Maximizing Covered Area in the Euclidean Plane with Connectivity Constraint*

Chien-Chung Huang[†] Mathieu Mari[‡] Claire Mathieu[§] Joseph S. B. Mitchell[¶]
Nabil H. Mustafa[‖]

April 3, 2022

Abstract

Given a set $\mathcal{D}$ of n unit disks in the plane and an integer $k \leq n$, the maximum area connected subset problem asks for a set $\mathcal{D}' \subseteq \mathcal{D}$ of size k maximizing the area of the union of disks in $\mathcal{D}'$, under the constraint that this union is connected. This problem is motivated by wireless router deployment and is a special case of maximizing a submodular function under a connectivity constraint.

We prove that the problem is NP-hard and analyze a greedy algorithm, proving that it computes a $\frac{1}{2}$-approximation. We then give a polynomial-time approximation scheme (PTAS) for this problem with resource augmentation, i.e., allowing an additional set of εk unit disks that are not drawn from the input. Additionally, for two special cases of the problem we design a PTAS without resource augmentation.

1 Introduction

Maximizing a submodular function[1] under constraints is a classical problem in computer science and operations research [29, 33]. The most commonly studied constraints are cardinality, knapsack and matroids constraints. A natural constraint that has received little attention is the *connectivity* constraint. In this paper, we study the following problem. Given a set $\mathcal{D}$ of n unit disks in the plane and an integer $k \leq n$, compute a set $\mathcal{D}' \subseteq \mathcal{D}$ of size k that maximizes the area of the union of disks in $\mathcal{D}'$, under the constraint that this union is connected. We call this problem *Maximum Area Connected Subset* problem (MACS). Notice that the area covered by the union of a set of disks is a monotone submodular function.

*An earlier version of this paper appeared in the *Proceedings of Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM), 2019.*

[†]DI ENS, École Normale supérieure, Université PSL, Paris, France and CNRS, Paris, France. cchuang@di.ens.fr.

[‡]University of Warsaw, Poland. mathieu.mari@ens.fr.

[§]CNRS and IRIF, Université de Paris, Paris, France. clairemmathieu@gmail.com.

[¶]Stony Brook University, Stony Brook, NY 11794, USA. joseph.mitchell@stonybrook.edu. Partially supported by the National Science Foundation (CCF-2007275, CCF-1540890), the US-Israel Binational Science Foundation (project 2016116), and DARPA (Lagrange).

[‖]Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France. mustafan@esiee.fr. Research partially supported by ANR ADDS (ANR-19-CE48-0005).

[1]Given a set X, a function $f : 2^X \to \mathbb{R}$ is *submodular* if given any two subsets $A, B \subseteq X$, $f(A) + f(B) \geq f(A \cap B) + f(A \cup B)$.

The problem is motivated by wireless router deployment, first introduced in [23]: the goal is to install a certain number of routers to maximize the number of clients covered while also ensuring that these routers are connected to each other. When the clients are spatially uniformly spread, the number of clients covered is proportional to the area and hence the objective is to maximize the area covered. We note that another motivation to connectivity constraint for submodular function maximization come from cancer genome studies (see related work section below).

Our Contributions. We first analyze a variant of the greedy algorithm and prove that it computes a $\frac{1}{2}$-approximation (Theorem 2); further we show that the analysis of the algorithm is tight. On the other hand, we show that the naive greedy algorithm that adds disks one at a time to maximize the area of the union computes, in the worst-case, a solution that is a $\Omega(k)$-factor smaller than the optimal one.

To improve upon the $\frac{1}{2}$-approximation ratio, we turn to the resource augmentation setting in which the algorithm is allowed to add a few additional disks that are not drawn from the input. We design a PTAS for the resource augmentation version of the problem (Theorem 3) using the m-guillotine method of Mitchell [28]. (In the conference version of this paper [18] we give a randomized PTAS using Arora's shifted dissection technique [1].) The correctness proof hinges on a structural statement which shows the existence of a near-optimal solution with $O(\varepsilon k)$ additional disks, and with additional structure that allows it to be computed efficiently by dynamic programming.

For two special cases of the MACS we obtain a PTAS without resource augmentation: (i) when the Euclidean distances between the input disk centers are well-approximated by shortest paths in the intersection graph (Theorem 4), and (ii) when every point in the convex-hull of the input disk centers is covered by at least one input disk (Corollary 1).

On the negative side, via a reduction from the Rectilinear Steiner Tree problem, we show that MACS is NP-hard (Theorem 1). We further show that if the goal is to compute MACS for a set of arbitrary quadrilaterals instead of disks, the problem is APX-hard (Theorem 5). We leave open the question of whether MACS is APX-hard or admits a PTAS without resource augmentation.

Related work. Maximizing a monotone submodular function under constraint(s) is a subject that has received a large amount of attention over the years. Kulik et al. [22] designed an approximation algorithm for maximizing a submodular function under multiple linear constraints with an approximation ratio that (almost) matches the bound of $1 - 1/e$. The greedy algorithm gives a $1/(k+1)$-approximation where the objective function is subject to k matroid constraints [29]. Lee et al. [25] later improved the approximation arbitrarily close to k when $k \geqslant 2$ using a local-search approach. When a monotone submodular function is subject to only one matroid constraint, there is a randomized $(1 - 1/e)$-approximation algorithm [3].

Our problem can be regarded as maximizing a submodular function under a cardinality (knapsack) constraint and a connectivity constraint. Notice that the connectivity constraint is central to the difficulty of our problem: without connectivity constraints, MACS admits a PTAS even for the more general case of convex pseudodisks [6].

Another motivation for studying the connectivity constraint is related to cancer genome studies. Suppose that a vertex represents an individual protein (and associated gene), an edge represents pairwise interactions, and each vertex has an associated set. Finding the connected subgraph of k genes that is mutated in the largest number of samples is equivalent to the problem of finding the connected subgraph with k nodes that maximizes the cardinality of the union of of the associated sets (see [32]).

In the general (non-geometric) setting where a general monotone submodular function is given and the connectivity constraint is that a feasible solution must induce a connected subgraph of a given input graph, a $\Omega\left(\frac{1}{\sqrt{k}}\right)$-approximation algorithm is given in Kuo et al. [23]. This approximation is obtained by computing, for each vertex in the graph, a set of $\sqrt{k}$ vertices in the $\sqrt{k}$-neighborhood around this root vertex that are then connected by shortest paths. The output is then the solution, among all root vertices, with maximum value. Our results show that when the submodular function and the connectivity are induced by a geometric configuration, the approximation ratio can be significantly improved.

We next discuss several related problems where the connectivity constraint is involved. An example is the node-cost budget problem introduced by Rabani and Scalosub [30], where the goal is to find a connected set of vertices in a general graph to collect the maximum profit on the vertices while guaranteeing that the total cost does not exceed a certain budget. Notice that in this setting the submodular function is a simple additive function of the profits.

Khuller et al. [21] study the budgeted connected dominating set problem, where given a general undirected graph, there is a budget k on the number of vertices that can be selected, and the goal is to induce a connected subgraph that dominates as many vertices as possible. It was pointed out to us that via a reduction, their algorithm gives a $\frac{1}{13}\left(1-\frac{1}{e}\right)$-approximate solution for MACS[2]. The approximation of the budgeted connected dominating set problem was very recently improved to $\frac{1}{11}(1-e^{-7/8})$ [24]. Hochbaum and Pathria [17] consider the problem of selecting k nodes of an input node-weighted graph to form a connected subgraph, with the aim of maximizing or minimizing the selected weight.

We now turn to the geometric setting. For the connected sensor coverage problem introduced by Gupta et al. [15], a logarithmic-factor approximation algorithm is known. In this problem, one selects at most k sensors in the plane forming a connected communication network and covering the desired region. Here the region covered by each sensor is not necessarily a disk but may be a convex region of the plane (see [13, 20]). Our resource augmentation PTAS relies on ideas used for Euclidean TSP and other geometric problems [1, 28]. Another related problem introduced by Chambers et al. [4], is to assign radii to a given set of points in the plane so that the union of the resulting disks is connected, the objective being to minimize the sum of radii. A $(1-\varepsilon)$-approximation algorithm in time $n^{O(1/\varepsilon)}$ for the maximum independent set problem on unit disk graphs is known [27]. Marathe et al. [26] present a constant-factor approximation algorithm for several problems on unit disk graphs, including maximum independent set. The maximum independent set problem is NP-hard even for unit disk graphs in the plane [7]. When the goal is to cover a specified set of clients (instead of the maximum area) with the minimum number of disks (instead of constraining the number of disks to at most k), and there is no connectivity constraint, the problem is NP-hard [7] but there exists a polynomial-time approximation scheme [19].

[2] Here is a brief outline of the reduction: given an instance of n unit disks in the plane, add a sufficiently large number of "client" points in the plane and edges between a disk x and any client that is covered by x so that the area covered corresponds to the number of client points dominated up to a fixed constant factor.

	general case	unit-disks
without connectivity	$(1-\frac{1}{e})$-approximation [29]	PTAS [6]
with connectivity	$\Omega\left(\frac{1}{\sqrt{k}}\right)$-approximation [23]	$\frac{1}{2}$-approximation (Theorem 2)

Figure 1: Previous approximation bounds for maximizing a positive monotone submodular function with and without the connectivity constraint.

2 Formal definitions and summary of results

Let $d_G(x, y)$ denote the distance between two vertices x and y in an edge-weighted graph G that defines a discrete metric (edge weights satisfy the triangle inequality). The Euclidean distance between two points x and y is denoted by $|xy|$. When there is no confusion, we will refer to a point x in the plane and the unit disk centered at x interchangeably.

Definition 1. *Given a finite set S in the plane, the* unit disk intersection graph UDG(S) *is an undirected graph on S where there is an edge, $\{x, y\}$, for $x, y \in S$ if and only if $|xy| \leq 2$.*

A set S of points in the plane is said to be *connected* if UDG(S) is a connected graph. We now formally define the Maximum Area Connected Subset (MACS) problem:

Definition 2. *Given a finite set of points $X \subseteq \mathbb{R}^2$ such that* UDG(X) *is connected and a non-negative integer k ($k \leq |X|$), the Maximum Area Connected Subset (*MACS*) problem is to find a subset $S \subseteq X$ of size at most k such that* UDG(S) *is connected and so that the area of union of the unit disks centered at points of S is maximized.*

For an input (X, k), an optimal solution (subset $S \subseteq X$) for MACS is denoted by $\mathbf{OPT}(X, k)$; when the context is clear, we refer to $\mathbf{OPT}(X, k)$ as $\mathbf{OPT}$, which is also used to denote the area covered by the optimal solution. (Observe that $\mathbf{OPT}$ is trivially upper-bounded by πk.) Any $S \subseteq X$ with $|S| \leqslant k$ for which UDG(S) is connected is called a *feasible solution* for MACS. We distinguish between the *decision version* of MACS, in which we are to decide for a given $q > 0$ if there exists a feasible solution of area at least q (i.e., if $\mathbf{OPT}(X, k) \geq q$), and the *optimization version* of MACS, in which we are to determine $\mathbf{OPT}(X, k)$.

Note that we have assumed in the definition of MACS that X is a connected set. This is without loss of generality since, for a set X that is not connected (i.e., for which UDG(X) has two or more connected components), the MACS on X can be solved separately on each connected component of UDG(X) that has cardinality at least k, and then the best solution among these can be taken as the solution to MACS on X.

Hardness and a constant-factor approximation for MACS are established in our first two theorems:

Theorem 1 (Hardness)**.** *The decision version of* MACS *is NP-hard.*

Theorem 2 (Approximation)**.** *The optimization version of* MACS *has a polynomial time algorithm (Algorithm 1) that achieves a* (1/2)*-approximation.*

In the resource augmentation version of MACS, we are to compute a subset $S \subseteq X$ of at most k points, together with a small subset S_{add} of additional points that augment the set of allowed disk centers, such that $\mathrm{UDG}(S \cup S_{add})$ is connected and so that the area of the union of the unit disks centered at points of S is maximized. The following theorem[3] states our main result for the resource augmentation version of MACS, for which we obtain a $(1-\varepsilon)$-approximation, using a set S_{add} of at most εk additional points:

Theorem 3 (Resource augmentation). *Let $\varepsilon > 0$ be a given parameter. Given a connected set $X \subseteq \mathbb{R}^2$ of n points and a positive integer k, let $\mathbf{OPT}(X,k)$ be the maximum possible area of a connected union of a set of k unit disks centered at a subset of k points of X. Then, there is a deterministic algorithm that computes, in time $n^{O(\varepsilon^{-1})}$, a subset $S \subseteq X$ of size at most k and a set $S_{add} \subseteq \mathbb{R}^2$ of at most εk points, such that $\mathrm{UDG}(S \cup S_{add})$ is connected, and the area covered by the unit disks centered at S is at least $(1-\varepsilon)\mathbf{OPT}(X,k)$.*

While it remains open whether or not MACS, without resource augmentation, has a PTAS (Theorem 2 gives a (1/2)-approximation), we are able to obtain a PTAS under certain assumptions about the set X, namely if X is "α-well-distributed" in the Euclidean plane.

Definition 3. *Given $\alpha \geq 1$, a finite set X of points in the plane is called* α-well-distributed *if for all $x, y \in X$, $d_{\mathrm{UDG}(X)}(x,y) \leqslant \lceil \alpha \cdot |xy| \rceil$, where the distance $d_{\mathrm{UDG}(X)}$ is based on Euclidean edge weights in* $\mathrm{UDG}(X)$.

In Section 5.6 we prove that MACS has a PTAS if the input points X are α-well-distributed.

Theorem 4 (α-well-distributed). *The* MACS *on α-well-distributed inputs (for a constant α) has a polynomial-time approximation scheme.*

A sufficient condition on the input X that implies that it is α-well-distributed is the property of being "pseudo-convex", which we define in this context as follows:

Definition 4. *A set X is called* pseudo-convex *if the convex-hull of X is covered by the union of the unit disks centered at points of X.*

Lemma 1. *A pseudo-convex set $X \subseteq \mathbb{R}^2$ is 3.82-well-distributed.*

The exact constant is $12/\pi < 3.82$ and is obtained by simple geometric observations and a disk packing argument (see Section 5). An immediate corollary of Theorem 4 is the following.

Corollary 1. MACS *on pseudo-convex inputs admits a polynomial-time approximation scheme.*

In contrast, by a reduction from 3-SET-COVER, we show that a similar problem stated with quadrilaterals instead of disks is hard to approximate.

Definition 5. *Given a finite set $\mathcal{T}$ of convex quadrilaterals in the plane, and an integer k ($k \leq |\mathcal{T}|$), the* QUAD-CONNECTED-COVER *problem is to find a subset $T \subseteq \mathcal{T}$ of size at most k such that the intersection graph[4] of T is connected and so that the area covered by the union of the quadrilaterals in T is maximized.*

Theorem 5. QUAD-CONNECTED-COVER *is APX-hard.*

[3] In the earlier conference version of this paper [18], we provide an alternative PTAS for the augmentation version of MACS, with a slightly worse running time, based on the (randomized) shifted quadtree approach of Arora [1].

[4] The intersection graph has vertex set T, and two quadrilaterals are adjacent in this graph if and only if they intersect.

3 NP-Hardness of MACS

We present a reduction from the NP-hard Rectilinear Steiner Tree (RST) problem to prove that the decision version of MACS is NP-hard.

> Rectilinear Steiner Tree problem: Given n terminals in the Euclidean plane and a number L, decide whether there exists a tree to connect all the n terminals using horizontal and vertical line segments of total length at most L.

The problem is NP-complete [14], even if all terminals have integer coordinates bounded by $V = \mathit{poly}(n)$. In the following, we assume that n is sufficiently large ($n \geq 8$ suffices).

Given an instance of MACS and a real $q > 0$, we show that it is NP-hard to decide whether there is a feasible solution to MACS that covers an area at least q (however, notice that in our reductions, the given set of points X do not necessarily have integer coordinates). We start from an instance of RST and construct an instance of MACS as follows.

- For all integers $0 \leq i, j \leq V$, place one disk with center at (in, jn). We call them *cardinal disks.*
- For all integers $0 \leq i, j \leq V$, place $n - 4$ disks centered at $\left(in + 2 + \left(1 + \frac{1}{n-5}\right) \cdot t, jn\right)$ where $t \in \{0, \ldots, n-5\}$ and $n - 4$ disks centered at $\left(in, jn + 2 + t\left(1 + \frac{1}{n-5}\right)\right)$ where $t \in \{0, \ldots, n-5\}$. We call them *path disks.*
- For each terminal (i, j) in the RST instance, we place $n^2/10$ *bonus disks*: the first one centered at $\left(in + \sqrt{2}, jn + \sqrt{2}\right)$ and the remaining centers forming a connected group in $[in + 2, (i+1)n - 2] \times [jn + 2, (j+1)n - 2]$ in such a way that each bonus disk is tangent[5] to other bonus disks, and can be connected to the first bonus disk. Notice that except the first one, no bonus disk intersects path disks. This defines the set of disks.
- Set $k = 1 + L(n-3) + n^3/10$.

This defines the MACS instance. See Figure 2 for an illustration. Note that the interior of a cardinal disk is disjoint from all other disks of the instance.

Notice that as the RST instance has all terminals inside a rectangle of size polynomial in n, the above reduction can be done in polynomial time.

Let Z denote the set of the cardinal disk at $(0, 0)$ and the $n-4$ path disks at $\left(2 + t\left(1 + \frac{1}{n-5}\right), 0\right)$ where $t \in \{0, \ldots, n-5\}$ and let $\mathcal{A}(Z)$ denote the area covered by Z.

Lemma 2. *The original* RST *instance has a feasible solution of total length at most L if and only if the derived* MACS *instance has a feasible solution of area of at least* $\pi + L \cdot \mathcal{A}(Z) + \left(\frac{n^3}{10} - \frac{n}{3}\right)\pi$.

Proof. First, consider the "only if" direction: Assume that the original RST instance has a feasible solution of total length at most L. We call a set of disks a *segment* if it consists of a cardinal disk and all the $n - 4$ path disks between it and one of its four adjacent cardinal disks. Thus, the area covered by a segment is exactly $\mathcal{A}(Z)$. Consider a feasible solution for the RST instance, of length exactly L, without loss of generality. We root it at an arbitrary integral point, direct it outwards

[5]their centers are at distance exactly 2.

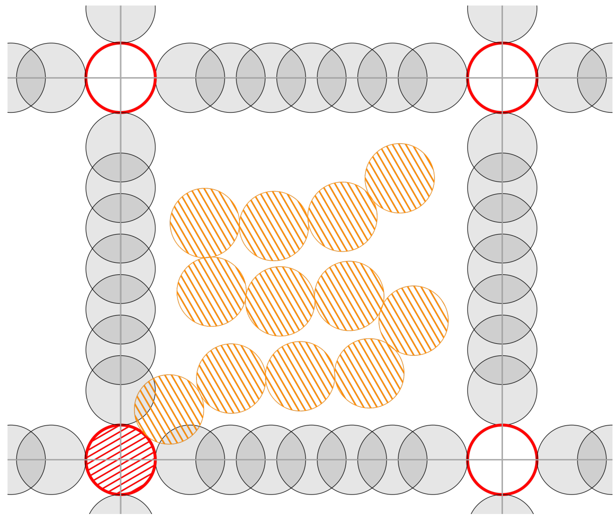


Figure 2: Filled (grey), hatched (orange) and empty (red) disks respectively represent *path*, *bonus* and *cardinal* disks. The hatched cardinal disk is associated with a terminal node.

from the root, and view it as a collection of horizontal or vertical directed edges of unit length. In the MACS instance, we take all bonus disks, the cardinal disk associated to the root of the RST solution, and, for each directed edge of the RST solution, all disks of the corresponding segment. The total number of disks is exactly k, and the area covered is at least $\pi + L \cdot \mathcal{A}(Z) + \frac{n^3\pi}{10} - 2n\gamma$, where γ is the area of the intersection of the first bonus disk associated with a terminal and the path disk just tangent to the corresponding cardinal disk of the latter. (Recall that the first bonus disk can overlap with up to two path disks). The distance between two such centers is $h = \sqrt{8 - 4\sqrt{2}}$. Furthermore, the area of the intersection can be expressed as

$$\gamma = 2 \arccos \frac{h}{2} - 2 \left(\frac{h}{2}\right) \sqrt{1 - \left(\frac{h}{2}\right)^2} \tag{1}$$

which is upper-bounded by 0.45. Therefore $2n\gamma \leq 0.9n \leq n\pi/3$. This gives the proof of one direction.

For the converse (the "if" direction), assume that a solution S for the MACS instance is given with area at least $\pi + L \cdot \mathcal{A}(Z) + \left(\frac{n^3}{10} - \frac{n}{3}\right)\pi$. By our construction, we can modify S, while conserving its connectivity and without diminishing covered area, so that the following properties hold.

(i) If any bonus disk corresponding to a terminal is part of S, so is the cardinal disk corresponding to this terminal.

(ii) The path and cardinal disks in S form a tree; furthermore, such a tree consists of a cardinal disk, a set of segments, and at most one sub-segment. (A *sub-segment* is a subset of a segment, so that it induces a connected component.)

Indeed, if (i) does not hold, then we add in S the missing terminal disk and remove any bonus disk that does not disconnect the solution. Such a bonus disk exists since bonus disks do not contribute to connect the terminals. After this step, the solution remains connected and since the interior of any cardinal disk is disjoint from any other disks, the area covered the new solution has not decreased.

To guarantee (ii), remark that a sub-segment does not contribute to connect terminals. Then, if there are at least two sub-segments, we can remove some disks in the shortest sub-segment and replace them by disks in another sub-segment until it is complete or the shortest sub-segment is empty. After this step, the number of sub-segments decreased by at least one. We repeat this operation until the solution has only one sub-segment. Now, suppose that the path and cardinal disks have a cycle. We can replace one segment of this cycle by any set of disks without disconnecting the solution or creating new cycles. For instance, one can create a new segment incident to exactly one terminal disk of the solution (such a segment always exists). Notice that during this step, no new sub-segments are created. We have established the existence of a solution S for the MACS instance with area at least $\pi + L \cdot \mathcal{A}(Z) + \left(\frac{n^3}{10} - \frac{n}{3}\right)\pi$ that satisfies properties (i) and (ii).

We claim that the number B of bonus disks in S is at least $\frac{n^3}{10} - \frac{9n}{10}$. Suppose not. Using properties (i) and (ii), we observe that the covered area of S can be upper-bounded as

$$\mathcal{A}(S) \leqslant B\pi + \pi + \frac{L|Z| + \frac{n^3}{10} - B}{|Z|}\mathcal{A}(Z). \tag{2}$$

(Here we ignore the possible intersection of a bonus disk with the path disks. The first term is the area covered by bonus disks; the second term is the area covered by a cardinal disk; the third term is the maximum area that can be covered by segments, and possibly the single sub-segment in S). Now, by assumption, the area covered by S is at least

$$\mathcal{A}(S) \geqslant \pi + L \cdot \mathcal{A}(Z) + \left(\frac{n^3}{10} - \frac{n}{3}\right)\pi. \tag{3}$$

The difference between the lower bound (3) and the upper bound (2) is

$$\left(\frac{n^3}{10} - B\right)\left(\pi - \frac{\mathcal{A}(Z)}{|Z|}\right) - \frac{n\pi}{3} \geq \frac{9n}{10}\left(\pi - \frac{\mathcal{A}(Z)}{|Z|}\right) - \frac{n\pi}{3}.$$

Here, in order to reach a contradiction (making the last term greater than 0), we need to calculate $\mathcal{A}(Z)$, which is $(n-3)\pi - (n-5)\gamma'$, where γ' is the area of the intersection of two disks whose centers have distance $1 + \frac{1}{n-5}$. Area γ' is easily shown to be at most 1.25. Therefore,

$$\begin{aligned}0.9n\left(\pi - \frac{\mathcal{A}(Z)}{|Z|}\right) - \frac{n\pi}{3} &\geq 0.9n\left(\pi - \frac{(n-3)\pi - (n-5)1.25}{n-3}\right) - \frac{n\pi}{3}\\ &= \left(0.9 * 1.25 - \frac{\pi}{3}\right)n - \frac{2 * 1.25}{n-3} \geqslant 0.07n - \frac{2.5}{n-3} \geq 0,\end{aligned}$$

which can be verified when $n \geq 8$ by the simple study of a quadratic polynomial.

So we know that S has at least $\frac{n^3}{10} - \delta$ bonus disks, where $\delta \leq n/10$. Ignoring the possible sub-segment of S, S includes a cardinal disk, L' segments and $\frac{n^3}{10} - \delta$ bonus disks. As a result,

$$k = 1 + L|Z| + \frac{n^3}{10} \geq 1 + L'|Z| + \frac{n^3}{10} - \delta \geq 1 + L'|Z| + \frac{n^3}{10} - \frac{n}{10},$$

implying that $n/10 \geq (L' - L)|Z| = (L' - L)(n-3)$. Thus $L' = L$ and the cardinal disks and path disks of S correspond to a tree of length L in the RST instance. This concludes the proof of Lemma 2. □

4 Proof of Theorem 2: The two-by-two algorithm

In this section we present a simple (1/2)-approximation for MACS based on a greedy approach, by iteratively adding two unit disks that maximize the additional area covered while maintaining feasibility. Interestingly, the algorithm that adds disks one at a time is not a constant approximation algorithm. See Figure 4 and the remark at the end of the section for an example. Moreover, trying all possible sets of s disks, for any $s \geq 3$, in the neighborhood of the current solution does not improve the approximation ratio. This can be seen on Figure 3 where the first disk chosen by the algorithm is not x, but x_s. Finally, optimizing the choice of the first disk(s) chosen in the solution does not improve the approximation ratio; see the first remark at the end of the section for details.

Let B_x denote the unit disk centered at $x \in \mathbb{R}^2$ and $B(S) = \bigcup_{x \in S} B_x$ denote the set of points at distance at most one from at least one point in a finite set $S \subset \mathbb{R}^2$ of points. The area covered by a set $C \subset \mathbb{R}^2$ is denoted by $\mathcal{A}(C)$. When $C = B(S)$, its area is simply written as $\mathcal{A}(S)$. Given a graph G, $G[S]$ denotes the subgraph induced by a subset S of vertices. A subset of the vertices of a graph is a *dominating set* if every vertex belongs to the set or is adjacent to some vertex of it.

Algorithm 1: The Two-by-two algorithm for MACS.

Input: $X \subseteq \mathbb{R}^2, k \geq 0$, where X is finite and $k \leq |X|$.
Output: a connected set of size k.

if k *is even* **then**
  $S \leftarrow$ any two intersecting disks of X;
else
  $S \leftarrow$ any one disk of X;
while $|S| \leqslant k - 2$ **do**
  $\{x, x'\} \leftarrow \arg\max \{\mathcal{A}(S \cup \{x, x'\}) \colon x, x' \in X,\ S \cup \{x, x'\} \text{ is feasible } \}$;
  $S \leftarrow S \cup \{x, x'\}$;
return S;

Theorem 2 (Approximation)**.** *The optimization version of* MACS *has a polynomial time algorithm (Algorithm 1) that achieves a* (1/2)*-approximation.*

We assume here that the input set X is connected. Otherwise, one may consider the largest solution over all connected components. For the analysis, we divide the execution of Algorithm 1 in two phases. An iteration belongs to the first phase as long as the current solution S is not a dominating set in the graph UDG(X).

During the first phase, in each iteration, the area covered increases by at least π. During the second phase, since the current solution is a dominating set, any disk can be added while keeping the solution feasible. Therefore, the algorithm is then a standard greedy algorithm to maximize a submodular function, and the analysis is similar to the proof that Nemhauser's algorithm is a $\left(1 - \frac{1}{e}\right)$-approximation for classic submodular functions, showing that the approximation ratio of Algorithm 1 stays greater than 1/2 during the second phase.

Proof. We first analyze the even case where $k = 2\kappa$, and then we reduce the odd case to the even one. Let $S_\kappa = \{x_1, x_2, \ldots, x_{2\kappa}\}$ be the solution returned by the algorithm. Let $S_i = \{x_1, \ldots, x_{2i}\}$ denote the set S right after $|S| = 2i$, and let d be the smallest integer such that S_d is a dominating set in UDG(X). If such an integer does not exist, i.e., S_κ is not a dominating set, then set $d = \kappa$.

Claim 1. *The area $\mathcal{A}(S_d)$ is at least πd.*

Proof. For $i < d$, S_i is not a dominating set. Then there exist two disks y, y' such that $B(S_i) \cap B_y = \emptyset$ and $S \cup \{y, y'\}$ is connected. Adding such a pair increases the area covered by at least $\mathcal{A}(B_y) = \pi$. Since (x_{2i+1}, x_{2i+2}) is chosen to maximize $\mathcal{A}(S_i \cup \{x, x'\})$ among all feasible pairs, $\mathcal{A}(S_{i+1}) \geqslant \mathcal{A}(S_i \cup \{y, y'\}) \geqslant \mathcal{A}(S_i) + \pi$. By induction, $\mathcal{A}(S_d) \geqslant \pi d$. □

Note that when $d = \kappa$, Claim 1 immediately implies that $\mathcal{A}(S_\kappa) \geqslant \frac{\mathbf{OPT}}{2}$. Also remark that, regardless of the initial choice, the area covered by the first two disks is at least π. This observation will be useful when we consider the case where k is odd.

Claim 2. *For all $d \leqslant i < \kappa$, $\mathcal{A}(\boldsymbol{OPT}) \leqslant \mathcal{A}(S_i) + \kappa \cdot (\mathcal{A}(S_{i+1}) - \mathcal{A}(S_i))$.*

Proof. It is easy to check that the function $\mathcal{A}(\cdot)$ satisfies the following properties for all $H \subseteq H' \subseteq X$:

(1) positivity : $\mathcal{A}(H) \geqslant 0$.

(2) monotonicity : $\mathcal{A}(H) \leqslant \mathcal{A}(H')$.

(3) submodularity : $\forall H'' \subseteq X, \ \mathcal{A}(H' \cup H'') \leqslant \mathcal{A}(H \cup H'') - \mathcal{A}(H) + \mathcal{A}(H')$.

Let $\mathbf{OPT} = \{y_1, \ldots, y_{2\kappa}\}$. We have for all $d \leqslant i \leqslant \kappa$:

$$\begin{aligned}
\mathcal{A}(\mathbf{OPT}) &\leqslant \mathcal{A}(S_i \cup \mathbf{OPT}) \\
&= \mathcal{A}(S_i) + (\mathcal{A}(S_i \cup \{y_1, y_2\}) - \mathcal{A}(S_i)) + \ldots \\
&\qquad + (\mathcal{A}(S_i \cup \{y_1, \ldots, y_{2\kappa}\}) - \mathcal{A}(S_i \cup \{y_1, \ldots, y_{2\kappa-2}\})) \\
&\leqslant \mathcal{A}(S_i) + (\mathcal{A}(S_i \cup \{y_1, y_2\}) - \mathcal{A}(S_i)) + \cdots + (\mathcal{A}(S_i \cup \{y_{2\kappa-1}, y_{2\kappa}\}) - \mathcal{A}(S_i)) \\
&\leqslant \mathcal{A}(S_i) + \kappa \cdot (\mathcal{A}(S_i \cup \{x_{2i+1}, x_{2i+2}\}) - \mathcal{A}(S_i)) \\
&= \mathcal{A}(S_i) + \kappa \cdot (\mathcal{A}(S_{i+1}) - \mathcal{A}(S_i)).
\end{aligned}$$

The first and the second inequality respectively come from *monotonicity* and *submodularity*, while the third one follows from the fact that for $i \geq d$ (x_{2i+1}, x_{2i+2}) is the pair of disks maximizing $\mathcal{A}(S_i \cup \{x, x'\})$ among **all pairs** (x, x') in X. As S_d is a connected dominating set in X, all pairs (y_{2j-1}, y_{2j}) for $1 \leqslant i \leqslant \kappa$ are considered. □

We can now re-write Claim 2 as

$$\text{For all } d \leqslant i < \kappa : \mathcal{A}(S_{i+1}) \geqslant \left(1 - \frac{1}{\kappa}\right) \mathcal{A}(S_i) + \frac{\mathbf{OPT}}{\kappa}.$$

Combined with Claim 1, simple algebra yields that, for $d < i \leqslant \kappa$, we have

$$\begin{aligned}
\mathcal{A}(S_i) &\geqslant \left(1 - \frac{1}{\kappa}\right)^{i-d} \mathcal{A}(S_d) + \frac{\mathbf{OPT}}{\kappa} \left(1 + \left(1 - \frac{1}{\kappa}\right) + \cdots + \left(1 - \frac{1}{\kappa}\right)^{i-d-1}\right) \\
&\geqslant \left(1 - \frac{1}{\kappa}\right)^{i-d} \pi d + \frac{\mathbf{OPT}}{\kappa} \cdot \frac{1 - (1 - 1/\kappa))^{i-d}}{1 - (1 - 1/\kappa)} \\
&\geqslant \left(1 - \frac{1}{\kappa}\right)^{i-d} \mathbf{OPT} \cdot \frac{d}{2\kappa} + \mathbf{OPT} \left(1 - \left(1 - \frac{1}{\kappa}\right)^{i-d}\right) \\
&\geqslant \left[1 - \left(1 - \frac{d}{2\kappa}\right) \left(1 - \frac{1}{\kappa}\right)^{i-d}\right] \mathbf{OPT}.
\end{aligned}$$

Therefore, for $i = \kappa$ we have

$$\mathcal{A}(S) = \mathcal{A}(S_\kappa) \geqslant \left[1 - \left(1 - \frac{d}{2\kappa}\right)\left(1 - \frac{1}{\kappa}\right)^{\kappa - d}\right]\mathbf{OPT} = \left[1 - \frac{1}{2}(1+t)\left(1 - \frac{1}{\kappa}\right)^{\kappa t}\right]\mathbf{OPT},$$

where $t = \dfrac{\kappa - d}{\kappa} \in [0, 1]$. As $1 + x \leqslant e^x$ for all $x \in \mathbb{R}$, we get

$$\mathcal{A}(S) \geqslant \left(1 - \frac{1}{2}(1+t)e^{-t}\right)\mathbf{OPT} \geqslant \left(1 - \frac{1}{2}e^t e^{-t}\right)\mathbf{OPT} = \frac{1}{2}\mathbf{OPT},$$

concluding the proof of the case in which k is an even number.

For the case in which k is an odd number, $k = 2\kappa - 1$, in the first iteration, instead of adding two disks to S_1, we add a single disk of X to S_1. This is equivalent to adding two copies of the same disk. This iteration belongs to the first phase, and the only properties we used in the first phase is that each iteration adds an area of π, and keeps the solution feasible; these are clearly true for the first iteration even with one disk. □

Since two disks are guessed at each step, Algorithm 1 has a running time $O(n^3)$, where n is the number of input disks.

Remark. The analysis of Algorithm 1 is tight. For any $\varepsilon > 0$, we construct an input set X as follows. See Figure 3. X contains $x = (0, 0)$ (stripe-shaded disk in Figure 3), $x_i = (2(i-1) + i\varepsilon, 0)$ and $x'_i = ((2+\varepsilon)i, 0)$ for $1 \leqslant i \leqslant k$ (blue disks) and $y_i = (-2i - \varepsilon/2, 0)$ for $0 \leqslant i \leqslant k$ (orange disks). Suppose that $k = 1 + 2\kappa$ is odd and the algorithm starts with $S_0 := \{x, x\}$. Then the algorithm will add $\{x_i, x'_i\}$ in iteration i since it covers more additional area than $\{y_0, y_1\}$. The solution returned (blue disks) covers an area of $\pi + \kappa(\pi + f(\varepsilon)) \approx \frac{k}{2}\pi$, for some function $f(\cdot)$ with $\lim_{\varepsilon \to 0} f(\varepsilon) = 0$, while **OPT** (orange disks) covers an area $k\pi$.

Using similar ideas, one can show that optimizing the choice of the first initial disk(s) does not improve the approximation ratio. To see this, consider several chains of $\lfloor 1/\epsilon \rfloor$ orange disks each, as in Figure 3, and connect them together in such a way that the last disk of one chain almost coincides with the first disk of the next chain. Then, attached to these two disks, add a chain of blue disks as in the figure. Wherever starts the greedy 2-by-2 greedy algorithm on this instance, it will branch to the blue disks as soon as it meets one.

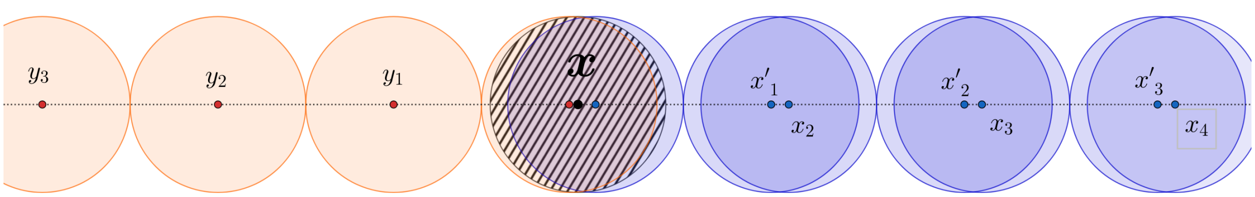


Figure 3: A tight example for Algorithm 1.

Remark. The similar greedy algorithm that adds disks one at a time is not a constant approximation algorithm. See Figure 4. For any $k \geqslant 0$ and $\varepsilon > 0$, consider the input where

$O = (0,0)$, and $y_i = (2(i-1)+\varepsilon, 0)$ for all i. Then, put all $x_1, \ldots, x_k$ evenly spaced (by an angle α) on a circle of radius 2 around O so that none of them intersect y_2. Each light grey region is covered by only one disk x_i so the marginal gain of adding x_i to any solution is at least the area of one of these regions, say $a > 0$. If ε is chosen such that $\mathcal{A}(B_{y_1} \setminus B_O) < a$, then if the algorithm starts by picking disk O, it will then choose all x_j, so that the area covered by the solution is upper-bounded by the area of a radius 3 disk, 9π, while the optimal solution (disks y_i) has area πk.

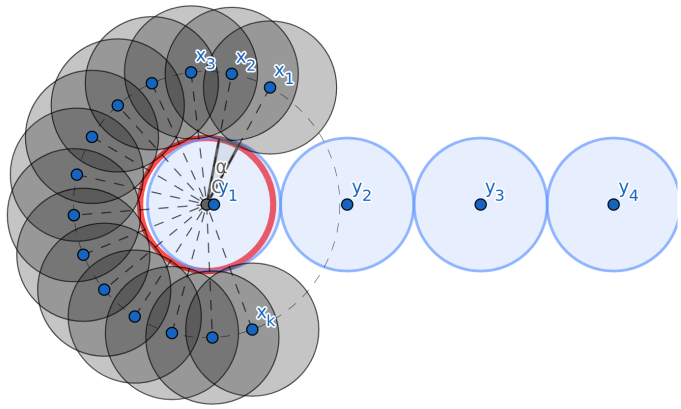


Figure 4: The greedy algorithm that adds only one connected disk maximising the marginal area covered is not a constant factor approximation algorithm.

5 Proof of Theorem 3: PTAS with resource augmentation

In this section we prove our main result.

Theorem 3 (Resource augmentation)**.** *Let $\varepsilon > 0$ be a given parameter. Given a connected set $X \subseteq \mathbb{R}^2$ of n points and a positive integer k, let $\mathbf{OPT}(X,k)$ be the maximum possible area of a connected union of a set of k unit disks centered at a subset of k points of X. Then, there is a deterministic algorithm that computes, in time $n^{O(\varepsilon^{-1})}$, a subset $S \subseteq X$ of size at most k and a set $S_{add} \subseteq \mathbb{R}^2$ of at most εk points, such that* $\mathrm{UDG}(S \cup S_{add})$ *is connected, and the area covered by the unit disks centered at S is at least* $(1-\varepsilon)\mathbf{OPT}(X,k)$.

5.1 Overview of the Method

Our approach is outlined as follows:

- We define the notion of a set of disks having a special recursive property, the "m-guillotine property", which closely follows the notion of *m-guillotine subdivisions*, introduced by Mitchell [28], and utilized in several geometric approximation results. For completeness here, we begin by defining and reviewing m-guillotine subdivisions.
- We prove a structural result (Lemma 4), showing that for any set of k unit disks whose union is connected, there is a set of at most εk "augmentation disks", so that the augmentation disks, together with a subset of the original set of k disks, have a connected union covering at least as much as the original disks, and have the "m-guillotine property".

- We give a dynamic programming algorithm to compute a maximum-area m-guillotine set of at most k input disks, utilizing at most $k' = \lceil \varepsilon k \rceil$ augmentation disks.

5.2 Review of m-guillotine subdivisions

We review some definitions and facts from [28]. Let G be a straight-edge embedding of a planar graph, and let L denote the total Euclidean length of its edge set, E. We can assume (without loss of generality) that G is restricted to the unit square, $\mathcal{U}$ (i.e., $E \subset int(\mathcal{U})$).

Consider an axis-parallel rectangle, a *window*, $W \subseteq \mathcal{U}$. An axis-parallel line ℓ that intersects the interior of W is called a *cut*. The intersection, $\ell \cap (E \cap int(W))$, of a cut ℓ with the restriction of E to the window W consists of a (possibly empty) set of subsegments (possibly singleton points) of ℓ. We let $p_1, \ldots, p_\xi \subset \ell$ denote the ξ endpoints of such subsegments, in order along ℓ. For integer $m \geq 1$, the *m-span*, $\sigma_m(\ell)$, of ℓ with respect to W is defined as follows: If $\xi \leq 2(m-1)$, then $\sigma_m(\ell) = \emptyset$; otherwise, $\sigma_m(\ell)$ is defined to be the line segment, $p_m p_{\xi-m+1}$, joining the mth endpoint, p_m, with the mth-from-the-last endpoint, $p_{\xi-m+1}$. (It may be that $\sigma_m(\ell)$ is the single point p_m, if $\xi = 2m-1$.) We say that the line ℓ is an *m-good cut with respect to W and E* if $\sigma_m(\ell) \subseteq E$. (Note that ℓ is trivially m-good if $\xi \leq 2(m-1)$.) The edge set E is said to *satisfy the m-guillotine property with respect to rectangle W* if either (1) no edge of E lies (completely) interior to W; or (2) there exists a cut ℓ, that is m-good with respect to W and E, such that ℓ splits W into W_1 and W_2, and, recursively, E satisfies the m-guillotine property with respect to both W_1 and W_2.

The structural result of [28] states that any set E of edges is "very nearly m-guillotine" in a precise sense: One can augment the edge set E with a set of new (axis-parallel) segments ("m-spans"), of total length $O(1/m)$ times the total length of E, such that the augmented edge set is m-guillotine. Proof of this fact is based on a charging argument that utilizes the notion of "m-darkness": a point $p \in W$ is *m-dark with respect to horizontal cuts of W* if the vertical rays going upwards/downwards from p each cross at least m edges of E before reaching the boundary of W. As in [28], the length of the m-dark portion of a cut is the "chargeable" length of the cut, in that one can charge the m-dark portion of the cut to the m levels of E on each side of the cut that become "exposed" after making cut ℓ: After a vertical cut along a line ℓ, the left side of a portion, e, of a vertical edge of E that lies to the right of ℓ becomes exposed after the cut along ℓ if a leftward horizontal ray with endpoint on e crosses fewer than m (vertical) edges before crossing ℓ. (Once a portion of an edge of E is charged, because of a cut, on one of its two sides (an amount equal to $1/2m$ of the portion's length), and that portion becomes exposed by the cut, that side of that portion is never charged again, by a cut deeper in the recursive partitioning, since once it is exposed to the boundary of a window, it remains exposed.)

Given an edge set E of a connected planar graph G, if E is not already satisfying the m-guillotine property with respect to W, then the following lemma of [28] (reproduced here, with proof, for completeness) shows that there exists a "favorable cut" for which we can afford to charge off (to the edges of E) the construction of any m-span that must be added to E in order to make the cut m-good with respect to W and E. To be precise, a cut ℓ is said to be *favorable* if the length of its m-dark portion (its chargeable length) is at least as large as the length $|\sigma_m(\ell)|$ of the m-span (the cost of making cut ℓ).

Lemma 3. *([28]) For any edge set E and any (rectangle) window W, there is a favorable cut. If E consists of a set of axis-parallel line segments, then there is a favorable cut (horizontal or vertical) that passes through an endpoint of a segment of E.*

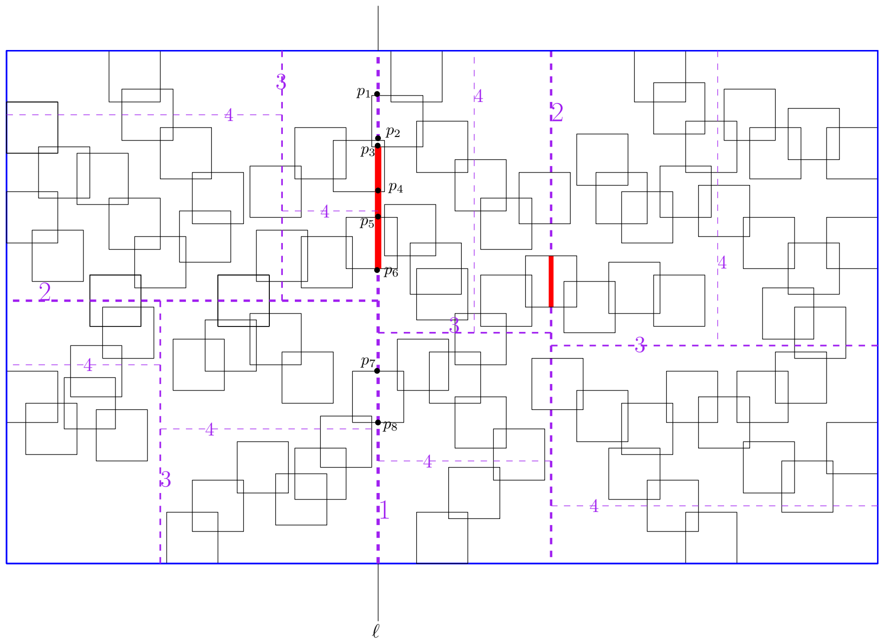


Figure 5: Example of an m-guillotine subdivision, with $m = 3$. The bold (red) segments are m-spans that are added to the edge set to make it m-guillotine. The first 4 levels of cuts are shown. Numbers "1", "2", etc by dashed (purple) cut lines indicate the level of the cut, with level "1" being the root (first cut).

Proof. Let $f(x) = |\sigma_m(\ell_x)|$ be the length of the m-span for the vertical cut, ℓ_x, along the vertical line through the point $(x, 0)$. Then, $f(x)$ is the cost of making cut ℓ_x, and its integral, $A_x = \int_0^1 f(x)dx$, is the area of the region, R_x^m, of points of $\mathcal{U}$ that are m-dark with respect to horizontal cuts. Similarly, let $g(y)$ be the cost of making horizontal cut ℓ_y along the horizontal line through the point $(0, y)$, and let $A_y = \int_0^1 g(y)dy$ be the corresponding area of the region of points in $\mathcal{U}$ that are m-dark with respect to vertical cuts. One of the two areas, A_x or A_y, is larger than the other; assume, without loss of generality, that $A_x \geq A_y$. Now the area A_x of the region R_x^m can be computed by integrating with respect to y: specifically, $A_x = \int_0^1 h(y)$, where $h(y)$ is the chargeable length of the horizontal cut ℓ_y through y. By the assumption that $A_x \geq A_y$, we get that $\int_0^1 h(y)dy \geq \int_0^1 g(y)dy \geq 0$; thus, it cannot be that $h(y) < g(y)$ for all values of $y \in [0, 1]$, so there must be a value $y = y^*$ for which $h(y^*) \geq g(y^*)$. This shows that there is a horizontal cut ℓ_{y^*} that is favorable. (If, instead, we had $A_x \leq A_y$, then we would have the existence of a *vertical* cut that is favorable.) Refer to Figure 6. Finally, we remark that if the segments E are axis-parallel (as in the example of Figure 6), then the functions f, g, and h are piecewise-constant, with discontinuities corresponding to the x- and y-coordinates of the endpoints of the line segments E. This implies the existence of a value $y = y^*$ for which $h(y^*) \geq g(y^*)$ at one of these discontinuities. $\square$

The charging scheme does the following, in the case of a vertical favorable cut (as in the example of Figure 6): For each unit of length of the cut within the "blue" region, we assign half of the unit to the right (distributed evenly over the m levels of edges we know, by m-darkness, to exist to the right), and similarly assign half of the unit to the left, distributed over m levels. (Here, and in Figure 6, we consider the case in which edges E are axis-parallel, since they are sides of axis-aligned

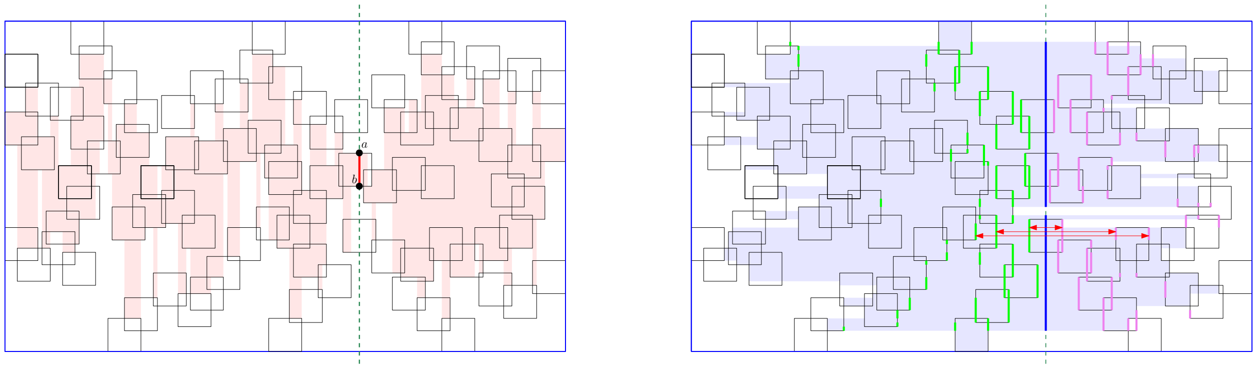


Figure 6: In this example, the edge set E consists of the line segments bounding a set of squares. We take $m = 3$. The region R_x^m is shown in red (on the left); the region in blue (on the right) are those points that are m-dark with respect to vertical cuts. On the left, we show a vertical cut (dashed, green), with its corresponding m-span highlighted in red; the same cut is shown on the right, with the m-dark portion highlighted in blue, and the offloading of the length of the blue chargeable length to the m levels of segments on either side of the cut indicated with magenta highlight on portions of E charged from the left, and with green highlight on portions of E charged from the right.

squares; [28] discusses the slightly more general case of arbitrarily oriented edges.) In total, each unit of length of an edge of E is charged at most $1/2m$ from each of its two sides, implying that the total charge overall (i.e., the total length of all m-spans added along favorable cuts) is at most $1/m$th of the sum of the lengths of the edges in E. (In Figure 6, the shown vertical cut results in vertical edge portions that are charged on their left (resp., right) sides, which are highlighted in magenta (resp., green).) Because we only offload charge along portions of cuts that are m-dark, and we offload charge to sides of edges of E that lie within m levels of the boundary in the new subproblems after the cut, no portion of E is ever re-charged. The result is the following structure theorem.

Theorem 6. *([28]) Let G be an embedded connected planar graph, with edge set E consisting of line segments of total length L. Then, for any positive integer m, there exists an edge set $E' \supseteq E$ that obeys the m-guillotine property with respect to the axis-aligned bounding box of E, and for which the length of E' is at most $(1 + O(1/m))L$.*

5.3 Defining the m-guillotine property for a set of disks

We now define what it means for a set of unit *disks* to have the "m-guillotine property". The notion is very similar to that of an edge set E having the m-guillotine property, as defined previously.

Let $X^* \subseteq X$ be an optimal (for MACS) subset of $k = |X^*|$ centers of unit disks, the union of which is connected and of maximum possible area $\mathbf{OPT}(X, k)$. Let $Z = \{(x, y) : x = i/2, y = j/2,$ for some integers $i, j\}$ be the set of points in the plane having half-integral coordinates. Our approximation algorithm will select augmentation disks centered at a subset of Z.

Let $\mathcal{Q} = \{Q_1, \dots, Q_k\}$ be the set of k axis-aligned bounding squares of the unit disks centered at the points X^*. Let E be the set of $4k$ axis-parallel line segments (each of length 2) that bound the squares $\mathcal{Q}$. The union of the segments E is connected, since the union of the (equal-size) squares $\mathcal{Q}$

is connected. Lemma 4 implies that the edge set E can be made to have the m-guillotine property, for any positive integer m, with the addition of m-spans (horizontal/vertical line segments, defined by coordinates of the endpoints of the segments E) whose total length is at most $O(k/m)$; with the appropriate choice of $m = \Theta(1/\varepsilon)$, the total length of all added m-spans is thus at most εk.

First, if an m-span has length less than 2, then any square that is contained in the corresponding window and that is intersected by the m-span is already among the squares that intersect the cut that are accounted for among the first m or last m edges crossed by the cut. Thus, the total number of squares intersected by the cut is at most $2m$, and we can afford to ignore this short m-span, since our goal is to have $O(m)$ information specified across a partitioning cut. Thus, we can assume that all m-span segments are of length at least 2.

Now, associated with each (remaining) horizontal/vertical m-span segment, ab, we define an *m-span rectangle*, which is axis-parallel, centered on ab, of width 2; i.e., if ab is vertical, with $a = (x_a, y_a)$ and $b = (x_b = x_a, y_b)$, the corresponding m-span rectangle is $[x_a - 1, x_a + 1] \times [y_a, y_b]$. It is readily seen that the m-span rectangle associated with ab is covered by a set of $O(|ab|)$ unit disks with centers at half-integral points Z; thus, the set of all m-span rectangles is covered by a set of $O(\varepsilon k)$ augmentation disks, centered at points of Z. Refer to Figure 7. The purpose of the m-span rectangle is to allow us, in a dynamic program, to decouple the subproblems on each side of a cut: any unit disk centered at a point of X to the right of a vertical cut contributes nothing to the union of disks centered at points left of the cut, unless it is one of the $O(m)$ specified disks crossing the cut, since, if it crosses an m-span segment on the cut, the m-span rectangle fully covers it, so that the augmentation disks fully cover it as well. Since the augmentation disks fully cover the m-span rectangle, they connect all disks, centered on points of X, that intersect the m-span rectangle.

Let U be a finite set of unit disks centered at points of $X \cup Z$. We now define the notion of a set of disks satisfying a property that we call the "m-guillotine property". An axis-parallel cut line ℓ is *m-good with respect to the set U of unit disks and an axis-aligned rectangle window W* if (1) $\ell \cap W$ intersects at most $2m$ disks of U that are centered at points of X; and (2) U includes all unit disks centered at points of Z that lie within the m-span rectangle associated with the m-span of ℓ with respect to the edge set E of segments bounding the axis-aligned bounding squares of unit disks of U that are centered at points of X. An m-good cut has a succinct specification of those disks of U that are intersected by the cut: $O(m)$ disks of U centered at points of X, together with at most one m-span segment (rectangle), which specifies the set of all augmentation disks (centered at half-integral points Z) within $\mathcal{U}$ that intersect the cut.

We say that a set U of unit disks centered at points of $X \cup Z$ *satisfies the m-guillotine property with respect to (axis-aligned) rectangle W* if either (1) no disk of U lies (completely) inside W; or (2) there exists an axis-parallel cut line ℓ that is m-good with respect to U and W, such that ℓ splits W into W_1 and W_2, and, recursively, U satisfies the m-guillotine property with respect to W_1 and with respect to W_2. We say that U *satisfies the m-guillotine property* if U satisfies the m-guillotine property with respect to the axis-aligned bounding rectangle of U.

5.4 A key structural lemma

The following key structural lemma allows us to prove our main claim, since it shows that an arbitrary set of input disks (e.g., an optimal set, centered at points $X^* \subseteq X$ that are an optimal solution to an instance of MACS) can be converted to a covering set of disks that satisfies the m-guillotine property, with only a small (factor $(1 + O(1/m))$) increase in the total number of disks.

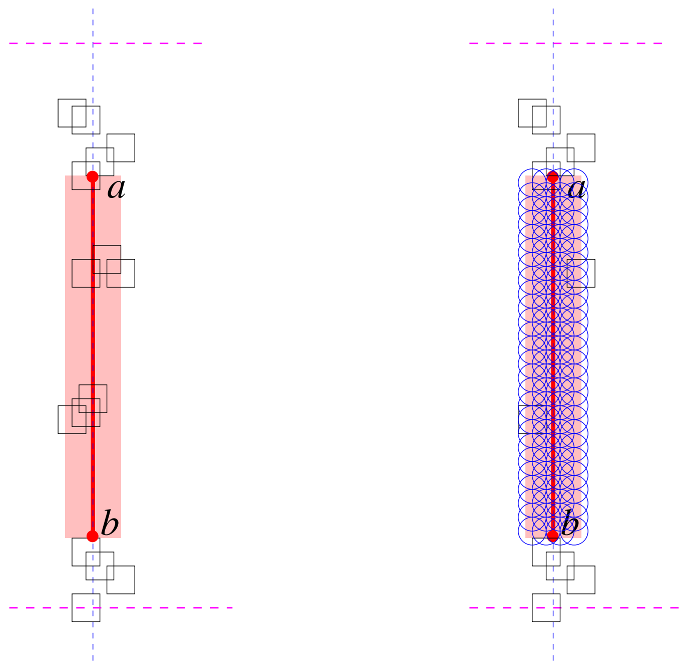


Figure 7: Left: A vertical cut, the m-span segment ab (for $m = 5$), and the associated m-span rectangle. Right: The set of augmentation unit disks centered at half-integral points of Z within the m-span rectangle; the augmentation disks cover the m-span rectangle and thus completely cover those squares bounding unit disks whose intersection with ab lies interior to ab (the 4 such squares shown on the left have been removed in the figure on the right).

Lemma 4. *For any positive integer m and finite set U of k unit disks, there exists a set U' of unit disks centered at (half-integral) points of Z, and a subset $U'' \subseteq U$, such that $U'' \cup U'$ satisfies the m-guillotine property, $|U'| = O(k/m)$, and the union of the disks $U'' \cup U'$ covers the union of the disks U.*

Proof. Theorem 6 implies that the edge set E, of edges of the bounding squares of the k unit disks U, can be made to be an m-guillotine subdivision through the addition of m-spans (horizontal/vertical line segments) whose total length is at most $O(k/m)$.

Now, each recursive (axis-parallel) cut, within a rectangular window W, in the associated m-guillotine hierarchy gives rise, potentially, to an m-span segment ab, which has an associated m-span rectangle, which is covered completely by the $O(|ab|)$ unit augmentation disks centered at half-integral points Z within the m-span rectangle (refer to Figure 7); the set U' consists of all such augmentation disks. Every disk of U whose intersection with ab lies interior to ab is fully covered by the m-span rectangle and, therefore, by the union of the augmentation disks U'; thus, such disks of U can be removed (while maintaining the area of coverage and the connectivity), and we define $U'' \subseteq U$ to be the remaining subset of the disks of U. For the set $U' \cup U''$ we have that each cut is m-good, since (1) it intersects at most $2m$ of the disks (centered at points of X) of U'', and (2) the augmentation disk set U' (centered at points of Z) includes all unit-radius disks centered at points of Z that lie within the m-span rectangle. Since the total length of all m-spans in the hierarchy is at most $O(k/m)$, the total number of augmentation disks U' is also $O(k/m)$ (recall that the length of each m-span is at least 2). Furthermore, the set $U'' \cup U'$ of disks has the m-guillotine property, as evidenced by the same set of hierarchical cuts, each of which is m-good, that realize the m-guillotine subdivision (whose existence comes from Theorem 6) for the edge set E (the axis-parallel segments bounding the unit disks U). □

5.5 The dynamic programming algorithm

Proof. To complete the proof of Theorem 3, we now provide a dynamic programming algorithm to compute, for given positive integers k, k', and m, and an input set of points X for which UDG(X) is connected, a set U of unit disks centered at points of $X \cup Z$ such that (i) U satisfies the m-guillotine property, (ii) U has at most k disks centered at X and at most k' centered at points of Z, and (iii) the union of the disks of U has the maximum possible area among all sets of disks satisfying (i) and (ii). The application of this algorithm, with $k' = O(k/m)$, yields the claimed PTAS, since we know, by Lemma 4 applied to a MACS-optimal set U of k disks, that, among the m-guillotine sets of disks over which the dynamic program optimizes, there is such a set that includes a MACS-optimal set of k disks.

The dynamic program proceeds in much the same way that similar algorithms are used to compute optimal m-guillotine subdivisions for TSP and other problems [2, 5, 8, 9, 10, 11, 12, 28, 31]. Subproblems will be specified by axis-aligned rectangles, W, the coordinates of which come from the left/right/top/bottom coordinates of the n input disks; specifically, we let $x_1 \leq x_2 \leq \cdots \leq x_{2n}$ and $y_1 \leq y_2 \leq \cdots \leq y_{2n}$ denote the sorted coordinates. The optimization of a subproblem is to select an axis-parallel cut, partitioning the rectangular window into two, along with the $O(m)$ data associated with the cut, including the $O(m)$ unit disks centered at points of X that intersect the cut, the connection requirements ($O(1)$, for fixed m) for the two new subproblems, and the defining coordinates of an m-span rectangle (if any), which succinctly encodes the set of augmentation disks centered at half-integral points of Z within the m-span rectangle. Overall, the approximation to the

MACS with augmentation will be given by the optimal solution of a subproblem associated with a root rectangle, W_0; there are only $O(n^4)$ possible choices of W_0, and these include the axis-aligned bounding box of an (exact) MACS-optimal set of k disks.

A subproblem, $(W, \mathcal{B}, \kappa, \kappa')$, is specified by (1) a rectangle $W \subseteq W_0$, with coordinates among the x_i's and y_j's, (2) a specification of certain *boundary information*, $\mathcal{B}$, that gives the information necessary to describe how the solution inside W interfaces with the solution outside of the window W, (3) a cardinality $\kappa \leq k$ indicating how many unit disks centered at input points X are to lie fully within the rectangle W (without intersecting its boundary), and (4) a cardinality $\kappa' = O(k/m)$ indicating how many unit (augmentation) disks centered at points Z are to lie fully within W. The boundary information $\mathcal{B}$ includes the following:

(a) For each of the four sides of W, we specify at most $2m$ unit disks, centered at input points X, that intersect the side. Additionally, each side can have one segment specified, which specifies the associated m-span rectangle, with the corresponding set of all unit (augmentation) disks centered at points of Z that lie within the rectangle. There are $n^{O(m)}$ choices for this information.

(b) We specify connectivity information by specifying a partition of the set of $O(m)$ boundary elements (disks centered at points of X, and clusters of augmentation disks covering the at most four m-span rectangles on the boundary of W), with each subset of the partition indicating which boundary elements are connected outside of W. Knowing what is connected outside of W implies the connectivity requirements within W for the subproblem. (If there are no boundary elements associated with W (meaning that $W = W_0$ is one of the choices of a root rectangle), then the connectivity requirement is simply that all disks within W must have a connected union.) Since the number of different partitions of the $O(m)$ boundary elements is purely a function of m, considered to be a constant, there are only a constant number, $\chi(m)$, of choices of connectivity information.

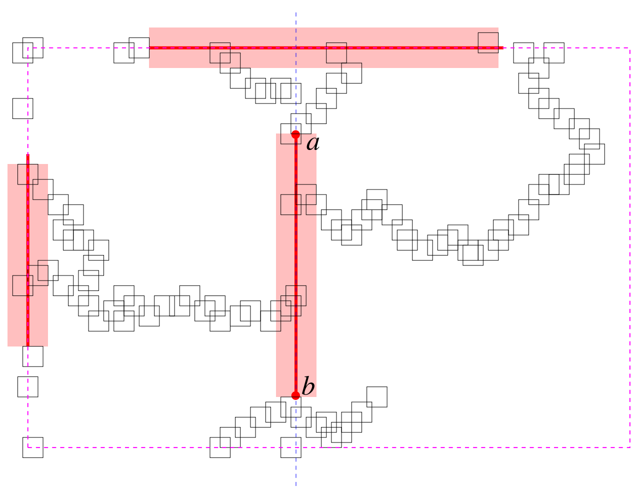


Figure 8: A subproblem in the dynamic program that optimizes over sets of disks that satisfy the m-guillotine property.

In total, then, the number of different subproblems is the product of the number of different choices for W, κ, and κ' ($O(n^4k^2)$) and the number of choices for boundary information (a) and (b), namely, $n^{O(1/m)} \cdot \chi(m) = n^{O(1/m)}$, since $m = O(1/\varepsilon)$ is a constant.

Let $f(W, \mathcal{B}, \kappa, \kappa')$, with $\kappa \leq k$ and $\kappa' \leq k' = O(k/m)$, denote the value of subproblem, the maximum area of the intersection of W with the union of the disks in a set U of unit disks that satisfy the following properties:

(a) U satisfies the m-guillotine property with respect to W; and,

(b) U satisfies the subproblem $(W, \mathcal{B}, \kappa, \kappa')$, obeying the boundary information $\mathcal{B}$, including the connectivity requirement, and has κ unit disks of U that are centered at input points X and lie fully within W and κ' unit disks of U that are centered at points of Z and lie fully within W.

Note that for a disk that lies only partly within the rectangle W, and thus is specified as part of the boundary information $\mathcal{B}$, the quantity $f(W, \mathcal{B}, \kappa, \kappa')$ counts the area of only that portion of the disk that lies within the rectangle W, since it is defined to be the area of the intersection of W with the union of the disks; this accounting assures that in the recursion (below), when W is cut into two subrectangles, there is no over- or under-counting of area. In order to tabulate values of f, we build up the solutions bottom-up, as usual, starting with subproblems that are trivial, and tabulating values corresponding to window W, defined by rectangle $[x_i, x_{i'}] \times [y_j, y_{j'}]$ with $i < i'$ and $j < j'$, in order of increasing values of $(i' - i)$ and $(j' - j)$, with each choice of boundary information, κ, and κ'. The value $f(W, \mathcal{B}, \kappa, \kappa')$ is computed recursively:

$$f(W, \mathcal{B}, \kappa, \kappa') = \max_{\xi, \mathcal{B}_\xi} \{ f(W_1, \mathcal{B}_1, \kappa_1, \kappa'_1) + f(W_2, \mathcal{B}_2, \kappa_2, \kappa'_2) \},$$

where ξ is an axis-parallel cut (at one of the discrete coordinates x_i, y_j), $\mathcal{B}_\xi$ is the boundary information across the cut ξ (including specification of κ_ξ unit disks, centered at points of $X \cap W$, that cross ξ and (possibly) specification of an m-span rectangle that contains κ'_ξ half-integral points of $Z \cap W$ where augmentation disks are centered), W_1 and W_2 are the subrectangles of W obtained when making cut ξ, $\mathcal{B}_1$ and $\mathcal{B}_2$ are boundary information consistent with $\mathcal{B}$ and $\mathcal{B}_\xi$, κ_1 and κ_2 satisfy $\kappa = \kappa_1 + \kappa_2 + \kappa_\xi$, and κ'_1 and κ'_2 satisfy $\kappa' = \kappa'_1 + \kappa'_2 + \kappa'_\xi$. The base cases of the recursion are the values $f(W, \mathcal{B}, 0, 0)$, which are readily found by computing the union of the disks specified in the boundary information $\mathcal{B}$ and then computing the area of intersection of the union with the rectangle W. (There are at most $\kappa + \kappa' = O(k)$ disks of U in total, and the union of interest in the base case consists of $O(m)$ disks centered at points of X and at most 4 sets of augmentation disks centered at Z, each specified by an m-span segment on one of the four sides of W.) Notice that in case that if the union of the disks is disconnected (i.e. the subproblem is infeasible), the value of $f(W, \mathcal{B}, 0, 0)$ is set to $-\infty$.

The correctness of the recursion is argued as follows. Assuming we are not in the base case, for an optimal set U that satisfies the subproblem $(W, \mathcal{B}, \kappa, \kappa')$ and achieves the area $f(W, \mathcal{B}, \kappa, \kappa')$, there must be, by definition of being m-guillotine with respect to W, an axis-parallel cut, ξ, that is m-good with respect to U and W. This cut ξ splits W into two subrectangles, W_1 and W_2, and must, by definition of being m-good, intersect a number, $\kappa_\xi \leq 2m$ (and $\kappa_\xi \leq \kappa$), of disks of U centered at $X \cap W$ and a number, $\kappa'_\xi \leq \kappa'$ of disks of U centered at $Z \cap W$, corresponding to a set of augmentation disks that cover an m-span rectangle centered on a subsegment of ξ within W. In order for U to be optimal for subproblem $(W, \mathcal{B}, \kappa, \kappa')$, then U must, recursively, be optimal for the two new subproblems $(W_1, \mathcal{B}_1, \kappa_1, \kappa'_1)$ and $(W_2, \mathcal{B}_2, \kappa_2, \kappa'_2)$, for κ_1, κ'_1, κ_2, κ'_2 satisfying $\kappa = \kappa_1 + \kappa_2 + \kappa_\xi$

and $\kappa' = \kappa'_1 + \kappa'_2 + \kappa'_\xi$, and for boundary information specifications $\mathcal{B}_1$ and $\mathcal{B}_2$ that are compatible with $\mathcal{B}$ and with the connectivity requirements that each of the two subproblems inherit across the cut ξ. Since the main recursion optimizes over all choices of ξ and $\mathcal{B}_\xi$, it includes the option to partition the subproblem into the two subproblems according to the cut that is guaranteed to exist by the m-guillotine property of the optimal set U with respect to W.

The overall solution to the problem is given by $f(W_0, \mathcal{B}_0, k, k')$ for W_0 chosen to be one of the $O(n^4)$ possible root rectangles, $\mathcal{B}_0$ specifying no crossed disks or augmentation disks for the boundary of W_0 (so that all disks are interior to W_0), k equal to the input parameter of the MACS instance, $k' = ck/m$ for a fixed constant c, and connectivity specifying that the union of all disks interior to W_0 must be connected.

The number of subproblems is $n^{O(m)}$, and the evaluation of $f(W, \mathcal{B}, \kappa, \kappa')$ for any one subproblem requires time $n^{O(m)}$ to optimize over all choices of cuts and boundary specifications. Thus, the overall running time is $n^{O(m)}$, which is $n^{O(1/\varepsilon)}$, with the choice of $m = O(1/\varepsilon)$. This concludes the proof of Theorem 3. □

Remark. The proof of Theorem 3 can be extended to a slightly more general setting of MACS than that of unit disks, namely, the case of input regions that have "fat" bounding boxes of comparable sizes, with the possibility of adding a small number ($O(\varepsilon k)$) of augmentation disks of comparable size (diameter). More precisely, consider a given set $\mathcal{R} = \{R_1, \ldots, R_n\}$ of n connected regions in the plane, each having a boundary consisting of a union of a constant number of algebraic curves, with a connected union $\bigcup_i R_i$, and assume that each region R_i has an axis-aligned bounding box, $BB(R_i)$, whose aspect ratio is at most ρ; i.e., the ratio of the length of the longer side of $BB(R_i)$ to the length of the shorter side of $BB(R_i)$ is at most ρ. Further, we assume that the sizes (diameters) of the regions R_i are all about the same, within a constant factor: the ratio $diam(R_i)/diam(R_j)$ is bounded by a constant. Then, if we allow at most εk augmentation disks of size $\Theta(\max_i diam(R_i))$, then the PTAS we described for regions that are unit disks generalizes immediately to this case. The structural result holds as before: the set E of edges of bounding boxes of the regions R_i can be made to be an m-guillotine subdivision with the addition of m-span segments of lengths totalling $O((1/m)\sum_i diam(R_i)) = O((k/m)\max_i diam(R_i))$. Each of the m-span segments ab yields an m-span rectangle, of width $\max_i diam(R_i)$, centered on it, which can be covered by $O(|ab|/\max_i diam(R_i))$ augmentation disks of size $\max_i diam(R_i)$. Thus, using only $O(\varepsilon k)$ augmentation disks, an optimal set of regions, with area-maximizing connected union, can be made to have an m-guillotine property (defined analogously to the case of input disks). (Note that the augmentation disks have diameters comparable to the diameters of the regions R_i, but they may have have much greater areas than the regions R_i, which may have areas far smaller than the areas of their bounding boxes, unless an assumption is added that the regions R_i themselves are also "fat" according to common definitions of fatness of convex or nonconvex regions. Note too that the fatness assumption for the bounding boxes of the regions depends on the orientation of the coordinate axes, unless the regions are themselves fat.) The dynamic programming algorithm then optimizes, as before, over sets of regions having the m-guillotine property.

5.6 Cases in which no augmentation is needed

If the input set X is α-well-distributed (Definition 3), then we obtain a PTAS for MACS, even without resource augmentation; we restate Theorem 4 here, and then give its proof.

Theorem 4 (α-well-distributed)**.** *The* MACS *on α-well-distributed inputs (for a constant α) has a polynomial-time approximation scheme.*

Proof. We utilize the dynamic programming algorithm that yields the PTAS with augmentation (Theorem 3). Specifically, we use the algorithm to find an m-guillotine set, U, of at most $\lceil Ck \rceil$ input disks, for an appropriate choice of $C < 1$, and a set, U', of $k' = O(\varepsilon k)$ augmentation disks, so that the union of the $Ck + O(\varepsilon k)$ disks is connected, with the area of the union of all disks, $U \cup U'$, at least $(1 - O(\varepsilon))\mathbf{OPT}(X, \lceil Ck \rceil)$. Recall that the augmentation disks are centered at grid points of Z that fully cover the width-2 rectangle centered on the m-span segment, with at most one such segment per side of a window W, which interconnect the disks that intersect the m-span (determined by the bounding boxes of the input disks). We argued that the solution uses in total only $O(\varepsilon k)$ augmentation disks to cover the m-span rectangles centered on m-span segments that arise in the specifications of the subproblems that make up the hierarchical subdivision, since the total length of all such segments needed to convert an input set of disks into an m-guillotine such set is $O(k/m) = O(\varepsilon k)$, for $m = \lceil 1/\varepsilon \rceil$ (see the proof of Lemma 4).

Consider an m-span segment ab on the boundary of a subproblem in the computed solution from the dynamic programming algorithm. The computed solution includes a set of $O(|ab|)$ augmentation disks (in the set U') that cover the m-span rectangle (of width 2) centered on ab. If we simply remove these augmentation disks, we may disconnect the overall union of disks. Thus, we consider the set $\{u_1, u_2, \ldots, u_K\} \subseteq U$ of input disks in the computed solution that intersect the m-span rectangle (of width 2, centered on ab), with centers on one side of ab (say, left of the oriented line through ab), indexed in order of the projections of their center points onto the (horizontal/vertical) line containing segment ab, and, for each consecutive pair, (u_i, u_{i+1}), we do the following: If u_i and u_{i+1} are disjoint (implying that their projected center points are separated by at least some constant $\Omega(1)$), then we add to U a set, $U_{u_i,u_{i+1}}$, of input disks that form a path in UDG(X) from the center of disk u_i to the center of disk u_{i+1}. By the α-well-distributed property, the number, $|U_{u_i,u_{i+1}}|$, of disks added is at most a constant (depending on α) times the distance between the center points of u_i and u_{i+1}, and thus at most a constant times the distance between the consecutive projected center points of u_i and u_{i+1}; thus, in total, the number of added input disks is $O(|ab|)$. These added input disks interconnect all of the disks of U that intersect the m-span rectangle (of width 2, centered on the segment ab) on one of its sides; similarly, we add input disks to interconnected the disks of U that intersect the m-span rectangle on the other side of ab, and, if there are disks of U on both sides of ab, we add also a set of (at most $O(|ab|)$) input disks interconnecting an input disk on one side to an input disk on the other side. In total, then, we have added enough input disks to assure that all disks of U that intersect the m-span rectangle centered on the segment ab are connected in the union of the newly enlarged set of input disks, yet we have added only $O(|ab|)$ such disks, exploiting the α-well-distributed property of X. This allows the augmentation disks covering the m-span rectangle to be removed, while maintaining connectivity. In total, then, the number of original disks U produced by the algorithm, together with the added disks that ensure connectivity, is $(1 + O(\varepsilon))Ck$, which is at most k, for an appropriate choice of constant $C < 1$. The union of these disks is connected and has area is at least $(1 - O(\varepsilon))\mathbf{OPT}(X, \lceil Ck \rceil)$, which, according to Lemma 5 below, is at least $(1 - O(\varepsilon))(1 - \varepsilon c)\mathbf{OPT}(X, k)$, which is $(1 - O(\varepsilon))\mathbf{OPT}(X, k)$. □

The following lemma, utilized in the proof of Theorem 4, shows that, for small values of $\varepsilon > 0$, we can use slightly less than k input disks (specifically, at most $\lceil k/(1+\varepsilon) \rceil$ of the input disks) to achieve a connected union with area very close to $\mathbf{OPT}(X, k)$ (specifically, area at least $(1 - \varepsilon c)\mathbf{OPT}(X, k)$, for some constant c).

Lemma 5. *For any integer $k \leq n$ and $\varepsilon > 0$, there is a constant c such that* $\mathbf{OPT}(X, \lceil k/(1+\varepsilon)\rceil) \geq (1-\varepsilon c)\mathbf{OPT}(X,k)$.

Proof. Let $X^* \subseteq X$ be an optimal (for MACS) subset of $k = |X^*|$ centers of unit disks, the union of which is connected and of maximum possible area $A = \mathbf{OPT}(X,k)$. Then, we will show that, for any $\varepsilon > 0$, there is a subset $X' \subseteq X^*$ of cardinality at most $k' = \lceil k/(1+\varepsilon)\rceil$ such that UDG(X') is connected and the area of the union of the unit disks centered at X' is at least $(1-\varepsilon c)\mathbf{OPT}(X,k)$, for some constant c. This will show the claim.

We will obtain X' from X^* by removing from X^* a subset of $k - k'$ points; we do so in a way that keeps the union area "large" while keeping the UDG of the remaining points connected.

Consider a uniform grid that partitions the plane into squares (*pixels*) of side length $1/2\sqrt{2}$. A pixel is said to be *occupied* if a point of X^* lies within it; let M be the number of occupied pixels. By the choice of pixel size, and the fact that UDG(X^*) is connected, we know that the union of occupied pixels is connected.

Case (1). If $M \leq k'$, then we "mark" (for retention) one point of X^* in each of the occupied pixels, (these marked points will be included in X'), and then iteratively remove $k-k'-M$ unmarked points, selecting at each stage an (unmarked) point of X^* whose deletion minimizes the decrease in area of the union of disks. The amount by which the area of the union of the unit disks decreases when a disk D is removed is exactly the area of the set of points in the plane that are "uniquely covered" by the disk D (i.e., the points of the plane covered by D, but not covered by any other disk centered on a remaining point of X^*). Initially, among the $k - M$ unmarked points of X^*, there must be one whose unit disk uniquely covers area at most $A/(k-M)$ (otherwise, the $k - M$ disks collectively cover area greater than A); then, after its removal, there must be one whose unit disk uniquely covers area at most $A/(k-M-1)$, etc. Continuing, the total area removed, over the $k-k'-M$ stages, is at most $A(1/(k-M)+1/(k-M-1)+\cdots+1/(k'+1)) \leq A(k-k'-M)/(k'+1) \leq A(k-k')/k' \leq \varepsilon A$. Thus, the area of the union of the unit disks centered on the remaining k' points of X^* is at least $(1-\varepsilon)A$.

Case (2). If $M > k'$, then we consider any connected subset of the M occupied pixels, then we remove points within a leaf pixel (of spanning tree in dual of grid graph), and continue doing this until there are only k' occupied pixels. This removes points from $M - k'$ pixels. The union of unit disks with centers within any one pixel has area at most $\beta = \pi + \sqrt{2} + (1/8) \approx 4.68$, as a simple calculation shows. Thus, the removal of the points results in a decrease in the area of the union of disks of at most $\beta(M-k') \leq \beta(k-(k/(1+\varepsilon))) \leq \beta\varepsilon k/(1+\varepsilon)$. Since $M > k' \geq k/(1+\varepsilon)$, and since $A \geq M/8$ (each pixel has area 1/8, and all occupied pixels are fully covered), we know that $k \leq 8(1+\varepsilon)A$. Thus, the total area of the disks centered at the removed points is at most $8\beta\varepsilon A$, and the area of the union of the disks centered at the remaining points, which lie within k' occupied pixels, is at least $(1-8\beta\varepsilon)A$. Since there are now only k' occupied pixels, we are in case (1), and can thus further reduce the number of points to at most k', while keeping the area of the union of disks at least $(1-O(\varepsilon))A$. □

One intuitive view of a well-distributed input is to look at the shape of the "holes" of the input, that are the different connected components of the complement of the union of the input disks in the plane. The assumption of well-distribution means that these holes are, in a sense, *fat*. One particularly interesting case arises when there are no holes at all in the union of the input disks. We call these sets *pseudo-convex*, and we prove that this is a special case of well-distributed inputs.

Definition 4. *A set X is called* pseudo-convex *if the convex-hull of X is covered by the union of the unit disks centered at points of X.*

Lemma 1. *A pseudo-convex set $X \subseteq \mathbb{R}^2$ is 3.82-well-distributed.*

Proof. (Lemma 1) Let X be a pseudo-convex set with unit-disk-graph $G = \mathrm{UDG}(X)$. Let $x, y \in X$ be any two points (disk centers) in X, and let $L = |xy|$. We show that $d_G(x, y) \leqslant \lceil \alpha L \rceil$ where $\alpha = 12/\pi < 3.82$.

If $L < 2$ then the two unit disks centered at x and y overlap, so that $d_G(x, y) = L \leqslant \lceil \alpha L \rceil$.

Now assume that $L \geqslant 2$. Since X is pseudo-convex, the convex hull of X is covered by the set of unit disks centered at points X; thus, for any $x, y \in X$, each point on the line segment xy is covered by a disk centered at a point in X. Let $S = \{z \in X \mid B_z \cap xy \neq \emptyset, |xz| > 2 \text{ and } |yz| > 2\}$ and let I be any maximal independent set in $S \cup \{x, y\}$. Since S is at distance at least 2 from x and y, we deduce that $x, y \in I$ and that all disks in $I \setminus \{x, y\}$ lie within the $L \times 4$ rectangle centered on xy; thus, $|I| \leqslant 4L/\pi$, since the unit disks centered at I are disjoint, each of area π. Let $I = \{x, z_1, \ldots, z_K, y\}$, with the points $z_i \in I$ indexed according to the ordering along xy of the projections of the points z_i onto xy. Then, by the maximality of the independent set I, for each $i \in \{1, \ldots, K-1\}$, the shortest unweighted path in $\mathrm{UDG}(S)$ from z_i to z_{i+1} has at most 3 edges (otherwise, there would be a point of S that could be added to I while maintaining independence). Thus, there exists a connected subset $D \subseteq X$ such that $I \subseteq D$ and $|D| \leqslant 3|I| - 2 \leqslant 12L/\pi - 2$. We conclude that $d_G(x, y) \leqslant (12L/\pi - 2) + 1 \leqslant \lceil \alpha L \rceil$. □

Corollary 1 is then a consequence of Theorem 4 and Lemma 1.

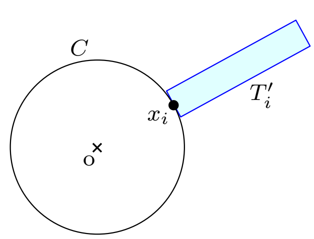
6 APX-hardness of Quad-Connected-Cover

Theorem 5. Quad-connected-cover *is APX-hard.*

The reduction will be from the following problem.

3-set-cover. Given a set X of n elements, and its subsets $\mathcal{S} = \{S_1, \ldots, S_m\}$ such that $|S_i| \leq 3$ for $i = 1, \ldots, m$, compute a minimum size subset of $\mathcal{S}$ that covers X. 3-set-cover is APX-hard (due to the fact that minimum vertex cover on graphs with maximum degree 3 is APX-hard).

Proof. (Theorem 5) The proof is by a reduction from 3-set-cover to quad-connected-cover. In particular, given a set $X = \{x_1, \ldots, x_n\}$ and subsets $\mathcal{S} = \{S_1, \ldots, S_m\}$, we show how to construct, in polynomial time and for any parameter $\varepsilon < \frac{1}{6}$, a $(1 + \varepsilon)$-approximation to 3-set-cover from a $\left(1 - \frac{\varepsilon}{6}\right)$-approximation to the quad-connected-cover.



Map the n points of X to n points placed uniformly on a circle C of unit area centered at the origin o; we will use the notation x_i for these points as well. Our set $\mathcal{T}$ will consist of convex quadrilaterals of two types:

- **center-quads.** These are, for each set $S_j \in \mathcal{S}$, the quadrilateral $T_j = \text{convexhull}(S_j \cup \{o\})$.
- **side-quads.** For each element x_i, let T'_i be the rectangle with width $\frac{1}{2n}$, length $4n$, containing x_i and tangent to C (see the figure).

Note that every pair of center-quads intersect (namely, at o), no two side-quads intersect, and a center-quad T_j intersects a side-quad T'_i if and only if $x_i \in S_j$. The area of the union of the center-quads is at most 1, and the area of each side-quad is 2.

Let s be the size of an optimal set-cover for X and $\mathcal{S}$. Let $\mathcal{T}'$ be a $(1-\frac{\varepsilon}{6})$-approximate solution to the quad-connected-cover problem on the set $\{T_1,\ldots,T_m,T'_1,\ldots,T'_n\}$ with $k=n+s$. Observe that to maintain connectivity of the intersection graph of $\mathcal{T}'$, if a point x_i is covered by a side-quad of $\mathcal{T}'$, it must also be covered by some center-quad of $\mathcal{T}'$, as a side-quad only intersects center-quads.

One possible solution consists of picking the s center-quads of the set-cover, and all the n side-quads to get the total area of at least $2n$; in particular, an optimal solution has value at least $2n$. Thus the area of the union of the quadrilaterals in $\mathcal{T}'$ is at least $\left(1-\frac{\varepsilon}{6}\right)\cdot 2n$. This implies that $\mathcal{T}'$ leaves at most $\frac{\varepsilon n}{6}$ elements of X uncovered by center-quads; otherwise at least $\frac{\varepsilon n}{6}+1$ side-quads are not picked, and so the area covered by $\mathcal{T}'$ can only be $1+2\left(n-\frac{\varepsilon n}{6}-1\right) \leq \left(1-\frac{\varepsilon}{6}\right)\cdot 2n-1$. Thus, out of the $n+s$ quadrilaterals in $\mathcal{T}'$, at least $n-\frac{\varepsilon n}{6}$ side-quads are present, and at most $(n+s)-\left(n-\frac{\varepsilon n}{6}\right)=s+\frac{\varepsilon n}{6}$ center-quads are present. Thus one can pick arbitrarily one set for each uncovered point to construct a set cover for X of size at most $\left(s+\frac{\varepsilon n}{6}\right)+\frac{\varepsilon n}{6} \leq s+\frac{2\varepsilon}{6}\cdot 3s \leq (1+\varepsilon)\cdot s$, where the first inequality follows from the fact that $s \geq \frac{n}{3}$. This completes the proof.

□

We conjecture that by finding a more specific reduction from APX-hard geometric covering problems in [16] for instance, the problem QUAD-CONNECTED-COVER remains APX-hard even when the quadrilaterals are replaced by triangles with area arbitrarily close to one.

References

[1] S. Arora, *Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems*, Journal of the ACM, 45 (1998), pp. 753–782.

[2] Y. Bartal, L. Gottlieb, and R. Krauthgamer, *The traveling salesman problem: low-dimensionality implies a polynomial time approximation scheme*, SIAM Journal on Computing, 45 (2016), pp. 1563–1581.

[3] G. Călinescu, C. Chekuri, M. Pál, and J. Vondrák, *Maximizing a monotone submodular function subject to a matroid constraint*, SIAM Journal on Computing, 40 (2011), pp. 1740–1766.

[4] E. W. Chambers, S. P. Fekete, H.-F. Hoffmann, D. Marinakis, J. S. Mitchell, V. Srinivasan, U. Stege, and S. Whitesides, *Connecting a set of circles with minimum sum of radii*, Computational Geometry, 68 (1991), pp. 62–76. Special issue in memory of Ferran Hurtado.

[5] T. H. Chan and S. H. Jiang, *Reducing curse of dimensionality: Improved PTAS for TSP (with neighborhoods) in doubling metrics*, in Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016, R. Krauthgamer, ed., SIAM, 2016, pp. 754–765.

[6] S. Chaplick, M. De, A. Ravsky, and J. Spoerhase, *Approximation schemes for geometric coverage problems*, in ESA, vol. 112 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018, pp. 17:1–17:15.

[7] B. N. Clark, C. J. Colbourn, and D. S. Johnson, *Unit disk graphs*, Discrete Math., 86 (1991), pp. 165–177.

[8] A. Dumitrescu and J. S. B. Mitchell, *Approximation algorithms for TSP with neighborhoods in the plane*, Journal of Algorithms, 48 (2003), pp. 135–159. Special issue devoted to 12th ACM-SIAM Symposium on Discrete Algorithms, Washington, DC, January, 2001.

[9] A. Dumitrescu and C. D. Tóth, *Constant-Factor Approximation for TSP with Disks*, Springer, 2017.

[10] K. M. Elbassioni, A. V. Fishkin, and R. Sitters, *On approximating the TSP with intersecting neighborhoods*, in Algorithms and Computation, 17th International Symposium, ISAAC 2006, Kolkata, India, December 18-20, 2006, Proceedings, T. Asano, ed., vol. 4288 of Lecture Notes in Computer Science, Springer, 2006, pp. 213–222.

[11] ——, *Approximation algorithms for the Euclidean traveling salesman problem with discrete and continuous neighborhoods*, International Journal of Computational Geometry Applications, 19 (2009), pp. 173–193.

[12] C. Feremans and A. Grigoriev, *Approximation schemes for the generalized geometric problems with geographic clustering*, in (Informal) Proceedings of the 21st European Workshop on Computational Geometry, Eindhoven, The Netherlands, March 9-11, 2005, Technische Universiteit Eindhoven, 2005, pp. 101–102.

[13] S. Funke, A. Kesselman, F. Kuhn, Z. Lotker, and M. Segal, *Improved approximation algorithms for connected sensor cover*, Wireless Networks, 13 (2007), pp. 153–164.

[14] M. M. Garey. and D. S. Johnson, *The rectilinear steiner tree problem is NP-complete*, SIAM Journal on Applied Mathematics, 32 (1977), pp. 826–834.

[15] H. Gupta, Z. Zhou, S. R. Das, and Q. Gu, *Connected sensor cover: Self-organization of sensor networks for efficient query execution*, IEEE/ACM Transactions on Networks, 14 (2006), pp. 55–67.

[16] S. Har-Peled, *Being fat and friendly is not enough*, CoRR, abs/0908.2369 (2009).

[17] D. S. Hochbaum and A. Pathria, *Node-optimal connected k-subgraphs*, Technical Report, UC Berkeley, https://hochbaum.ieor.berkeley.edu/html/pub/k-connect.pdf, 1994.

[18] C. Huang, M. Mari, C. Mathieu, J. S. B. Mitchell, and N. H. Mustafa, *Maximizing covered area in the euclidean plane with connectivity constraint*, in Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2019, September 20-22, 2019, Massachusetts Institute of Technology, Cambridge, MA, USA, D. Achlioptas and L. A. Végh, eds., vol. 145 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019, pp. 32:1–32:21.

[19] H. B. Hunt, M. V. Marathe, V. Radhakrishnan, S. Ravi, D. J. Rosenkrantz, and R. E. Stearns, *NC-approximation schemes for NP- and PSPACE-Hard problems for geometric graphs*, J. Algorithms, 26 (1998), p. 238274.

[20] K. Kar and S. Banerjee, *Node placement for connected coverage in sensor networks*, 2003.

[21] S. Khuller, M. Purohit, and K. K. Sarpatwar, *Analyzing the optimal neighborhood: Algorithms for budgeted and partial connected dominating set problems*, in Proceedings of the Twenty-fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '14, Philadelphia, PA, USA, 2014, Society for Industrial and Applied Mathematics, pp. 1702–1713.

[22] A. Kulik, H. Shachnai, and T. Tamir, *Maximizing submodular set functions subject to multiple linear constraints*, in Symposium on Discrete Algorithms (SODA), SIAM, 2009, pp. 545–554.

[23] T.-W. Kuo, K. C.-J. Lin, and M.-J. Tsai, *Maximizing submodular set function with connectivity constraint: Theory and application to networks*, IEEE/ACM Transactions on Networks, 23 (2015), pp. 533–546.

[24] I. Lamprou, I. Sigalas, and V. Zissimopoulos, *Improved budgeted connected domination and budgeted edge-vertex domination*, in Combinatorial Algorithms - 31st International Workshop, IWOCA 2020, Bordeaux, France, June 8-10, 2020, Proceedings, L. Gasieniec, R. Klasing, and T. Radzik, eds., vol. 12126 of Lecture Notes in Computer Science, Springer, 2020, pp. 368–381.

[25] J. Lee, M. Sviridenko, and J. Vondrák, *Submodular maximization over multiple matroids via generalized exchange properties*, Mathematics of Operations Research, 35 (2010), pp. 795–806.

[26] M. V. Marathe, H. Breu, H. B. Hunt III, S. S. Ravi, and D. J. Rosenkrantz, *Simple heuristics for unit disk graphs*, Networks, 25 (1995), pp. 59–68.

[27] T. Matsui, *Approximation algorithms for maximum independent set problems and fractional coloring problems on unit disk graphs*, in Discrete and Computational Geometry, J. Akiyama, M. Kano, and M. Urabe, eds., Berlin, Heidelberg, 2000, Springer Berlin Heidelberg, pp. 194–200.

[28] J. S. B. Mitchell, *Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for geometric TSP, k-MST, and related problems*, SIAM Journal on Computing, 28 (1999), pp. 1298–1309.

[29] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, *An analysis of approximations for maximizing submodular set functions-I*, Mathematical programming, 14 (1978), pp. 265–294.

[30] Y. Rabani and G. Scalosub, *Bicriteria approximation tradeoff for the node-cost budget problem*, ACM Transactions on Algorithms, 5 (2009), pp. 19:1–19:14.

[31] F. van Nijnatten, R. Sitters, G. J. Woeginger, A. Wolff, and M. de Berg, *The traveling salesman problem under squared Euclidean distances*, in 27th International Symposium on Theoretical Aspects of Computer Science, STACS 2010, March 4-6, 2010, Nancy, France, J. Marion and T. Schwentick, eds., vol. 5 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2010, pp. 239–250.

[32] F. Vandin, E. Upfal, and B. J. Raphael, *Algorithms for detecting significantly mutated pathways in cancer*, Journal of Computational Biology, 18 (2011), pp. 507–522.

[33] L. A. Wolsey, *Maximising real-valued submodular functions: primal and dual heuristics for location problems*, Mathematics of Operations Research, 7 (1982), pp. 410–425.