, label=}i] \cdot \Pr[\text{label=}i \mid \text{active}] \\
&= \sum_{i\in F} \left(\frac{x_i^{\min}}{x_i^{\max}}\right) \cdot \left(\frac{x_i^{\max}}{\sum_{j\in F} x_j^{\max}}\right) \\
&= \frac{1}{\sum_{j\in F} x_j^{\max}} \sum_{i\in F} x_i^{\min} \\
&\geq \frac{1}{k|S|} \sum_{i\in L} y^*_{iS} \qquad (3.1) \\
&\geq \frac{z^*_S}{k|S|}\,, \qquad (3.2)
\end{aligned}$$

where inequality (3.1) follows from the LP constraints (2.1) and (2.2), and inequality (3.2) follows from the LP constraint (2.3).

Since the iterations are independent, the first inequality in the lemma follows by considering the first $k$ iterations that are active (note there are indeed at least $k$). The second inequality is derived by applying the (crude) inequality

$$\left(1-\frac{z}{k}\right)^k \le e^{-z} \le 1-z+\frac{z^2}{2} \le 1-\frac{z}{2}$$

for $z \in (0,1)$. □

Using this lemma and linearity of expectation, we immediately obtain the approximation bounds for the MAX CL and MAX FAIR CL problems.

**Theorem 3.2.** *There are randomized polynomial-time* $(1/2f_{\max})$*-approximation algorithms for weighted* MAX CL *and* MAX FAIR CL.

The bound $1/2f_{\max}$ in Theorem 3.2 can be sharpened; for example, it is $1/f_{\max}$ when $k=1$.

Theorem 3.2 does not immediately give a useful approximation algorithm for computing separating or padded decompositions; we next give the necessary refinements.

## 3.2 Separating decomposition

Theorem 3.2 gives an approximation guarantee for the maximum consistency probability (as in (1.2)), rather than for the minimum inconsistency probability (as in (1.1)). These two objectives are equivalent for exact optimization, but not for approximation. We now show how to modify our LP relaxation and analysis for MAX FAIR CL (but using the same rounding algorithm), to obtain an $f_{\max}$-approximation for the objective (1.1). Choosing the weight $w_S$ of a set $S=\{x,y\}$ to be $\Delta/d(x,y)$, we immediately get a 2-approximation algorithm for computing an optimal separating decomposition, which matches the integrality gap for our LP relaxation. The precise statements appear in Theorems 3.4 and 3.5.

We address the problem of minimizing the inconsistency probability using the LP (3.3) below. This LP differs from the one used for MAX FAIR CL in that we fix $k=1$; that $y_{iS}$ represents the probability of an inconsistency involving label $i$; $z_S$ represents the probability that $S$ is inconsistently labeled; and we bound the $z_S$'s from above (rather than from below) using $\alpha$.

$$\begin{array}{lll}
\text{Min} & \alpha & \\
\text{s.t.} & \sum_{i\in L} x_{ai} = 1 & \forall a \in A \\
 & y_{iS} \ge x_{ai} - x_{a'i} & \forall S \in \mathcal{C};\ a,a' \in S \\
 & z_S \ge \frac{1}{|S|}\sum_{i\in L} y_{iS} & \forall S \in \mathcal{C} \\
 & x_{ai} = 0 & \forall a \in A;\ i \notin L_a \\
 & \alpha \ge w_S z_S & \forall S \in \mathcal{C}.
\end{array} \tag{3.3}$$

It is straightforward to verify that this LP is indeed a relaxation for the problem of minimizing inconsistencies. Given a distribution over partitions, one defines $x_{ai}$ as the probability that $a$ receives label $i$; $y_{iS}$ as the probability that a set $S$ has at least one but not all elements labeled $i$ (which is at least $\max_a x_{ai} - \min_a x_{ai}$); and $z_S$ as the probability that $S$ is inconsistently labeled. When $S$ is inconsistently labeled it involves at most $|S|$ distinct labels, which justifies the inequality $|S|z_S \geq \sum_{i\in L} y_{iS}$. This LP can be viewed as a generalization of the Kleinberg-Tardos relaxation from the case $|S| = 2$ to general sets $S$.

Let $(x^*, y^*, z^*, \alpha^*)$ be the optimal fractional solution to LP (3.3); then $\alpha^*$ is a lower bound on the value of an optimal solution. We now apply to this LP solution the rounding algorithm of Figure 1.

**Lemma 3.3.** *When executing the algorithm of Figure 1 on the solution of LP (3.3), for every set $S \in \mathcal{C}$,*

$$\Pr[S \textit{ is not consistently labeled}] \leq |S|z_S^*.$$

*Proof.* Fix a set $S \in \mathcal{C}$. For each label $i \in L$, let $x_i^{\max}$ and $x_i^{\min}$ denote $\max_{a\in S} x_{ai}^*$ and $\min_{a\in S} x_{ai}^*$, respectively. Let $F \subseteq L$ denote the set of labels for which $x_i^{\max} > 0$. For a given iteration, denote by $\mathcal{E}_S$ the event that $S$ is *not* consistently labeled. Consider henceforth the first iteration in which some object of $S$ receives a label. We then have (conditioned on this event):

$$\begin{aligned}
\Pr[\mathcal{E}_S] &= \sum_{i\in F} \Pr[\mathcal{E}_S \,|\, \text{label=}i] \cdot \Pr[\text{label=}i] \\
&= \sum_{i\in F} \left(\frac{x_i^{\max} - x_i^{\min}}{x_i^{\max}}\right) \cdot \left(\frac{x_i^{\max}}{\sum_{j\in F} x_j^{\max}}\right) \\
&= \frac{1}{\sum_{j\in F} x_j^{\max}} \sum_{i\in F} (x_i^{\max} - x_i^{\min}).
\end{aligned}$$

By the first LP constraint, the denominator must be at least 1, and by the second LP constraint, each summand is $x_i^{\max} - x_i^{\min} \leq y_{iS}^*$. ¿From these, together with the third LP constraint, we conclude that $\Pr[\mathcal{E}_S] \leq \sum_{i\in L} y_{iS}^* \leq |S|z_S^*$. □

This inequality immediately implies an $f_{\max}$-approximation for minimizing the (weighted) inconsistency probability of all sets. In particular, we obtain the following theorem.

**Theorem 3.4.** *There is a randomized polynomial-time 2-approximation algorithm for computing a separating decomposition.*

Our next result shows that no better approximation ratio is possible using our linear programming relaxation as a lower bound.

**Theorem 3.5.** *The LP relaxation (3.3) has integrality gap arbitrarily close to 2, even in the special case of computing a separating decomposition.*

*Proof.* Let the radius bound be $\Delta = 1$, and let $(X, d)$ be an $n$-point metric where the pairwise distances equal 2 along one specific perfect matching between the points, and equal 1 otherwise. Formally, let $X = \{1, 2, \ldots, n\}$, where $n > 2$ is even, and for each $j \neq j'$ let $d(j, j') = 2$ if $|j - j'| = n/2$ and $d(j, j') = 1$ otherwise.

We claim that the optimal value of the LP is $\alpha \leq 1/(n-1)$. Indeed, assign each object (point) $j$ equally in a fractional sense to all $n-1$ labels (points) $j'$ with $d(j, j') \leq 1$, i.e., $x_{jj'} = 1/(n-1)$. For each $i$ and $S$ set accordingly $y_{iS} = \max_{a\in S} x_{ai} - \min_{a\in S} x_{ai}$, and for each $S$ set $z_S = (\sum_{i\in L} y_{iS})/2$. It is easy to verify that for every $S = \{j, j'\} \in \mathcal{C}$ we have $z_S = 1/(n-1)$, and recall $w_S = 1/d(j, j') \leq 1$.

Consider now a $\Delta$-bounded $\alpha^*$-separating decomposition, and let us give a lower bound on $\alpha^*$. Let $P$ be a random partition drawn from this decomposition, and define the following random variable:

$$Z = \sum_{j=1}^{n} 1_{\{P(j) \neq P(j+1)\}} ,$$

where point $n+1$ is understood to be point 1. On the one hand, linearity of expectation implies that $\mathbb{E}[Z] \leq n\alpha^*$. On the other hand, with probability 1 we have $Z \geq 2$, since a $\Delta$-bounded partition $P$ must contain at least two clusters. Thus $\alpha^* \geq \mathbb{E}[Z]/n \geq 2/n$, proving that the integrality ratio is at least $2(n-1)/n = 2 - 1/n$. □

### 3.3 Padded decomposition

Building on our previous techniques, we now design an algorithm for computing a padded decomposition; the precise statement of the guarantees appears in Theorem 3.9. Recall that the input is a metric space $(X,d)$ and a parameter $q > 0$. The following LP formulation is similar to the one we used for the MAX FAIR CL problem:

$$\boxed{\begin{array}{ll} \sum_{i \in X} x_{ij} \leq 1 & \forall j \in X \\ y_{ij} \leq x_{ij'} & \forall i,j \in X;\ j' \in B(j,\Delta/\beta) \\ x_{ij} = 0 & \forall j \in X;\ i \in X \setminus B(j,\Delta) \\ \sum_{i \in X} y_{ij} \geq q & \forall j \in X\,. \end{array}} \tag{3.4}$$

Here, objects correspond to points in $X$, labels represent cluster centers (all points of $X$), the allowed labels for an object are those within distance $\Delta$, and the consistency sets $\mathcal{C}$ corresponds to all balls of radius $\Delta/\beta$. This LP has nonnegative variables $x_{ij}$, which represent an assignment of a point $j \in X$ to a cluster centered at $i \in X$ (i. e., labeling an object), and variables $y_{ij}$, which represent the consistency of the ball around $j$ with respect to the cluster represented by center $i$ (i. e., consistency of a set). Notice that $j \in X$ has two roles (simultaneously), of an object and of a consistency set.

**Lemma 3.6.** *The linear program (3.4) is a relaxation of the padded decomposition problem—that is, it is feasible provided $\beta \geq \beta^*(X,\Delta,q)$.*

*Proof.* Whenever $\beta \geq \beta^*(X,\Delta,q)$ there exists a $\Delta$-bounded $(q,\beta)$-padded decomposition $\mu$. We first construct an LP solution from a single such partition $P$ in the support of $\mu$: For every cluster $S \in P$, designate a center point that attains the radius bound. Set $x_{ij} = 1$ if $i$ is the designated center of $P(j)$; otherwise, set $x_{ij} = 0$. Now set $y_{ij} = 1$ if both $x_{ij} = 1$ and $j$ is $\Delta/\beta$-padded in its cluster (formally, $B(j,\Delta/\beta) \subseteq P(j)$); otherwise, set $y_{ij} = 0$. This solution clearly satisfies the first and third constraints. To see that the second constraint is satisfied, it suffices to consider the case when $y_{ij} = 1$ (otherwise it is trivial). This means that $j$ is padded in $P$, so all nearby points $j'$ belong to the same cluster, and thus $x_{ij'} = 1$. For the moment we ignore the last constraint, which might be unsatisfied, since $\sum_{i \in X} y_{ij}$ is 1 if $j$ is padded in $P$ and is 0 otherwise.

Now take a convex combination of the solutions constructed for the different $P \in \mathrm{supp}(\mu)$, weighted by their probabilities (according to $\mu$). This solution still satisfies the first three constraints. The fourth constraint is now satisfied because in the decomposition $\mu$, every point $j \in X$ is padded with probability at least $q$. □

**Input: an instance of padded decomposition**
1. Find the smallest $\beta > 0$ such that LP (3.4) is feasible. Let $(x^*, y^*)$ denote a feasible LP solution.
2. Initialize a cluster $C_i = \emptyset$ for every $i \in X$.
3. Repeat $n$ times
4. Choose uniformly at random $i \in X$ and a threshold $t \in [0,1]$.
5. Add to cluster $C_i$ every unclustered point $j \in X$ for which $t < y^*_{ij}$.
6. Let $D^* = \{j \in X : B(j, \Delta/\beta) \text{ meets more than one cluster } C_i\}$.
7. For every $i \in X$, let $C'_i = C_i \setminus D^*$.
8. For every $i \in X$, let $C''_i = \{j \in X : d(j, C'_i) \le \Delta/2\beta\}$.
9. Output the partition induced by $\{C''_i : i \in X\}$, using singleton clusters as needed.

Figure 2: The Padded Decomposition algorithm.

Our algorithm's first step is to find the smallest $\beta > 0$ such that the LP (3.4) is feasible, which can be done via binary search over the $\binom{n}{2}$ distance values appearing in the input metric. (Note that $\beta$ is not a variable of the LP.)

The rounding procedure for LP (3.4) has three steps (see Figure 2). First, we use a procedure similar to that in the algorithm of Figure 1, except that exactly $n$ assignment rounds are performed to obtain a collection of disjoint clusters $\{C_i : i \in X\}$. This need not be a partition, since some points might not be assigned at all. Notice that this procedure uses the $y$-variables rather than the $x$-variables. Next, we check for which points $j \in X$ the ball $B(j, \Delta/\beta)$ meets more than one cluster $C_i$, and remove all these points (simultaneously) from the clustering. Finally, we expand each of the (non-empty) clusters remaining to its $\Delta/2\beta$-neighborhood, and output the partition induced by these clusters (points that belong to no cluster form singleton clusters).

We analyze the performance of this algorithm in the next two lemmas.

**Lemma 3.7.** *The algorithm in Figure 2 always outputs a $\Delta$-bounded partition of $X$.*

*Proof.* To prove that the produced clustering is indeed a partition, it suffices to verify that the clusters $\{C''_i : i \in X\}$ are disjoint. Assume for contradiction that some $j \in X$ belongs to two such clusters, $C''_i$ and $C''_h$. Then by the definition in step 8, $j$ is at distance at most $\Delta/2\beta$ from a point in $C'_i$ and a point in $C'_h$. But then these two points should have been included in $D^*$, contradicting their inclusions in $C'_i$ and $C'_h$.

To prove the radius bound, consider $j \in C''_i$. Then there is a $j' \in C'_i \subseteq C_i$ with $d(j, j') \le \Delta/2\beta$. By step 5, $j' \in C_i$ satisfies $y^*_{ij'} > 0$. By the second LP constraint $x^*_{ij} \ge y^*_{ij'} > 0$, implying by the third LP constraint $d(i, j) \le \Delta$, which completes the proof. □

**Lemma 3.8.** *Let $P$ denote the partition output by the algorithm in Figure 2. Then for every $j \in X$,*

$$\Pr_P[B(j, \Delta/2\beta) \subseteq P(j)] \ge q/12\,.$$

*Proof.* Fix a point $j \in X$. Observe that once $j \in C'_i$, the entire ball $B(j, \Delta/2\beta)$ will end up inside the cluster $C''_i$. Thus,

$$\Pr[B(j, \Delta/2\beta) \subseteq P(j)] \ge \Pr[j \in \cup_{i \in X} C'_i] = \sum_{i \in X} \Pr[j \in C'_i]\,, \tag{3.5}$$

where the equality is due to the fact that the clusters $\{C'_i\}_{i\in X}$, and thus also the respective events, are disjoint. We next examine the $n$ iterations over steps 4–5, and refine our earlier analysis of the randomized assignment procedure.

Fix now also $i^* \in X$. For the event $j \in C'_{i^*}$ to occur, we must have that both $j \in C_{i^*}$ and $j \notin D^*$; the latter means that $B(j,\Delta/\beta)$ is the disjoint union $\bigcup_{i\neq i^*} C_i$. For the purpose of a lower bound on $\Pr[j \in C'_{i^*}]$, it suffices to consider the case that $i^* \in X$ is chosen (in step 4) in exactly one of the $n$ iterations, which happens with probability

$$\binom{n}{1}\frac{1}{n}\Big(1-\frac{1}{n}\Big)^{n-1} \geq \frac{1}{e}\,.$$

Assuming this is the case, in the iteration in which $i^*$ is the chosen center, we need it to "capture" point $j$, which happens with probability $\Pr_t[t < y^*_{i^*j}] = y^*_{i^*j}$. In each of the other $n-1$ iterations, we need the chosen center $i \neq i^*$ to capture no point in $B(j,\Delta/\beta)$, which happens with probability

$$1 - \max\{y^*_{ij'} :\ j' \in B(j,\Delta/\beta)\} \geq 1 - x^*_{ij}\,,$$

where the inequality is by the second constraint of LP (3.4). Recalling that $i \neq i^*$ is chosen uniformly at random from $n-1$ values, we obtain (for our fixed $i^*$)

$$\Pr[j \in C'_{i^*}] \geq \frac{1}{e}\cdot y^*_{i^*j}\cdot\Big(\sum_{i\neq i^*}\frac{1-x^*_{ij}}{n-1}\Big)^{n-1} = \frac{1}{e}\cdot y^*_{i^*j}\cdot\Big(1-\frac{\sum_{i\neq i^*} x^*_{ij}}{n-1}\Big)^{n-1}\,.$$

The first LP constraint enforces $\sum_{i\neq i^*} x_{ij} \leq 1$, and we obtain (assuming $n \geq 3$)

$$\Pr[j \in C'_{i^*}] \geq \frac{1}{e}\cdot y_{i^*j}\cdot\Big(1-\frac{1}{n-1}\Big)^{n-1} \geq \frac{1}{4e}\cdot y_{i^*j}\,.$$

Finally, plugging the last inequality into (3.5) and then using the last constraint of the LP, we conclude that

$$\Pr[B(j,\Delta/2\beta) \subseteq P(j)] \geq \sum_{i^*\in X}\Pr[j \in C'_{i^*}] \geq \sum_{i^*\in X}\left(\frac{1}{4e}\cdot y_{i^*j}\right) \geq \frac{q}{12}\,.$$

□

The two lemmas above immediately yield the following.

**Theorem 3.9.** *There is a randomized polynomial-time algorithm that, given a metric $(X,d)$ and $\Delta, q > 0$, produces a $\Delta$-bounded $(\beta', q/12)$-padded decomposition, where $\beta' \leq 2\beta^*(X,\Delta,q)$.*

# 4 Minimum consistent labeling

This section gives two approximation algorithms for the minimization version of Consistent Labeling, where the goal is to consistently label a prescribed fraction of the sets while using as few labels as possible. The first algorithm is tailored to the MIN CCL problem, where all of the sets must be consistently labeled. Our algorithm achieves an $O(\log(|A|+|\mathcal{C}|))$-approximation for the general problem (Theorem 4.1). Applying this result to the case of SPARSE COVER in a distributed network (where $|A| = |\mathcal{C}| = n$),

**Input: an instance of MIN CCL.**
1. Minimize $k$ subject to constraints (2.1)–(2.5) and in addition that (2.4) holds with equality for every set $S \in \mathcal{C}$. Let $(x^*, y^*, z^*, k^*)$ denote the optimal LP solution.
2. Repeat $|L|\ln(2|\mathcal{C}|)$ times:
3. Choose a label $i \in L$ and a threshold $t \in [0,1]$ uniformly at random.
4. For each object $a \in A$, if $x^*_{ai} > t$, then add $i$ to the set of labels assigned to $a$.

Figure 3: The MIN CCL algorithm.

we immediately obtain an $O(\log n)$ approximation (Corollary 4.2). We also obtain a similar result for approximating the Nagata dimension of a finite metric space (Remark 4.3). Our second algorithm computes, for a given $\varepsilon \in (0, 1/4)$, a solution that consistently labels a $(1-3\varepsilon)$-fraction of the sets using $O(\ln(f_{\max}/\varepsilon))$ times more labels per object than the minimum necessary to consistently label at least $(1-\varepsilon)$ fraction of the sets. (Recall that $f_{\max} = \max_{S\in\mathcal{C}} |S|$; the constant 3 is quite arbitrary, and we make no attempt to optimize it.) This bicriteria guarantee is particularly appropriate for the NETWORK TRIANGULATION problem, where one typically permits a small fraction of pairs of points to have inaccurate distance estimates.

## 4.1 COMPLETE CONSISTENT LABELING and SPARSE COVER

Our approximation algorithm for MIN CCL is shown in Figure 3. The only difference between this algorithm and that for MAX CL and MAX FAIR CL (Figure 1) is the stopping condition: instead of explicitly controlling the number of labels assigned to each object, we stop after a fixed number of iterations.

**Theorem 4.1.** *The algorithm for* MIN CCL *in Figure 3 computes, with constant probability, an* $O(\log(|A|+|\mathcal{C}|))$*-approximation.*

*Proof.* Let $(x^*, y^*, z^*, k^*)$ denote the optimal LP solution. For a set $S \in \mathcal{C}$, let $x^{\min}_{iS}$ denote $\min_{a\in S} x^*_{ai}$. The probability that $S$ is consistently labeled in a given iteration equals

$$\frac{1}{|L|}\sum_{i\in L} x^{\min}_{iS} \geq \frac{1}{|L|}\sum_{i\in L} y^*_{iS} \geq \frac{z^*_S}{|L|} = \frac{1}{|L|},$$

with the inequalities following from the LP constraints (2.2)–(2.4). Using the independence of the different iterations and a union bound over all sets $S \in \mathcal{C}$, the probability that the algorithm terminates with an infeasible solution is at most

$$|\mathcal{C}|\left(1-\frac{1}{|L|}\right)^{|L|\ln(2|\mathcal{C}|)} \leq \frac{1}{2}.$$

On the other hand, constraint (2.1) ensures that the probability that an object $a \in A$ receives a label in a given iteration is

$$\frac{1}{|L|}\sum_{i\in L} x^*_{ai} \leq \frac{k^*}{|L|}.$$

Thus, the expected number of an object receives over all iterations is at most $k^* \ln(2|\mathcal{C}|+|A|)$. Since $k^* \geq 1$, applying Chernoff bounds and a union bound over all $a \in A$, it follows that with high probability, say $3/4$, the algorithm in Figure 3 terminates with each object receiving at most $k^* \cdot O(\log|A|+\log|\mathcal{C}|)$ labels. A final union bound now completes the proof. □

Of course, the success probability in Theorem 4.1 can be amplified arbitrarily via independent repetitions of the algorithm.

Modeling the SPARSE COVER problem as a special case of MIN CCL, as explained in the Introduction, we see that $|A| = |\mathcal{C}| = n$ (where $n = |X|$ is the size of the metric space), and the following corollary is immediate.

**Corollary 4.2.** *There is a randomized polynomial-time algorithm that, given an instance of* SPARSE COVER*, outputs, with high probability, a feasible cover with maximum degree $O(\log n)$ times that of optimal.*

**Remark 4.3.** Another similar application is that of computing the analog of the *Nagata dimension* [2, 25] of a finite metric space. The notion of bounded Nagata dimension generalizes that of bounded doubling dimension and of hyperbolic spaces (see [25]). With respect to a parameter $\gamma > 1$, the corresponding Nagata dimension of a finite metric space $(X,d)$, denoted $\dim_N(X,\gamma)$, is defined to be the smallest $r > 0$ such that for all $\Delta > 0$ there exists a $\Delta$-bounded cover of $X$ with the following property: every subset of $X$ with radius at most $\Delta/\gamma$ meets at most $r$ clusters in the cover.

Our MIN CCL algorithm—trivially generalized so that each set $S \in \mathcal{C}$ has its own restricted set $L_S$ of labels that can be used to consistently label it—gives an $O(\log n)$-approximation for computing the Nagata dimension. In more detail, consider the input $(X,d)$ and $\gamma$. There are only $\binom{n}{2}$ relevant values of $\Delta$, and we can consider each one separately; so fix a value of $\Delta$.

Define a MIN CCL instance by defining objects and labels as the points of $X$, and the sets $\mathcal{C}$ to be the $\Delta/\gamma$-balls around each point of $X$. The allowable labels for a set $S$ centered at $x$ are the points in the $\Delta$-ball around $x$. First suppose that there is a $\Delta$-bounded cover for which every $(\Delta/\gamma)$-ball meets at most $k$ different clusters of the cover; we can extract a feasible labeling as follows. For every cluster $S$ in the cover with center $x$, label all points within distance $\Delta/\gamma$ of $S$ by $x$. Since every point belongs to some cluster of the cover, every $(\Delta/\gamma)$-ball is consistently labeled. Since every $(\Delta/\gamma)$-ball meets at most $k$ different clusters of the cover, every point is assigned at most $k$ different labels.

Conversely, consider a feasible labeling to the consistent labeling problem. Form clusters by the following rule: Whenever $B(x,\Delta/\gamma)$ is consistently labeled with $y$, put $x$ in a cluster centered at $y$. Since every $(\Delta/\gamma)$-ball is consistently labeled, this defines a cover. The restricted label sets guarantee that the cover is $\Delta$-bounded. Finally, if $B(x,\Delta/\gamma)$ meets a cluster of the cover that is centered at $y$—so this ball contains a point $z$ such that $B(z,\Delta/\gamma)$ is consistently labeled by $y$—then $x$ is labeled with $y$ in the feasible labeling. Hence, the maximum number of labels at a point upper bounds the maximum number of clusters in the cover meeting a single $(\Delta/\gamma)$-ball. This correspondence between feasible labelings and covers implies that we can use Theorem 4.1 to approximate the Nagata dimension $\dim_N(X,\gamma)$ to within an $O(\log n)$ factor in polynomial time.

### 4.2 A bicriteria guarantee and application to METRIC TRIANGULATION

For a consistent labeling instance and a parameter $\alpha \in (0,1)$, let $k_{\text{opt}}(\alpha)$ be the smallest $k > 0$ for which there is a feasible labeling that assigns as most $k$ labels per object and is consistent for an $\alpha$ fraction of the sets. The following theorem achieves a bicriteria guarantee that is often reasonable for $\alpha$ close to 1; other trade-offs are also possible.

**Theorem 4.4.** *There is a randomized polynomial-time algorithm that, given an instance of* CONSISTENT LABELING *and* $0 < \varepsilon \le 1/4$*, computes with high probability a labeling that uses at most* $O(\ln(f_{\max}/\varepsilon)) \cdot k_{\text{opt}}(1-\varepsilon)$ *labels per object and is consistent for a* $(1-3\varepsilon)$ *fraction of the sets.*

*Proof.* The algorithm we use is shown in Figure 4. It is based on the MIN CCL algorithm (Figure 3), but differs from it as follows: Step 1 solves a slightly different LP relaxation, which includes the constraint $\sum_S z_S \ge (1-\varepsilon)|\mathcal{C}|$. The iterations work as before, and we perform exactly $m = 8|L|\ln(1/\varepsilon)$ iterations. Finally, for each object, we output only the first $\ell = \max\{2\ln(f_{\max}/\varepsilon), 16ek^*\ln(1/\varepsilon)\}$ labels it received.

It is easy to verify that the LP in step 1 is indeed a relaxation. Hence $k^* \le k_{\text{opt}}(1-\varepsilon)$ and, by definition, the algorithm outputs at most $\ell \le 16e\ln(f_{\max}/\varepsilon) \cdot k_{\text{opt}}(1-\varepsilon)$ labels per object.

Now, call a set $S$ *good* if $z_S^* \ge 1/4$ in the optimal LP solution. At least $(1-2\varepsilon)|\mathcal{C}|$ sets are good, for otherwise

$$\sum_S z_S^* < (1-2\varepsilon)|\mathcal{C}| \cdot 1 + \frac{2\varepsilon}{4}|\mathcal{C}| < (1-\varepsilon)|\mathcal{C}|,$$

which would contradict the last LP constraint. For every good set $S$, by the calculation in Theorem 4.1, the probability that none of the $m$ iterations labels $S$ consistently equals

$$\left(1 - \frac{1}{|L|}\sum_{i\in L} x_{iS}^{\min}\right)^m \le \left(1 - \frac{z_S^*}{|L|}\right)^m \le e^{-2\ln\frac{1}{\varepsilon}} = \varepsilon^2 .$$

Again following the proof of Theorem 4.1, the expected number of labels that a given object $a \in A$ receives during the $m$ iterations is

$$m \cdot \frac{1}{|L|}\sum_{i\in L} x_{ai}^* \le \frac{mk^*}{|L|} = 8k^* \ln\frac{1}{\varepsilon} .$$

By a Chernoff bound of the form

$$\Pr\Big[X \ge t \cdot \mathbb{E}[X]\Big] \le 2^{-t\,\mathbb{E}[X]}$$

for all $t \ge 2e$ (see, e. g., [34, Exercise 4.1]), the probability that a given $a \in A$ receives more than $\ell$ labels is at most $2^{-\ell} \le (\varepsilon/f_{\max})^2$. Applying a union bound, for every good set $S$, the probability that $S$ is not labeled consistently by the algorithm's output (either because none of the $m$ iterations labels it consistently or because the final step removes a consistent label) is at most

$$\varepsilon^2 + |S| \cdot \left(\frac{\varepsilon}{f_{\max}}\right)^2 \le 2\varepsilon^2 .$$

By linearity of expectation, the expected fraction of good sets that the algorithm does not label consistently is at most $2\varepsilon^2$, and using Markov's inequality, the probability that this fraction exceeds $\varepsilon$ is at most

**Input: an instance of MIN CL.**
1. Minimize $k$ subject to constraints (2.1)–(2.5) and in addition the constraint $\sum_S z_S \geq (1-\varepsilon)|\mathcal{C}|$. Let $(x^*, y^*, z^*, k^*)$ denote the optimal LP solution.
2. Repeat $m = 8|L| \ln \frac{1}{\varepsilon}$ times:
3. Choose a label $i \in L$ and a threshold $t \in [0,1]$ uniformly at random.
4. For each object $a \in X$, if $x^*_{ai} > t$, then add $i$ to the set of labels assigned to $a$.
5. For each object, output only the first $\ell = \max\{2 \ln \frac{f_{\max}}{\varepsilon}, 16ek^* \ln \frac{1}{\varepsilon}\}$ labels it received.

Figure 4: A bicriteria algorithm for MIN CL.

$2\varepsilon \leq 1/2$. Altogether we conclude that with probability at least $1/2$, the algorithm labels consistently at least a $(1-\varepsilon)(1-2\varepsilon) > 1-3\varepsilon$ fraction of the sets in $\mathcal{C}$. Obviously, we can amplify the success probability via independent repetitions. □

**Metric triangulation** Recall from the Introduction the problem of computing a triangulation of a metric that has low order. We can model this as a CONSISTENT LABELING problem with the same slight generalization as in Remark 4.3, and use Theorem 4.4 to prove the following.

**Corollary 4.5.** *There is a randomized polynomial-time algorithm that, given a* METRIC TRIANGULATION *instance (including $\rho$ and $\varepsilon$), outputs a $(1-3\varepsilon, \rho)$-triangulation of order $O(\ln \frac{1}{\varepsilon}) \cdot k_{\mathrm{opt}}(X, \varepsilon, \rho)$.*

*Proof.* We first model the METRIC TRIANGULATION problem as a slight generalization of MIN CCL. Objects correspond to the points $X$ and labels correspond to beacons (generally all of $X$). For every pair of nodes $x, y$ we want a consistency constraint that reflects our desire that $x, y$ have at least one beacon in $S_x \cap S_y$ attaining $D^+(x,y)$, and similarly at least one common beacon attaining $D^-(x,y)$. We model this by using set-dependent allowable labels $L_S$, and furthermore replacing the set of constraints (2.3) by two sets of constraints, one with allowable label set

$$L^+_{\{x,y\}} = \{b \in X : \ d(x,b) + d(b,y) \leq \rho \cdot d(x,y)\}$$

and one with allowable label set

$$L^-_{\{x,y\}} = \{b \in X : \ |d(x,b) - d(b,y)| \geq d(x,y)/\rho\};$$

the extra set of constraints only increases the hidden constants in our analysis. The correspondence between this variant of CONSISTENT LABELING and METRIC TRIANGULATION is immediate (notice that $f_{\max} = 2$), and we can thus use our algorithm from Theorem 4.4 to obtain a bicriteria bound for METRIC TRIANGULATION. □

## 5 Hardness results

This section shows that the consistent labeling problems studied in the preceding sections are APX-hard. We require only relatively simple reductions from well-known problems such as SET COVER. We have not made serious attempts to optimize these hardness results and it is quite possible that they can be strengthened.

**Minimum Consistent Labeling** We start with hardness results for minimum consistent labeling problems; these match, up to constant factors, the guarantees of our approximation algorithms in Theorems 4.1 and 4.4.

**Theorem 5.1.** *There exists a constant $c_0 > 0$ such that it is NP-hard to approximate the* MIN CCL *problem within a factor of $c_0 \log(|A| + |\mathcal{C}|)$.*

*Furthermore, for every fixed $\varepsilon \in (0, 1/4)$, it is NP-hard to find, given a* MIN CCL *instance, a labeling that is consistent for a $1-3\varepsilon$ fraction of the sets and uses at most $c_0 k_{\text{opt}}(1-\varepsilon) \cdot \log(1/\varepsilon)$ labels. These results hold even when $f_{\max} = 2$.*

*Proof.* The proof is by reduction from the SET COVER problem, defined as follows. The input is a set $E$ of elements and a collection $\mathcal{U} \subseteq 2^U$ of subsets. The goal is to find a a minimum-cardinality subcollection $\mathcal{U}' \subset \mathcal{U}$ that *covers* $E$, meaning that $\bigcup_{U \in \mathcal{U}'} U = E$. There is a constant $c_1 > 0$ such that it is NP-hard to decide whether (i) an input comprising a SET COVER instance and an integer $t > 0$ admits a cover $\mathcal{U}'$ of size at most $t$; or (ii) the size of every cover $\mathcal{U}'$ is at least $c_1 t \log |E|$ [32, 12, 36].

Our reduction from SET COVER to MIN CCL works as follows. Create an object for every element (so $E \subseteq A$) and a label for every SET COVER set (so $L = \mathcal{U}$). For each object (element) $e$, the permissible labels are the SET COVER sets that contain it: $L_e = \{U \in \mathcal{U} : \ e \in U\}$. Create one additional *root* object $r$ with $L_r = \mathcal{U}$. Finally, for every object $e$, create a consistency set $\{e, r\}$. Recall that the goal in the MIN CCL problem is to minimize the maximum number of labels assigned to an object. Notice that $f_{\max} = 2$, the number of objects is $|A| = |E| + 1$, and the number of consistency sets is $|\mathcal{C}| = |E|$.

First, suppose that the SET COVER instance has a cover $\mathcal{U}'$ of size $t$. Then $\mathcal{U}'$ naturally induces a feasible labeling: label each object $e$ with some set of $\mathcal{U}'$ that contains it, and label the root object $r$ with every set in $\mathcal{U}'$. This feasible labeling uses at most $|\mathcal{U}'| = t$ labels per object.

Second, suppose that every cover of the SET COVER instance has size at least $c_1 t \log |E|$. Consider a solution for the corresponding MIN CCL instance that uses only $k$ labels per object. Since every object $e$ participates in exactly one consistency set, we can assume that it is assigned a single label. The same label must be assigned also to the root object. It follows that the labeling of the root object corresponds to a feasible solution $\mathcal{U}'$ to the SET COVER instance, and hence at least $|\mathcal{U}'| \geq c_1 t \log |E|$ labels are assigned to the root object. It is therefore NP-hard to determine whether the value of a MIN CCL instance is at most $t$ or at least $(c_1/4) t \log(|A| + |\mathcal{C}|)$.

We prove the second assertion of the theorem statement using the following fact which holds for every fixed $\varepsilon \in (0, 3/4)$: Given as input $t > 0$ and a SET COVER instance that admits a cover of size at most $t$, it is NP-hard to find a subcollection $\hat{\mathcal{U}} \subseteq \mathcal{U}$ that covers a $1-\varepsilon$ fraction of the elements in $E$ and has size at most $(c_1/2) t \log(1/\varepsilon)$. This fact follows from the aforementioned hardness results, because a polynomial-time procedure that finds such a subcollection $\hat{\mathcal{U}}$ can be used (iteratively on the yet uncovered elements) to cover all of $E$ using less than $c_1 t \log |E|$ sets (see, e. g., [12, Proposition 5.2]). We will also use the fact that all of the sets in these hard SET COVER instances have essentially the same size $|U|/t$ (in fact, the optimal solution uses exactly $t$ sets disjoint of each other). Thus, for $\varepsilon < 1/4$, every subcollection $\hat{\mathcal{U}} \subseteq \mathcal{U}$ that covers a $1-\varepsilon$ fraction of the elements contains at least $t/2$ sets.

Now apply our reduction from SET COVER to MIN CCL, starting with the SET COVER instances described in the previous paragraph. It is straightforward to verify that: For $\varepsilon \in (0, 1/4)$ and $t > 0$, and MIN CCL instances that admit a solution of value $k_{\text{opt}}(1) = t$, it is NP-hard to find a labeling that is consistent for at least a $1-3\varepsilon$ fraction of the sets and uses at most $(c_0/2)t\log(1/\varepsilon)$ labels per object. Moreover, in these MIN CCL instances $k_{\text{opt}}(1-\varepsilon) \geq t/2$, and the second assertion of Theorem 5.1 follows. □

**Maximum Consistent Labeling** We next show hardness results for the maximum consistent labeling problems studied in Section 3.

**Theorem 5.2.** *The* MAX CL *and* MAX FAIR CL *problems are APX-hard, even when* $f_{\max} = 2$.

*Proof.* We use essentially the same reduction as in Theorem 5.1, starting from the MAX $k$-COVER problem, where given the same data as in a SET COVER instance and also a "budget" $t > 0$, the goal is to find a subcollection $\mathcal{U}' \subseteq \mathcal{U}$ of size $t$ that maximizes the number of elements covered. For every fixed $0 < c_2 < 1/e$, it is NP-hard to decide whether $t$ sets can cover all the elements of a MAX $k$-COVER instance, or whether they can cover at most a $1-c_2$ fraction of the elements [12, Proposition 5.3]. Arguing as in the proof of Theorem 5.1 shows the following: for every fixed $0 < c_2 < 1/e$, it is NP-hard to decide whether a MAX CL instance has value (i. e., fraction of consistently labeled sets) 1 or value at most $1-c_2$.

Extending the argument to MAX FAIR CL is immediate, using the exact same reduction. First, suppose that the MAX $k$-COVER instance can be covered using $t$ sets. Then the MAX CL instance admits a labeling which is consistent for all sets, and thus can be viewed as a solution to the MAX FAIR CL instance with value 1 (the distribution over labelings uses only one labeling). Second, for $c_2 < 1/e$, suppose that every $t$ sets can cover at most $(1-c_2)t$ elements. As argued earlier, it follows that every solution to the respective MAX CL instance consistently labels at most a $1-c_2$ fraction of the sets. Since a solution to MAX FAIR CL is just a distribution over solutions to MAX CL, we immediately see that the former has value at most $1-c_2$. This shows that approximating MAX FAIR CL to within a factor of $1-c_2$ is NP-hard. □

We provide next two hardness results for MAX CL that apply even when the number $k$ of allowable labels per object is 1.

**Theorem 5.3.** *For every fixed* $f_{\max} \geq 3$*, the* MAX CL *problem is NP-hard to approximate to within a factor of* $\Omega(f_{\max}/\log f_{\max})$*, even when* $k = 1$.

*Proof sketch.* We show a reduction from the $t$-SET PACKING problem, which is NP-hard to approximate to within a factor of $\Omega(t/\log t)$ when $t$ (the size of the sets) is fixed [16]. The reduction takes a $t$-SET PACKING instance, which comprises elements and sets, and constructs a MAX CL instance by letting the elements be our objects and the sets our labels. The labels (sets) that are allowed for an object (element) are the sets that contain the element. The consistency sets are just the $t$-SET PACKING sets, hence $f_{\max} = t$. We also set the number $k$ of labels allowed per object to 1. The theorem follows by observing that a packing of $p$ sets naturally induces a feasible labeling that is consistent for $p$ sets, and conversely. □

**Theorem 5.4.** *The* MAX CL *problem is NP-hard, even when* $f_{\max} = 2$ *and* $k = 1$.

*Proof sketch.* We give a reduction from the MULTIWAY CUT problem, which is NP-hard even with three terminals [9]. Given an input graph $G$ for multiway cut with three terminals $\{t_1, t_2, t_3\} \subset V$, build a MAX

CL instance by letting $G$'s vertices be our objects, and the three terminals be our labels. Every object is allowed every label, except that each terminal vertex $t_i$ is allowed only itself as a label, i. e., $L_{t_i} = \{t_i\}$. Every edge in $G$ becomes a consistency set of cardinality $f_{\max} = 2$ in the obvious way. In addition, the number of labels per object is $k = 1$.

The theorem follows by observing that the multiway cuts of $G$ are in one-to-one correspondence with feasible labelings, where the uncut edges correspond to consistently labeled sets. □

## 6 Concluding remarks

For most of the problems studied in this paper, we leave open the question of whether our guarantees are close to the best possible. While we know that the abstract consistent labeling problems are APX-hard (see Section 5), we know little about the four special cases that are listed in Table 1 and motivated this work. For example, we are not aware of any hardness of approximation results that exclude a true (non-bicriteria) approximation for computing PADDED DECOMPOSITION and METRIC TRIANGULATION, even for general metric spaces. We also cannot rule out a constant-factor approximation algorithm for computing SPARSE COVER. The one slight exception is Theorem 3.5, which gives a lower bound on the integrality gap of our linear program for computing a separating decomposition.

Another direction for future research is to study optimization problems inspired by the metric decomposition of Arora, Rao and Vazirani [1]. For example, one could seek a relative guarantee, analogous to the absolute guarantee in [1], for the following problem: The input is a metric space $(X,d)$ and parameter $\delta > 0$, and the goal is to find $A, B \subset X$ satisfying $|A|, |B| \geq \delta|X|$, so as to maximize $d(A,B) = \min_{a \in A, b \in B} d(a,b)$. This problem does not seem to fall within our consistent labeling framework.

### Acknowledgments

We thank Anupam Gupta and James Lee for preliminary discussions about the various concepts used in the paper. We thank Laci Babai and four anonymous journal referees for their helpful comments on an earlier draft.

## References

[1] SANJEEV ARORA, SATISH RAO, AND UMESH V. VAZIRANI: Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2):1–37, 2009. [doi:10.1145/1502793.1502794] 70

[2] PATRICE ASSOUAD: Sur la distance de Nagata. *C. R. Acad. Sci. Paris Ser. I Math.*, 1(294):31–34, 1982. 54, 65

[3] BARUCH AWERBUCH AND DAVID PELEG: Sparse partitions. In *Proc. 31st FOCS*, pp. 503–513. IEEE Comp. Soc. Press, 1990. [doi:10.1109/FSCS.1990.89571] 51, 52, 53, 54

[4] YAIR BARTAL: Probabilistic approximation of metric spaces and its algorithmic applications. In *Proc. 37th FOCS*, pp. 184–193. IEEE Comp. Soc. Press, 1996. [doi:10.1109/SFCS.1996.548477] 51, 52

[5] YAIR BARTAL AND ROBERT KRAUTHGAMER: Unpublished, 2007. 52

[6] MIHAI BĂDOIU, PIOTR INDYK, AND ANASTASIOS SIDIROPOULOS: Approximation algorithms for embedding general metrics into trees. In *Proc. 18th Ann. ACM-SIAM Symp. Discrete Algorithms (SODA'07)*, pp. 512–521. Society for Industrial and Applied Mathematics, 2007. 50, 52

[7] GRUIA CALINESCU, HOWARD J. KARLOFF, AND YUVAL RABANI: An improved approximation algorithm for MULTIWAY CUT. *J. Comput. System Sci.*, 60(3):564–574, 2000. [doi:10.1006/jcss.1999.1687] 55

[8] MOSES CHARIKAR, CHANDRA CHEKURI, ASHISH GOEL, SUDIPTO GUHA, AND SERGE PLOTKIN: Approximating a finite metric by a small number of tree metrics. In *Proc. 39th FOCS*, pp. 379–388. IEEE Comp. Soc. Press, 1998. [doi:10.1109/SFCS.1998.743488] 51, 52, 53

[9] ELIAS DAHLHAUS, DAVID S. JOHNSON, CHRISTOS H. PAPADIMITRIOU, PAUL D. SEYMOUR, AND MIHALIS YANNAKAKIS: The complexity of multiterminal cuts. *SIAM J. Comput.*, 23(4):864–894, 1994. [doi:10.1137/S0097539792225297] 69

[10] JITTAT FAKCHAROENPHOL, SATISH RAO, AND KUNAL TALWAR: A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. System Sci.*, 69(3):485–497, 2004. [doi:10.1016/j.jcss.2004.04.011] 52

[11] JITTAT FAKCHAROENPHOL AND KUNAL TALWAR: Improved decompositions of graphs with forbidden minors. In *Proc. 6th Intern. Workshop Approx. Algorithms for Comb. Opt. (APPROX '03)*, volume 2764 of *Lecture Notes in Comput. Sci.*, pp. 871–882. Springer, 2003. [doi:10.1007/978-3-540-45198-3_4] 50, 52

[12] URIEL FEIGE: A threshold of $\ln n$ for approximating set cover. *J. ACM*, 45(4):634–652, 1998. [doi:10.1145/285055.285059] 68, 69

[13] NAVEEN GARG, VIJAY V. VAZIRANI, AND MIHALIS YANNAKAKIS: Approximate max-flow min-(multi)cut theorems and their applications. *SIAM J. Comput.*, 25(2):235–251, 1996. [doi:10.1137/S0097539793243016] 52

[14] ANUPAM GUPTA, ROBERT KRAUTHGAMER, AND JAMES R. LEE: Bounded geometries, fractals, and low-distortion embeddings. In *Proc. 44th FOCS*, pp. 534–543. IEEE Comp. Soc. Press, 2003. [doi:10.1109/SFCS.2003.1238226] 52

[15] JAMES D. GUYTON AND MICHAEL F. SCHWARTZ: Locating nearby copies of replicated internet servers. In *Proc. Conf. Appl., Technol., Architectures, and Protocols for Comput. Commun. (SIGCOMM '95)*, pp. 288–298, New York, NY, USA, 1995. ACM Press. [doi:10.1145/217382.217463] 54

[16] ELAD HAZAN, SHMUEL SAFRA, AND ODED SCHWARTZ: On the complexity of approximating $k$-set packing. *Comput. Complexity*, 15(1):20–39, 2006. [doi:10.1007/s00037-006-0205-6] 69

[17] CLAIRE KENYON, YUVAL RABANI, AND ALISTAIR SINCLAIR: Low distortion maps between point sets. *SIAM J. Comput.*, 39(4):1617–1636, 2009. [doi:10.1137/080712921] 50, 52

[18] PHILIP N. KLEIN, SERGE A. PLOTKIN, AND SATISH RAO: Excluded minors, network decomposition, and multicommodity flow. In *Proc. 25th STOC*, pp. 682–690. ACM Press, May 1993. [doi:10.1145/167088.167261] 50, 52

[19] JON M. KLEINBERG, ALEKSANDRS SLIVKINS, AND TOM WEXLER: Triangulation and embedding using small sets of beacons. *J. ACM*, 56(6):1–37, 2009. Earlier version in FOCS '04, pp. 444–453. [doi:10.1145/1568318.1568322] 54

[20] JON M. KLEINBERG AND ÉVA TARDOS: Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. *J. ACM*, 49(5):616–639, 2002. [doi:10.1145/585265.585268] 55, 57

[21] ROBERT KRAUTHGAMER: On triangulation of simple networks. In *Proc. 19th Ann. ACM Symp. Parallel Algorithms and Architectures (SPAA'07)*, pp. 8–15. ACM Press, 2007. [doi:10.1145/1248377.1248380] 54

[22] ROBERT KRAUTHGAMER AND JAMES R. LEE: Algorithms on negatively curved spaces. In *Proc. 47th FOCS*, pp. 119–132. IEEE Comp. Soc. Press, 2006. [doi:10.1109/FOCS.2006.9] 52

[23] ROBERT KRAUTHGAMER, JAMES R. LEE, MANOR MENDEL, AND ASSAF NAOR: Measured descent: A new embedding method for finite metrics. *Geom. Funct. Anal.*, 15(4):839–858, 2005. [doi:10.1007/s00039-005-0527-6] 51, 52

[24] ROBERT KRAUTHGAMER AND TIM ROUGHGARDEN: Metric clustering via consistent labeling. In *Proc. 19th Ann. ACM-SIAM Symp. Discrete Algorithms (SODA'08)*, pp. 809–818, 2008. 49

[25] URS LANG AND THILO SCHLICHENMAIER: Nagata dimension, quasisymmetric embeddings, and Lipschitz extensions. *Int. Math. Res. Not.*, 2005(58):3625–3655, 2005. [doi:10.1155/IMRN.2005.3625] 54, 65

[26] MICHAEL LANGBERG, YUVAL RABANI, AND CHAITANYA SWAMY: Approximation algorithms for graph homomorphism problems. In *Proc. 9th Intern.. Workshop Approx. Algorithms for Comb. Opt. (APRROX '06)*, volume 4110 of *Lecture Notes in Comput. Sci.*, pp. 176–187. Springer, 2006. [doi:10.1007/11830924_18] 57

[27] JAMES R. LEE AND ASSAF NAOR: Metric decomposition, smooth measures, and clustering. Manuscript, available at `http://www.cims.nyu.edu/~naor/homepage%20files/cluster.pdf`, 2004. 53

[28] JAMES R. LEE AND ANASTASIOS SIDIROPOULOS: Genus and the geometry of the cut graph. In *Proc. 20th Ann. ACM-SIAM Symp. Discrete Algorithms (SODA '10)*, pp. 193–201. SIAM, 2010. 51, 52

[29] FRANK THOMSON LEIGHTON AND SATISH RAO: An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms. In *Proc. 29th FOCS*, pp. 422–431. IEEE Comp. Soc. Press, 1988. [doi:10.1109/SFCS.1988.21958] 52

[30] NATHAN LINIAL, ERAN LONDON, AND YURI RABINOVICH: The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. [doi:10.1007/BF01200757] 50

[31] NATHAN LINIAL AND MICHAEL SAKS: Low diameter graph decompositions. *Combinatorica*, 13(4):441–454, 1993. [doi:10.1007/BF01303516] 52

[32] CARSTEN LUND AND MIHALIS YANNAKAKIS: On the hardness of approximating minimization problems. *J. ACM*, 41(5):960–981, 1994. [doi:10.1145/185675.306789] 68

[33] JIŘÍ MATOUŠEK AND ANASTASIOS SIDIROPOULOS: Inapproximability for metric embeddings into $R^d$. In *Proc. 49th FOCS*, pp. 405–413. IEEE Comp. Soc. Press, 2008. [doi:10.1109/FOCS.2008.21] 50

[34] RAJEEV MOTWANI AND PRABHAKAR RAGHAVAN: *Randomized Algorithms*. Cambridge University Press, 1995. 66

[35] SATISH RAO: Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proc. 15th Ann. Symp. Comput. Geom. (SoCG '99)*, pp. 300–306. ACM Press, 1999. [doi:10.1145/304893.304983] 50, 51, 52

[36] RAN RAZ AND SHMUEL SAFRA: A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proc. 29th STOC*, pp. 475–484. ACM Press, 1997. [doi:10.1145/258533.258641] 68

[37] ALEKSANDRS SLIVKINS: Distributed approaches to triangulation and embedding. In *Proc. 16th Ann. ACM-SIAM Symp. Discrete Algorithms (SODA'05)*, pp. 640–649, 2005. 54

[38] ALEKSANDRS SLIVKINS: Distance estimation and object location via rings of neighbors. *Distrib. Comput.*, 19(4):313–333, 2007. Earlier version in PODC '05, pp. 41–50. [doi:10.1007/s00446-006-0015-8] 54

## AUTHORS

Robert Krauthgamer
Department of Computer Science and Applied Mathematics
The Weizmann Institute of Science
Rehovot, Israel
robert.krauthgamer@weizmann.ac.il
`http://www.wisdom.weizmann.ac.il/~robi`

Tim Roughgarden
Assistant Professor
Department of Computer Science
Stanford University
Stanford, CA USA
tim@cs.stanford.edu
http://theory.stanford.edu/~tim

## ABOUT THE AUTHORS

ROBERT KRAUTHGAMER received his Ph. D. at the Weizmann Institute of Science in 2001 under Uriel Feige. He was subsequently a postdoc in Berkeley's theory group, and then a Research Staff Member at the theory group in the IBM Almaden Research Center. Since 2007, he has been a faculty member at the Weizmann Institute of Science. Robert's main research area is the design of algorithms for problems involving combinatorial optimization, finite metric spaces, high-dimensional geometry, data analysis, and related areas. His favorite sport since youth is swimming, and once he swam across the Sea of Galilee in a 10km race, and was the last one to arrive at the finish line.

TIM ROUGHGARDEN received his Ph. D. at Cornell University in 2002 under Éva Tardos. His research interests are in algorithms, and especially in algorithmic game theory.