

Approximating Maximum Diameter-Bounded Subgraph in Unit Disk Graphs

A. Karim Abu-Affash¹

Software Engineering Department, Shamoon College of Engineering
Beer-Sheva 84100, Israel
abuaa1@sce.ac.il

Paz Carmi²

Department of Computer Science, Ben-Gurion University
Beer-Sheva 84105, Israel
carmip@cs.bgu.ac.il

Anil Maheshwari³

School of Computer Science, Carleton University
Ottawa, Canada
anil@scs.carleton.ca

Pat Morin⁴

School of Computer Science, Carleton University
Ottawa, Canada
morin@scs.carleton.ca

Michiel Smid⁵

School of Computer Science, Carleton University
Ottawa, Canada
michiel@scs.carleton.ca

Shakhar Smorodinsky⁶

Department of Mathematics, Ben-Gurion University
Beer-Sheva 84105, Israel
shakhar@math.bgu.ac.il

Abstract

We consider a well studied generalization of the maximum clique problem which is defined as follows. Given a graph G on n vertices and an integer $d \geq 1$, in the *maximum diameter-bounded subgraph* problem (MaxDBS for short), the goal is to find a (vertex) maximum subgraph of G of diameter at most d . For $d = 1$, this problem is equivalent to the maximum clique problem and thus it is NP-hard to approximate it within a factor $n^{1-\epsilon}$, for any $\epsilon > 0$. Moreover, it is known that, for any $d \geq 2$, it is NP-hard to approximate MaxDBS within a factor $n^{1/2-\epsilon}$, for any $\epsilon > 0$.

In this paper we focus on MaxDBS for the class of unit disk graphs. We provide a polynomial-time constant-factor approximation algorithm for the problem. The approximation ratio of our algorithm does not depend on the diameter d . Even though the algorithm itself is simple, its analysis is rather involved. We combine tools from the theory of hypergraphs with bounded VC-dimension, k -quasi planar graphs, fractional Helly theorems and several geometric properties of unit disk graphs.

¹ Work was supported by Grant 2016116 from the United States – Israel Binational Science Foundation.

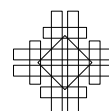
² Work was supported by Grant 2016116 from the United States – Israel Binational Science Foundation.

³ Work was supported by NSERC

⁴ Work was supported by NSERC

⁵ Work was supported by NSERC

⁶ Work was partially supported by Grant 635/16 from the Israel Science Foundation



2012 ACM Subject Classification Theory of computation → Computational geometry, Theory of computation → Numeric approximation algorithms

Keywords and phrases Approximation algorithms, maximum diameter-bounded subgraph, unit disk graphs, fractional Helly theorem, VC-dimension

Digital Object Identifier 10.4230/LIPIcs.SoCG.2018.2

Acknowledgements The authors would like to thank the Fields Institute for hosting the workshop in Ottawa and for their financial support.

1 Introduction

Computing a maximum size clique in a graph is one of the fundamental problems in theoretical computer science [7]. It is not only NP-hard but even hard to approximate within a factor of $n^{1-\epsilon}$, for any $\epsilon > 0$, unless $P = NP$ [16]. A clique, equivalently, is a subgraph of diameter 1. A natural, well studied, generalization of the maximum clique problem is the *maximum diameter-bounded subgraph* problem (MaxDBS for short), in which the goal is to compute a maximum (vertex) subgraph of diameter d , for a given $d \geq 2$. A subgraph with diameter $\leq d$ is sometimes referred to in the literature as a *d-club*. MaxDBS is also known to have hardness of approximation of $n^{1/2-\epsilon}$, unless $P = NP$ [5].

In this paper, we study MaxDBS in the class of *unit disk graphs*. A *unit disk graph* is defined as the intersection graph of disks of equal (e.g., unit) diameter in the plane. Unit disk graphs provide a graph-theoretic model for ad hoc wireless networks, where two wireless nodes can communicate if they are within the unit Euclidean distance away from each other.

Many classical NP-Complete problems including chromatic number, independent set and dominating set are still NP-complete even for unit disk graphs [13, 14]. However, the class of unit disk graphs is one of the non-trivial classes of graphs for which the maximum clique problem is in P . Indeed, in a celebrated result, Clark, Colbourn and Johnson [13] provide a beautiful polynomial time algorithm to compute the maximum clique in unit disk graphs. Unfortunately, we do not know how to extend this algorithm to the MaxDBS problem.

Our main contribution in this paper is a polynomial-time algorithm that approximates MaxDBS within a constant factor in unit disk graphs. Moreover, our constant factor approximation ratio does not depend on the diameter d . To the best of our knowledge, it is not known whether this problem is NP-hard. Our algorithm works as follows: Given a unit disk graph $G = (V, E)$ and an integer d , compute for each vertex $v \in V$ the BFS-tree of radius $\frac{d}{2}$ (or $\frac{d+1}{2}$ if d is odd) centered at v , and return the tree of maximum size. Although, the algorithm is very natural and simple, its analysis for unit disk graphs is rather involved. We note that this algorithm might perform very bad for arbitrary graphs. For example, a 2-subdivision of a clique with n vertices is the graph obtained by subdividing each edge of the clique K_n into a path of length 2. It is easily seen that such a graph with $n + \binom{n}{2}$ vertices has diameter 4 but every BFS tree of radius 2 contains at most $2n$ vertices.

1.1 Related work

MaxDBS has been studied extensively in general graphs in the last two decades. Bourjolly et al. [8] showed that MaxDBS is NP-hard. Balasundaram et al. [6] proved that for any d , MaxDBS is NP-hard in graphs of diameter $d + 1$. Asahiro et al. [5] showed that, for any $\epsilon > 0$ and $d \geq 2$, it is NP-hard to approximate MaxDBS within a factor of $n^{1/2-\epsilon}$,

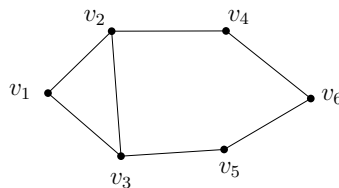
and they gave an $n^{1/2}$ -approximation algorithm for the problem. Chang et al. [11] provide an algorithm that finds a maximum subgraph of diameter d in $O(1.62^n \cdot \text{poly}(n))$ time. There are more results on solving MaxDBS by using various integer and linear programming formulations [4, 6, 9, 8, 10, 21].

Asahiro et al. [5] studied the MaxDBS in other subclasses of graphs, including chordal graphs, interval graphs, and s -partite graphs. For chordal graphs, they showed that the problem can be solved in polynomial-time for odd d 's, and cannot be approximated within factor $n^{1/3-\epsilon}$, for any $\epsilon > 0$ for even d 's. For interval graphs, they showed that the problem can be solved in polynomial-time. For s -partite graphs, they showed that the problem cannot be approximated (unless $P = NP$) within a factor of $n^{1/3-\epsilon}$, for any $\epsilon > 0$, when $s = 2$ and $d \geq 3$, and when $s \geq 3$ and $d \geq 2$.

For unit disk graphs, the hardness of MaxDBS is still open, for $d \geq 2$ we are not aware of any previous work. As mentioned already, for $d = 1$, the problem is equivalent to the maximum clique problem and it can be solved in polynomial-time [13].

1.2 Motivation

MaxDBS is a relaxation of the maximum clique problem and is motivated by cluster-detection that arise in wide variety applications. For instance, finding clusters in networks helps in understanding and analyzing the structure of the network. Another well studied notion is that of a d -clique [6, 20, 21]. A d -clique of a graph G is a subset S of vertices of G , such that, for every two vertices in S , the shortest distance between them in G is at most d . Clearly, every d -club is a d -clique, but not vice versa, as shown in the example given by Alba [3] in Figure 1.



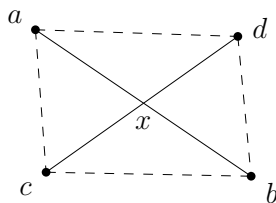
■ **Figure 1** $S = \{v_1, v_2, v_3, v_4, v_5\}$ is a 2-clique but not a 2-club since the graph induced by S has a diameter 3.

2 Preliminaries

Let V be a finite set of points in the plane. For two points $u, v \in V$, let $|uv|$ denote the Euclidean distance between u and v . The unit disk graph on V is the undirected graph $G = (V, E)$, such that $(u, v) \in E$ if and only if $|uv| \leq 1$. The following lemma (though very simple) turns out to be crucial to prove our main result.

► **Lemma 1.** *For every two crossing edges (a, b) and (c, d) in G , at least one of the edges (a, c) and (b, d) is in G , and at least one of the edges (a, d) and (b, c) is in G ; see Figure 2 for an illustration.*

Proof. To prove the lemma, it suffices to show that $\min\{|ac|, |bd|\} \leq 1$ and $\min\{|ad|, |bc|\} \leq 1$; see Figure 2. Let x be the intersection point of (a, b) and (c, d) . By the triangle inequality, $|ac| \leq |ax| + |xc|$ and $|bd| \leq |bx| + |xd|$. Thus, $|ac| + |bd| \leq |ab| + |cd| \leq 2$. Therefore, $\min\{|ac|, |bd|\} \leq 1$. By a similar argument, we prove that $\min\{|ad|, |bc|\} \leq 1$. ◀



■ **Figure 2** An illustration for the proof of Lemma 1.

2.1 Tool box

A *range space* (or a set system) (X, \mathcal{R}) is a pair consisting of a set X of objects (called the *space*) and a family \mathcal{R} of subsets of X (called *ranges*). We say that a subset A of X is *shattered* by \mathcal{R} , if for each subset A' of A , there exist a range $R \in \mathcal{R}$, such that $A \cap R = A'$. The *Vapnik-Chervonenkis* dimension (or VC-dimension for short) of a range space (X, \mathcal{R}) is the size of the largest (finite) shattered subset of X ; see [15] for examples of range spaces of bounded VC-dimension.

The *dual range space* of (X, \mathcal{R}) is a range space (Y, \mathcal{R}^*) , where $Y = \{y_R : R \in \mathcal{R}\}$ and, for each $x \in X$, the set $\{y_R : x \in R\}$ is a range in \mathcal{R}^* . It is well known [17] that, if the VC-dimension of (X, \mathcal{R}) is k , then the VC-dimension of the dual range space (Y, \mathcal{R}^*) is at most 2^k .

A range space (X, \mathcal{R}) has fractional Helly number k , if for every $\alpha > 0$, there exists $\beta > 0$, such that if at least $\alpha \binom{|\mathcal{R}|}{k}$ subsets of size k of \mathcal{R} have a non-empty intersection, then there exists an element of X that is contained in at least $\beta \cdot |\mathcal{R}|$ sets of \mathcal{R} . In [18], Matoušek proved the following theorem showing that every range space of bounded VC-dimension has a fractional Helly property.

► **Theorem 2** ([18]). *Let (X, \mathcal{R}) be a range space such that the VC-dimension of the dual range space of (X, \mathcal{R}) is at most $k - 1$. Then, (X, \mathcal{R}) has a fractional Helly number k .*

A range space (X, \mathcal{R}) satisfies the (p, q) -property if among every p ranges of \mathcal{R} some q have a non-empty intersection. Matoušek [18] established the following (p, q) -theorem for range spaces of bounded VC-dimension.

► **Theorem 3** ([18]). *Let (X, \mathcal{R}) be a range space such that the VC-dimension of the dual range space of (X, \mathcal{R}) is at most $k - 1$ and let $p \geq k$. Then, there exists a constant t (depending only on p and k), such that if (X, \mathcal{R}) satisfies the (p, k) -property, then there exists a subset X' of X of size at most t intersecting all the ranges of \mathcal{R} , i.e., $X' \cap R \neq \emptyset$, for every $R \in \mathcal{R}$.*

A (simple) *topological graph* is a graph drawn in the plane, such that its vertices are represented by a set of distinct points and its edges are Jordan arcs connecting the corresponding points, so that (i) each edge does not contain any other vertex as an interior point, (ii) every pair of edges intersect at most once, and (iii) no three edges have a common intersection point. Agarwal et al. [2] showed that any topological graph with n vertices and without k pairwise crossing edges has $O(n \log^{2k-6} n)$ edges. This bound was further improved to $O(n \log^{2k-8} n)$ by Ackerman [1]. Hence, if G is a complete topological graph on n vertices and without k pairwise crossing edges, then $\binom{n}{2} = n(n-1)/2 \leq c' n \log^{2k-8} n$, where c' is the constant in the big ‘ O ’, depending only on k . This implies that $k \geq \frac{\log(n-1)-c}{2 \log \log n} + 4$, where c is a constant depending on c' . Therefore, we have the following corollary.

► **Corollary 4.** *Any complete topological graph on n vertices contains at least $\frac{\log(n-1)-c}{2 \log \log n} + 4$ pairwise crossing edges, where c is a constant.*

3 Approximation algorithm

Let $G = (V, E)$ be the unit disk graph of a set of points V in the plane. For two vertices u and v in V , let $d(u, v)$ denote the shortest (hop) distance between u and v in G .⁷ Assuming that G is connected, the diameter of G is defined as the maximum (hop) distance between any two vertices in G , i.e., $\max_{u,v \in V} d(u, v)$. A subgraph of G is called d -club if its diameter is equal to d . Let G_{opt} denote a maximum d -club of G . In this section, we first present a polynomial-time approximation algorithm that computes a d -club of size at least c times the size of G_{opt} , where c is a constant and d is even. Later, we show how to generalize this algorithm for odd d 's.

Set $r = \frac{d}{2}$. For a vertex $u \in V$, let $T_r(u)$ denote the tree of center u and radius r that contains all vertices of distance at most r from u in G . Namely, a vertex v is in $T_r(u)$ if and only if $d(u, v) \leq r$. Given a vertex $u \in V$, $T_r(u)$ can be computed using the breadth first search (BFS) algorithm in $O(|V| + |E|)$ time. Our algorithm computes all the trees of radius r centered at vertices of G and returns a tree T of the maximum size, i.e., the tree that contains the maximum number of vertices. It is clear that T is a d -club of G and can be computed in polynomial time. Let V_{opt} denote the set of vertices of G_{opt} and $n = |V_{opt}|$ denote the size of G_{opt} . In the following, we prove that T contains at least cn vertices, where c is a constant.

Let $\mathcal{T}_r(G_{opt})$ denote the set of all trees of radius r centered at vertices of G_{opt} , and let T_r^* be a tree in $\mathcal{T}_r(G_{opt})$ that contains the maximum number of vertices of G_{opt} among trees of $\mathcal{T}_r(G_{opt})$. First, observe that the size of T is at least as the size of T_r^* , since G_{opt} is a subgraph of G . Therefore, it is sufficient to prove that T_r^* contains at least cn vertices.

Let (X, \mathcal{R}) be the range space where $X = V_{opt}$ and for each tree $T_r(u)$ in $\mathcal{T}_r(G_{opt})$, the set $\{v \in V_{opt} : v \in T_r(u)\}$ is a range in \mathcal{R} . Thus, in the dual range space (Y, \mathcal{R}^*) of (X, \mathcal{R}) , we have $Y = \{y_{T_r(u)} : u \in V_{opt}\}$, and, for each point $v \in V_{opt}$, the set $\{y_{T_r(u)} : v \in T_r(u)\}$ is a range in \mathcal{R}^* .

► **Observation 5.** *(X, \mathcal{R}) and (Y, \mathcal{R}^*) are isomorphic.*

Proof. Since $T_r(u)$ contains v if and only if $T_r(v)$ contains u , we have

$$\{y_{T_r(u)} : v \in T_r(u)\} = \{y_{T_r(u)} : u \in T_r(v)\}.$$

Now, if we set $y_{T_r(u)} = u$, then we obtain that $X = Y$ and

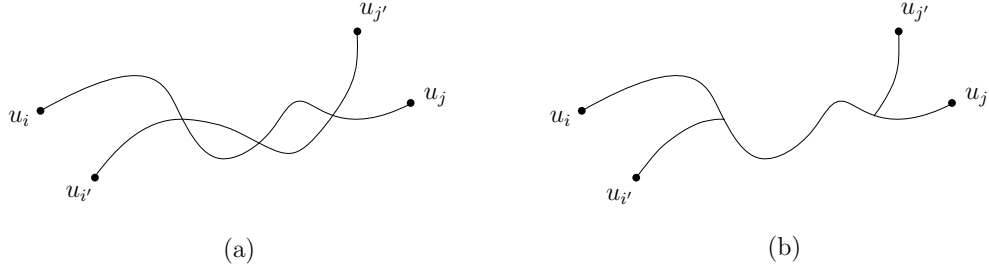
$$\{y_{T_r(u)} : v \in T_r(u)\} = \{u \in V_{opt} : v \in T_r(u)\} = \{u \in V_{opt} : u \in T_r(v)\},$$

for each $v \in V_{opt}$. Therefore, for each $v \in V_{opt}$, the set $\{u \in V_{opt} : u \in T_r(v)\}$ is a range in \mathcal{R}^* , which implies that $\mathcal{R}^* = \mathcal{R}$. ◀

► **Theorem 6.** *T_r^* contains at least cn vertices, where c is a constant.*

Proof. The proof plan is as follows. We show (later in Section 4) that the VC-dimension of the range space $(V_{opt}, \mathcal{T}_r(G_{opt}))$ is 4. Thus, by Observation 5, the VC-dimension of the dual

⁷ Note that for any $\epsilon > 0$, it could hold that the Euclidean distance between u and v is $1 + \epsilon$ but they are in different connected components of G and hence $d(u, v)$ is not necessarily bounded.



■ **Figure 3** (a) $\delta(u_i, u_j)$ and $\delta(u_{i'}, u_{j'})$ intersect in more than one point, and (b) replacing subpaths of $\delta(u_{i'}, u_{j'})$ by subpaths of $\delta(u_i, u_j)$ between the intersection points.

range space of $(V_{opt}, \mathcal{T}_r(G_{opt}))$ is also 4. Then, we use Corollary 4 to show that there exists a constant $m \geq 5$, such that $(V_{opt}, \mathcal{T}_r(G_{opt}))$ satisfies the $(m, 5)$ -property, which means that at least $\binom{n}{m} / \binom{n-5}{m-5} = \binom{n}{5} / \binom{m}{5}$ subsets of 5 trees of $\mathcal{T}_r(G_{opt})$ share a common point. Thus, by Theorem 2, there exists a point that is contained in at least βn trees of $\mathcal{T}_r(G_{opt})$, and therefore, there is a tree in $\mathcal{T}_r(G_{opt})$ that contains at least βn points of V_{opt} , which proves that $c \geq \beta > 0$.

Let m be an integer such that $\frac{\log(m-1)-c'}{2 \log \log m} + 4 \geq 6$, where c' is a constant. Let A be a set of m trees of $\mathcal{T}_r(G_{opt})$ and let $C = \{u_1, u_2, \dots, u_m\}$ be the centers of the trees of A . For two points $u_i, u_j \in C$, let $\delta(u_i, u_j)$ be a shortest path between u_i and u_j in G_{opt} , and let $d(u_i, u_j)$ be the length of $\delta(u_i, u_j)$. For every four distinct points $u_i, u_j, u_{i'}, u_{j'} \in C$, we assume that the intersection of the paths $\delta(u_i, u_j)$ and $\delta(u_{i'}, u_{j'})$ is either empty or a path (otherwise, we replace every subpath of $\delta(u_{i'}, u_{j'})$ between every two consecutive intersection points of the paths by the subpath of $\delta(u_i, u_j)$ between the same points; see Figure 3). Since $d(u_i, u_j) \leq 2r$, there is at least one point $u_{i,j}$ on $\delta(u_i, u_j)$ that is contained in $T_r(u_i)$ and in $T_r(u_j)$.

We now construct a drawing of a complete graph H on the points of C in which the edges are drawn as the Jordan arcs $\delta(u_i, u_j)$. Notice that, H is not necessarily a topological graph. However, we can transform it into a topological graph H' , such that, for every four distinct points $u_i, u_j, u_{i'}, u_{j'} \in C$, $\delta(u_i, u_j)$ and $\delta(u_{i'}, u_{j'})$ are crossing in H' if and only if they are crossing in H . This transformation is obtained using standard operations as in [12] and we omit the technical details here. Since H' is a complete topological graph on m vertices, by Corollary 4, H' has at least 6 pairwise crossing edges. Let $P = \{\delta(u_1, u_{1'}), \delta(u_2, u_{2'}), \dots, \delta(u_6, u_{6'})\}$ be the set of the corresponding 6 pairwise crossing paths in H' .

► **Lemma 7.** *Let $\delta(u_i, u_{i'})$ and $\delta(u_j, u_{j'})$ be two crossing paths of P and let x be their intersection point. Assume, w.l.o.g., that x is between $u_{i,i'}$ and u_i , and between $u_{j,j'}$ and u_j ; see Figure 4. Then, either $T_r(u_i)$ contains $u_{j,j'}$ or $T_r(u_j)$ contains $u_{i,i'}$.*

Proof. We distinguish between two cases.

Case 1: x is a point of V_{opt} ; see Figure 4(a). Assume, w.l.o.g., that $d(u_i, x) \leq d(u_j, x)$. Thus,

$$d(u_i, u_{j,j'}) \leq d(u_i, x) + d(x, u_{j,j'}) \leq d(u_j, x) + d(x, u_{j,j'}) = d(u_j, u_{j,j'}) \leq r.$$

Therefore, $u_{j,j'}$ is of distance at most r from u_i and, hence, is contained in $T_r(u_i)$.

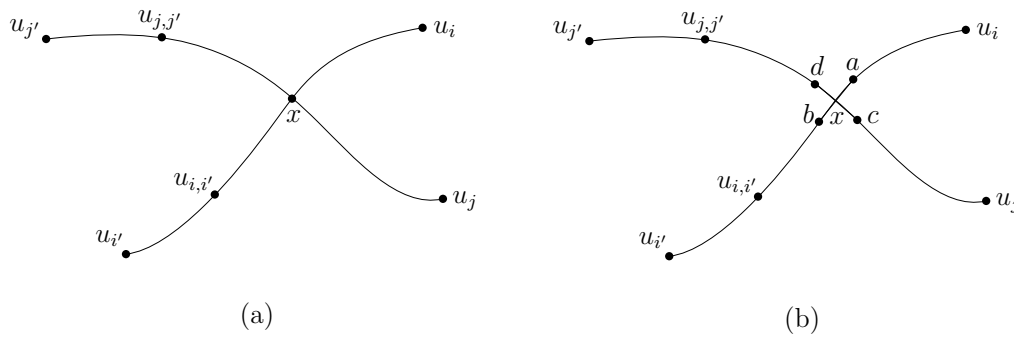


Figure 4 $\delta(u_i, u_{i'})$ and $\delta(u_j, u_{j'})$ intersect at x . (a) x is a point of V_{opt} , and (b) x is an intersection point of the edges (a, b) and (c, d) .

Case 2: x is not a point of V_{opt} . Thus, x is an intersection point of two edges (a, b) and (c, d) of G . Assume, w.l.o.g., that a is between x and u_1 and c is between x and u_2 ; see Figure 4(b).

- If $d(u_i, a) = d(u_j, c)$, then, by Lemma 1, at least one of the edges (a, d) and (b, c) is in G_{opt} .

- If (a, d) is in G_{opt} , then $d(a, u_{j,j'}) = d(c, u_{j,j'})$, and hence,

$$d(u_i, u_{j,j'}) \leq d(u_i, a) + d(a, u_{j,j'}) = d(u_j, c) + d(c, u_{j,j'}) = d(u_j, u_{j,j'}) \leq r.$$

Therefore, $u_{j,j'}$ is of distance at most r from u_i and, hence, is contained in $T_r(u_i)$.

- If (b, c) is in G_{opt} , then $d(a, u_{i,i'}) = d(c, u_{i,i'})$, and hence,

$$d(u_j, u_{i,i'}) \leq d(u_j, c) + d(c, u_{i,i'}) = d(u_i, a) + d(a, u_{i,i'}) = d(u_i, u_{i,i'}) \leq r.$$

Therefore, $u_{i,i'}$ is of distance at most r from u_j and, hence, is contained in $T_r(u_j)$.

- Otherwise, assume, w.l.o.g., that $d(u_i, a) < d(u_j, c)$. By Lemma 1, at least one of the edges (a, c) and (b, d) is in G_{opt} .

- If (a, c) is in G_{opt} , then $d(u_i, c) \leq d(u_j, c)$. Hence,

$$d(u_i, u_{j,j'}) \leq d(u_i, c) + d(c, u_{j,j'}) \leq d(u_j, c) + d(c, u_{j,j'}) = d(u_j, u_{j,j'}) \leq r.$$

- If (b, d) is in G_{opt} , then $d(u_i, d) \leq d(u_j, d)$. Hence,

$$d(u_i, u_{j,j'}) \leq d(u_i, d) + d(d, u_{j,j'}) \leq d(u_j, d) + d(d, u_{j,j'}) = d(u_j, u_{j,j'}) \leq r.$$

In both cases, $u_{j,j'}$ is of distance at most r from u_i and, hence, is contained in $T_r(u_i)$. ◀

► **Lemma 8.** $(V_{opt}, \mathcal{T}_r(G_{opt}))$ satisfies the $(m, 5)$ -property.

Proof. By Lemma 7, for every two paths $\delta(u_i, u_{i'})$ and $\delta(u_j, u_{j'})$ in P , either at least one of the trees $T_r(u_i)$ and $T_r(u_{i'})$ contains $u_{j,j'}$ or at least one of the trees $T_r(u_j)$ and $T_r(u_{j'})$ contains $u_{i,i'}$. We construct a directed graph on the vertices $\{u_1, u_2, \dots, u_6\}$, such that there is a directed edge from u_i to u_j if and only if at least one of the trees $T_r(u_i)$ and $T_r(u_{i'})$ contains $u_{j,j'}$. Since we have 6 pairwise crossing paths, there are at least 15 edges in this graph, which means that there is a vertex u_l in this graph, $1 \leq l \leq 6$, of in-degree at least 3. Hence, there is a point $u_{l,l'}$ that is covered by at least 3 other trees, in addition to the trees $T_r(u_l)$ and $T_r(u_{l'})$. Thus, $u_{l,l'}$ is contained in at least 5 trees of A . Therefore, $(V_{opt}, \mathcal{T}_r(G_{opt}))$ satisfies the $(m, 5)$ -property. ◀

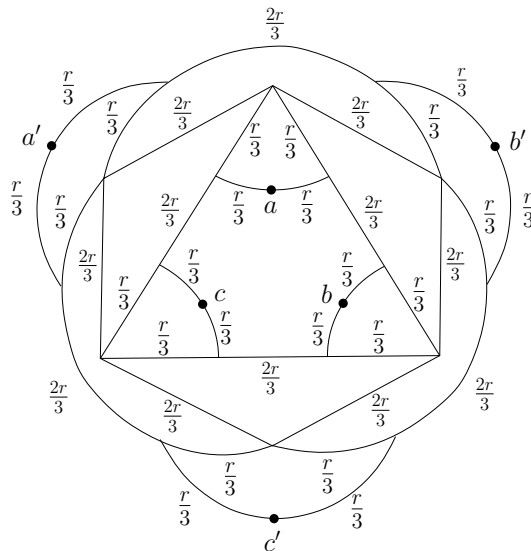
Since $(V_{opt}, \mathcal{T}_r(G_{opt}))$ satisfies the $(m, 5)$ -property, out of every m trees of $\mathcal{T}_r(G_{opt})$ there are 5 trees that share a common point. Thus, there are at least $\binom{n}{m} / \binom{n-5}{m-5} = \binom{n}{5} / \binom{m}{5}$ sets containing 5 trees that share a common point. Moreover, in Theorem 12 (Section 4), we show that the VC-dimension of the range space $(V_{opt}, \mathcal{T}_r(G_{opt}))$ is 4. Thus, by Observation 5, the VC-dimension of the dual range space of $(V, \mathcal{T}_r(G))$ is also 4 and therefore, by Theorem 2, the fractional Helly number of $(V_{opt}, \mathcal{T}_r(G_{opt}))$ is 5. Now, by setting $\alpha = 1/\binom{m}{5}$ (in the fractional Helly theorem), we have a point $u \in V_{opt}$ that is contained in at least βn trees of $\mathcal{T}_r(G_{opt})$, which means that $T_r(u)$ contains at least βn points of V_{opt} , where $\beta > 0$. Therefore, $c \geq \beta > 0$, which completes the proof of the theorem. ◀

► **Corollary 9.** *For any even $d \geq 2$, every d -club in any unit disk graph can be covered by a constant-number of trees of radius $\frac{d}{2}$.*

Proof. To prove the corollary, we show that there exists a constant ρ , such that G_{opt} can be covered by at most ρ trees of $\mathcal{T}_r(G_{opt})$. By Theorem 12, the VC-dimension of the dual range space of $(V_{opt}, \mathcal{T}_r(G_{opt}))$ is 4, and, by Lemma 8, $(V_{opt}, \mathcal{T}_r(G_{opt}))$ satisfies the $(m, 5)$ -property. Therefore, by Theorem 3, there exists a set of at most t trees of $\mathcal{T}_r(G_{opt})$ that cover all vertices of V_{opt} , which proves that $\rho \leq t$. ◀

Upper bound on c

We show, in Figure 5, a unit disk graph G on n vertices for which the tree computed by our algorithm does not contain more than $\frac{n}{3}$. G contains $n = 16r$ points and its diameter is $d = 2r$. Each tree of radius r in G covers at most $6r$ points. This proves that $c \leq \frac{3}{8}$. To show that $c \leq \frac{1}{3}$, we locate six cliques of size $\frac{n-16r}{6}$ on the points a, b, c, a', b' , and c' . Now, each tree of radius r can cover at most 2 cliques. Therefore, for sufficiently large n , we have $c \leq \frac{1}{3}$.



■ **Figure 5** G contains $16r$ points and its diameter is $d = 2r$. Each tree of radius r covers at most $6r$ points.

Generalization for odd d

In this section, we extend our algorithm to approximate MaxDBS for odd d 's. Given a unit disk graph G of a set V of points in the plane and an odd integer $d \geq 3$, let G_d be a maximum d -club and let G_{d+1} be a maximum $(d + 1)$ -club of G . Let n_d and n_{d+1} be the sizes of G_d and G_{d+1} , respectively, and observe that $n_{d+1} \geq n_d$. We set $r = \frac{d+1}{2}$ and we use our algorithm to compute a tree $T_r(u)$ of size at least $cn_{d+1} \geq cn_d$. Notice that $T_r(u)$ is a $(d + 1)$ -club but may not be a d -club. In the following lemma, we show that there is a subtree of $T_r(u)$ of diameter $d - 1$ that contains at least $1/12$ of the vertices of $T_r(u)$.

► **Lemma 10.** *The vertices of tree $T_r(u)$ can be covered by at most 12 trees of radius $r - 1$.*

Proof. Let $V_r(u)$ be the set of vertices of $T_r(u)$ and let $D_2(u) = \{v \in V_r(u) : d(u, v) = 2\}$, i.e., the set of all vertices of $V_r(u)$ of distance two from u . Let I be a maximal independent set of $D_2(u)$. By the packing argument in unit disk graphs, we have $|I| \leq 12$. Let v be a vertex in $V_r(u)$, and let $\delta(u, v)$ be a shortest path between u and v in $T_r(u)$. Since $d(u, v) \leq r$, there is at least one vertex $u' \in D_2(u)$, such that every vertex in $\delta(u, v)$ is of distance at most $r - 2$ from u' . Hence, there is at least one vertex $x \in I$, such that every vertex in $\delta(u, v)$ is of distance at most $r - 1$ from x . Thus, every vertex in $V_r(u)$ is contained in $T_{r-1}(x)$, for some $x \in I$, and therefore, $V_r(u)$ is covered by $\bigcup_{x \in I} T_{r-1}(x)$. ◀

By Lemma 10, we can find a tree $T_{r-1}^*(x)$ that contains at least $1/12$ of the vertices of $T_r(u)$. Since $r = \frac{d+1}{2}$, the diameter of $T_{r-1}^*(x)$ is at most $2(r - 1) = d - 1$. Therefore, $T_{r-1}^*(x)$ is a d -club of G and its size is at least $\frac{c}{12}n_d$.

The following theorem summarizes the result of this section.

► **Theorem 11.** *Given a unit disk graph G in the plane and an integer $d \geq 2$, one can find in polynomial time a d -club of G of size at least $\frac{c}{12}$ the size of a maximum d -club of G , where $0 < c \leq \frac{1}{3}$.*

4 The VC-Dimension of $(V_{opt}, \mathcal{T}_r(G_{opt}))$

In this section, we prove the following theorem.

► **Theorem 12.** *The range space $(V_{opt}, \mathcal{T}_r(G_{opt}))$ has VC-dimension 4.*

Proof. We first prove that the VC-dimension of $\mathcal{T}_r(G_{opt})$ is at most 4. For the sake of contradiction, suppose that there exist a set of points P and a subset $S = \{u_1, u_2, u_3, u_4, u_5\}$ of V_{opt} , such that S is shattered by $\mathcal{T}_r(G_{opt})$. Thus, for each $1 \leq i < j \leq 5$, there is a tree $T_r(c_{i,j})$ in $\mathcal{T}_r(G_{opt})$, such that $T_r(c_{i,j}) \cap S = \{u_i, u_j\}$. Let $P_{i,j}$ be the path between u_i and u_j in $T_r(c_{i,j})$. Note that $P_{i,j} \cap S = \{u_i, u_j\}$. Moreover, since S contains five points, by planarity constraints, at least two paths $P_{i,j}$ and $P_{k,l}$, for distinct two pairs (i, j) and (k, l) , intersect. Assume, w.l.o.g., that $P_{1,3}$ and $P_{2,4}$ intersect and let x be their intersection point. Assume also that x is between u_1 and $c_{1,3}$, and between u_2 and $c_{2,4}$; see Figure 6.

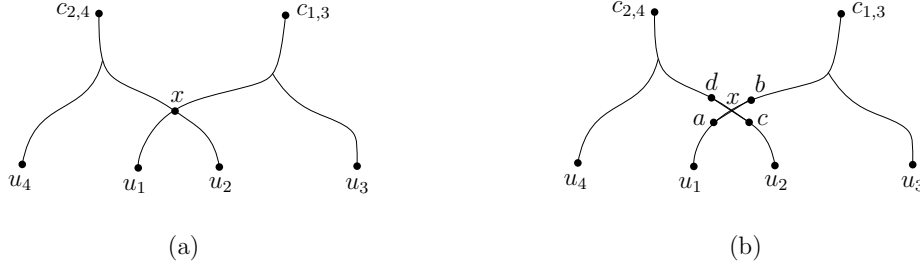
► **Lemma 13.** *Either u_1 is contained in $T_r(c_{2,4})$ or u_2 is contained in $T_r(c_{1,3})$.*

Proof. The proof is similar to the proof of Lemma 7. We distinguish between two cases.

Case 1: x is a point of V_{opt} ; see Figure 6(a). Assume, w.l.o.g., that $d(x, u_1) \leq d(x, u_2)$.

Thus,

$$d(c_{2,4}, u_1) \leq d(c_{2,4}, x) + d(x, u_1) \leq d(c_{2,4}, x) + d(x, u_2) = d(c_{2,4}, u_2) \leq r.$$



■ **Figure 6** $P_{1,3}$ and $P_{2,4}$ intersect at x . (a) x is a point of V_{opt} , and (b) x is an intersection point of the edges (a, b) and (c, d) .

Therefore, u_1 is of distance at most r from $c_{2,4}$ and, hence, is contained in $T_r(c_{2,4})$.

Case 2: x is not a point of V_{opt} . Thus, x is an intersection point of two edges (a, b) and (c, d) of G . Assume, w.l.o.g., that a is between x and u_1 and c is between x and u_2 ; see Figure 6(b).

- If $d(a, u_1) = d(c, u_2)$, then, by Lemma 1, at least one of the edges (a, d) and (b, c) is in G_{opt} . If (a, d) is in G_{opt} , then $d(c_{2,4}, a) = d(c_{2,4}, c)$, and hence,

$$d(c_{2,4}, u_1) \leq d(c_{2,4}, a) + d(a, u_1) = d(c_{2,4}, c) + d(c, u_2) = d(c_{2,4}, u_2) \leq r.$$

Therefore, u_1 is of distance at most r from $c_{2,4}$ and, hence, is contained in $T_r(c_{2,4})$. If (b, c) is in G_{opt} , then $d(c_{1,3}, a) = d(c_{1,3}, c)$, and hence,

$$d(c_{1,3}, u_2) \leq d(c_{1,3}, c) + d(c, u_2) = d(c_{1,3}, a) + d(a, u_1) = d(c_{1,3}, u_1) \leq r.$$

Therefore, u_2 is of distance at most r from $c_{1,3}$ and, hence, is contained in $T_r(c_{1,3})$.

- Otherwise, assume, w.l.o.g., that $d(a, u_1) < d(c, u_2)$. By Lemma 1, at least one of the edges (a, c) and (b, d) is in G_{opt} . If (b, d) is in G_{opt} , then $d(c_{2,4}, b) \leq d(c_{2,4}, c)$ and $d(b, u_1) \leq d(c, u_2)$. Hence,

$$d(c_{2,4}, u_1) \leq d(c_{2,4}, b) + d(b, u_1) \leq d(c_{2,4}, c) + d(c, u_2) = d(c_{2,4}, u_2) \leq r.$$

If (a, c) is in G_{opt} , then $d(c, u_1) \leq d(c, u_2)$. Thus,

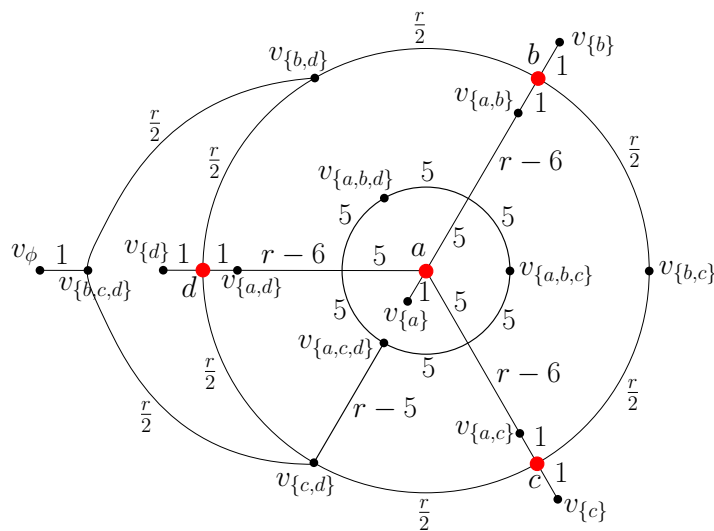
$$d(c_{2,4}, u_1) \leq d(c_{2,4}, c) + d(c, u_1) \leq d(c_{2,4}, c) + d(c, u_2) = d(c_{2,4}, u_2) \leq r.$$

In both cases, u_1 is of distance at most r from $c_{2,4}$ and, hence, is contained in $T_r(c_{2,4})$. ◀ Since $T_r(c_{2,4}) \cap S = \{u_2, u_4\}$ and $T_r(c_{1,3}) \cap S = \{u_1, u_3\}$, we have a contradiction. Therefore, the VC-dimension of $(V_{opt}, \mathcal{T}_r(G_{opt}))$ is at most 4.

To prove that the VC-dimension of $(V_{opt}, \mathcal{T}_r(G_{opt}))$ is at least 4, we show in Figure 7 a unit disk graph on a set of points V of diameter $2r$ and a subset $S = \{a, b, c, d\}$ of V , such that S can be shattered by $\mathcal{T}_r(G_{opt})$. The distance between every two points of S is r . For each subset $S' \subset S$, $S \cap T_r(v_{S'}) = S'$, and $S \cap T_r(a) = S$. ◀

5 Concluding remarks

In this paper, we consider the problem of computing a maximum subgraph of diameter d . We present the first constant-factor approximation algorithm for the problem in unit disk graphs, for any $d \geq 2$.



■ **Figure 7** Shattering the points $a, b, c,$ and d . $S \cap T_r(v_{S'}) = S'$, for each $S' \subset S$, and $S \cap T_r(a) = S$.

Our algorithm is simple and efficient, however, its analysis is not trivial and based on tools from the theory of hypergraphs with bounded VC-dimension, k -quasi planar graphs, fractional Helly theorems and several geometric properties of unit disk graphs. Unfortunately, the constant obtained is rather large. On the other hand, the most important feature of our algorithm is that its approximation factor is a constant independent of the diameter d . It is very easy to obtain an $O(d^2)$ approximation factor. Indeed, by a packing argument, any graph of diameter d can be covered by $O(d^2)$ cliques and as mentioned already the max-clique problem is in P .

Moreover, our algorithm works also for an abstract input of the unit disk graph without the geometric representation. It remains an open problem to determine whether MaxDBS for unit disk graphs is in P for $d \geq 2$.

Another interesting fact to note is that the shortest path metric on unit disk graphs does not have the so-called *constant doubling dimension*. It is easily seen that our algorithm has a constant factor approximation for graph families with constant doubling dimension.

Recall that a d -clique of a graph G is a set S of vertices of G , such that, for every two vertices in S , the shortest distance between them in G is at most d . Finding the maximum d -clique problem is closely related to MaxDBS. Unfortunately, our algorithm can not be directly extended to the maximum d -clique problem. Except for the $\frac{1}{2}$ -approximation algorithm of Pattillo et al. [19], for $d = 2$, there is no related work discussing the maximum d -clique problem in unit disk graphs. Hence, approximating the maximum d -clique problem in unit disk graphs is also an interesting open problem.

References

- 1 E. Ackerman. On the maximum number of edges in topological graphs with no four pairwise crossing edges. *Discrete Comput. Geom.*, 41:365–375, 2009.
- 2 P. K. Agarwal, B. Aronov, J. Pach, and M. Sharir. Quasi-planar graphs have linear number of edges. *Combinatorica*, 17:1–9, 1997.
- 3 R. D. Alba. A graph-theoretic definition of a sociometric clique. *J. Math. Sociol.*, 3:113–126, 1973.

- 4 M. T. Almeida and F. D. Carvalho. Integer models and upper bounds for the 3-club problem. *Networks*, 60:155–166, 2012.
- 5 Y. Asahiro, E. Miyano, and K. Samizo. Approximating maximum diameter-bounded subgraphs. In *LATIN, LNCS 6034*, pages 615–626, 2010.
- 6 B. Balasundaram, S. Butenko, and Trukhanov S. Novel approaches for analyzing biological networks. *J. Combin. Optim.*, 10:23–39, 2005.
- 7 I. M. Bomze, M. Budinich, P. M. Pardalos, and M. Pelillo. The maximum clique problem. *Handbook of Combinatorial Optimization*, pages 1–74, Kluwer Academic Publishers, 1999.
- 8 J.-M. Bourjolly, G. Laporte, and G. Pesant. An exact algorithm for the maximum k -club problem in an undirected graph. *European J. Oper. Res.*, 138:21–28, 2002.
- 9 A. Buchanan and H. Salemi. Parsimonious formulations for low-diameter clusters. http://www.optimization-online.org/DB_HTML/2017/09/6196.html, 2017.
- 10 F. D. Carvalho and M. T. Almeida. Upper bounds and heuristics for the 2-club problem. *European J. Oper. Res.*, 210:489–494, 2011.
- 11 M.-S. Chang, L.-J. Hung, C.-R. Lin, and P.-C. Su. Finding large k -clubs in undirected graphs. *Computing*, 95:739–758, 2013.
- 12 V. Chepoi, B. Estellon, and Y. Vaxès. Covering planar graphs with a fixed number of balls. *Discrete Comput. Geom.*, 37:237–244, 2007.
- 13 B. N. Clark, C. J. Colbourn, and D. S. Johnson. Unit disk graphs. *Discrete Math.*, 86:165–177, 1990.
- 14 A. Gräf, M. Stumpf, and G. Weißenfels. On coloring unit disk graphs. *Algorithmica*, 20(3):277–293, 1998.
- 15 S. Har-Peled. *Geometric Approximation Algorithms*. American Mathematical Society, Boston, USA, 2011.
- 16 J. Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. In *FOCS*, pages 627–636, 1996.
- 17 J. Matoušek. *Lectures on Discrete Geometry*. Springer, New York, USA, 2002.
- 18 J. Matoušek. Bounded VC-dimension implies a fractional Helly theorem. *Discrete Comput. Geom.*, 31:251–255, 2004.
- 19 J. Pattillo, Y. Wang, and S. Butenko. Approximating 2-cliques in unit disk graphs. *Discrete Appl. Math.*, 166:178–187, 2014.
- 20 J. Pattillo, N. Youssef, and S. Butenko. On clique relaxation models in network analysis. *European J. Oper. Res.*, 226:9–18, 2013.
- 21 A. Veremyev and V. Boginski. Identifying large robust network clusters via new compact formulations of maximum k -club problems. *European J. Oper. Res.*, 218:316–326, 2012.