Approximating Maximum Diameter-Bounded Subgraph in Unit Disk Graphs

A. Karim Abu-Affash[*] Paz Carmi[†] Anil Maheshwari[‡] Pat Morin[§]
Michiel Smid[¶] Shakhar Smorodinsky[||]

May 2, 2021

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 a fixed parameter $d \geq 1$, 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 Introduction

Let $G = (V, E)$ be a connected graph. The distance between two vertices u and v in G, denoted by $d(u, v)$, is the minimum number of edges on any path between u and v in G. The diameter of G is defined as the maximum distance between any two vertices in G, i.e., $\max_{u,v \in V} d(u, v)$. For a subset $V' \subseteq V$, let $G[V']$ be the subgraph of G induced by V'. In the *maximum diameter-bounded subgraph* problem (MaxDBS for short), we are given a connected graph G and an integer $d \geq 1$, and the goal is to compute a maximum subset $V' \subseteq V$, such that $G[V']$ is a graph of diameter at most d. A subgraph of diameter at most d is referred to in the literature as a *d-club*.

[*]Software Engineering Department, Shamoon College of Engineering, Beer-Sheva 84100, Israel, abuaa1@sce.ac.il.
[†]Department of Computer Science, Ben-Gurion University, Beer-Sheva 84105, Israel, carmip@cs.bgu.ac.il.
[‡]School of Computer Science, Carleton University, Ottawa, Canada, anil@scs.carleton.ca.
[§]School of Computer Science, Carleton University, Ottawa, Canada, morin@scs.carleton.ca.
[¶]School of Computer Science, Carleton University, Ottawa, Canada, michiel@scs.carleton.ca.
[||]Department of Mathematics, Ben-Gurion University, Beer-Sheva 84105, Israel, shakhar@math.bgu.ac.il.

For $d = 1$, MaxDBS is equivalent to the maximum clique problem, which is one of the fundamental problems in theoretical computer science [8]. It is not only NP-hard but even hard to approximate the maximum clique problem within a factor of $n^{1-\epsilon}$, for any $\epsilon > 0$, unless $P = NP$ [22]. MaxDBS is also known to have hardness of approximation of $n^{1/2-\epsilon}$, unless $P = NP$ [6].

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, 15]. 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 work, Clark, Colbourn and Johnson [13] provide a beautiful polynomial time algorithm to compute the maximum clique in unit disk graphs.

1.1 Related work

MaxDBS has been studied extensively in general graphs in the last two decades. Bourjolly et al. [9] showed that MaxDBS is NP-hard. Balasundaram et al. [7] proved that for any diameter parameter d, MaxDBS is NP-hard in graphs of diameter $d+1$. Asahiro et al. [5] showed that, for any $\epsilon > 0$ and a fixed $d \geq 2$, 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 $O(1.62^n \cdot poly(n))$ time. There are more results on solving MaxDBS by using various integer and linear programming formulations [4, 7, 9, 10, 21].

Asahiro et al. [6] 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 $d \geq 2$.

Chepoi et al. [12] studied MaxDBS in planar graphs. They showed that there exists a constant ρ, such that any planar graph G of diameter at most d can be covered with at most ρ balls of radius $d/2$ (a ball of center u and radius r consists of all vertices of G of distance at most r from u).

To the best of our knowledge, the hardness of MaxDBS in unit disk graphs is still open, for $d \geq 2$. 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

In MaxDBS, given a graph G and an integer $d \geq 2$, the goal is to compute a maximum d-club in G, i.e., a d-club with the maximum number of vertices. MaxDBS is a relaxation of the maximum clique problem and is motivated by cluster-detection that arise in a wide variety of 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* [7, 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 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.

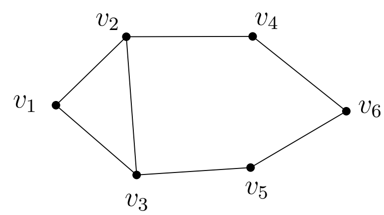

Figure 1: $S = \{v_1, v_2, v_3, v_4, v_5\}$ is a 2-clique but not a 2-club since the graph induced by S has a diameter 3.

1.3 Our contribution

In this paper, we adapt and extend the result in [12] for planar graphs to unit disk graphs. More precisely, we show that for any unit disk graph G and for any integer d, with $d = 2r$ or $d = 2r + 1$, there exists a constant $c < 1$ (which does not depend on d), such that if a maximal d-club of G contains n vertices, then a maximal ball of radius r in G contains at least cn vertices. An easy consequence of this result is a constant-factor approximation for MaxDBS in unit disk graphs.

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 is a well-known result, however, we give a simple proof to make the paper self contained.

Lemma 2.1. *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 $\min\{|ad|, |bc|\} \leq 1$. □

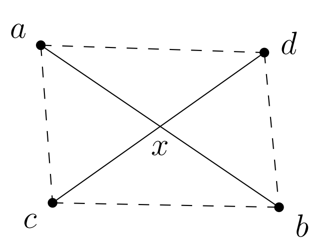

Figure 2: An illustration for the proof of Lemma 2.1.

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 every subset A' of A, there exists 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 shattered subset of X (if it exists); see [16] 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 $(X, \mathcal{R})$ 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 2.2 ([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 $k-1$ 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 corresponding points, so that (i) no edge contains 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 2.3. *Any complete topological graph on n vertices contains at least $\frac{\log(n-1)-c}{2 \log \log n} + 3$ 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. The distance between two vertices u and v in G, denoted by $d(u, v)$, is the minimum number of edges on any path between u and v in G.[1] Assuming that G is connected, the diameter of G is defined as the maximum distance between any two vertices in G, i.e., $\max_{u,v \in V} d(u, v)$. A *d-club* of G is an induced subgraph of G of diameter at most d. For a subset $V' \subseteq V$, let $G[V']$ be the subgraph of G induced by V'. Notice that $G[V']$ is the unit disk graph of the set V'. Given an integer $d \geq 2$, let V_{opt} be a maximum subset of V of size n, such that $G_{opt} = G[V_{opt}]$ is a d-club of G. In this section, we first present a polynomial-time approximation algorithm that computes a d-club of size at least cn, where c is a constant and d is even. Later, we show how to generalize this algorithm to the case of odd d's.

[1] 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.

Set $r = \frac{d}{2}$. For a vertex $u \in V$, let $B_G(r,u)$ denote the ball of radius r centered at u in G. That is, $B_G(r,u)$ contains all the vertices of distance at most r from u in G (i.e., $d(u,v) \leq r$).[2] Given a vertex $u \in V$, $B_G(r,u)$ can be computed using the breadth first search (BFS) algorithm in $O(|V|+|E|)$ time. Our algorithm computes $B_G(r,u)$, for each $u \in V$, and returns a ball B of maximum cardinality among these balls. It is clear that the subgraph of G induced by B is a d-club of G. We now prove that $|B| \geq cn$, where c is a constant.

Let $\mathcal{B}_r = \{B_{G_{opt}}(r,u) : u \in V_{opt}\}$ and let B^* be a ball of maximum cardinality in $\mathcal{B}_r$. Notice that, since G_{opt} is the subgraph of G induced by V_{opt}, G_{opt} is also a unit disk graph. Moreover, the distance between any two vertices u and v in G_{opt} is at least as the distance between them in G, which implies that $B_{G_{opt}}(r,u) \subseteq B_G(r,u)$, for every $u \in V_{opt}$. Therefore, $|B^*| \leq |B|$, and it is sufficient to prove that $|B^*| \geq cn$. From now on, we may consider the graph G_{opt} as the underlying graph G and thus, we refer to V_{opt} as V, and $B_{G_{opt}}(r,u)$ as $B(r,u)$. In the following, we show that G_{opt} can be covered by at most t balls of $\mathcal{B}_r$, which immediately proves that $|B^*| \geq cn = n/t$.

Let $(V,\mathcal{R})$ be the range space, where $\mathcal{R} = \mathcal{B}_r = \{B(r,u) : u \in V\}$. Thus, in the dual range space $(Y,\mathcal{R}^*)$ of $(V,\mathcal{R})$, we have $Y = \{y_R : R \in \mathcal{R}\} = \{y_{B(r,u)} : u \in V\}$, and, for each point $v \in V$, the set $\{y_{B(r,u)} : v \in B(r,u)\}$ is a range in $\mathcal{R}^*$. The following lemma has been proven in [12], however, we give the proof for completeness.

Lemma 3.1. *$(V,\mathcal{R})$ and $(Y,\mathcal{R}^*)$ are isomorphic.*

Proof. Since $v \in B(r,u)$ if and only if $u \in B(r,v)$, for fixed v and u, we have

$$\{y_{B(r,u)} : v \in B(r,u)\} = \{y_{B(r,u)} : u \in B(r,v)\}.$$

Now, by mapping $y_{B(r,u)}$ to u, we obtain that

$$Y = \{y_{B(r,u)} : u \in V\} = \{u : u \in V\} = V,$$

and, for each $v \in V$,

$$\{y_{B(r,u)} : v \in B(r,u)\} = \{u \in V : v \in B(r,u)\} = \{u \in V : u \in B(r,v)\} = B(r,v).$$

Hence, for each $v \in V$, the set $B(r,v)$ is a range in $\mathcal{R}^*$, which implies that $\mathcal{R}^* = \mathcal{R}$. Therefore, $(V,\mathcal{R})$ and $(Y,\mathcal{R}^*)$ are isomorphic. □

Theorem 3.2. *G_{opt} can be covered by at most t balls of $\mathcal{B}_r$, where t is a constant.*

Proof. The proof plan is as follows. We show (later in Section 4) that the VC-dimension of the range space $(V,\mathcal{R})$ is 4. Thus, by Lemma 3.1, the VC-dimension of the dual range space $(Y,\mathcal{R}^*)$ of $(V,\mathcal{R})$ is also 4. Then, we use Corollary 2.3 to show that there exists a constant $m \geq 5$, such that $(V,\mathcal{R})$ satisfies the $(m,5)$-property. Thus, by Theorem 2.2, there exists a set of at most t balls of $\mathcal{B}_r$ that cover all vertices of G_{opt}.

Let m be an integer such that $\frac{\log(m-1)-c'}{2\log\log m} + 3 \geq 6$, where c' is the constant from Corollary 2.3. Let A be a set of m balls of $\mathcal{B}_r$ and let $C = \{u_1, u_2, \ldots, u_m\}$ be the set of centers of the balls in 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

[2] We assume that the diameter of G is greater than d, otherwise, G is a d-club.

either empty or a path. Otherwise, $\delta(u_i, u_j)$ and $\delta(u_{i'}, u_{j'})$ intersect more than once and the subpaths between the intersection points are disjoint and have the same length; see Figure 3(a). In this case, we replace the subpaths of $\delta(u_{i'}, u_{j'})$ by the subpaths of $\delta(u_i, u_j)$ between the intersection points; see Figure 3(b). We can do this safely, since the 'new' path between $u_{i'}$ and $u_{j'}$ is also a shortest path between them.

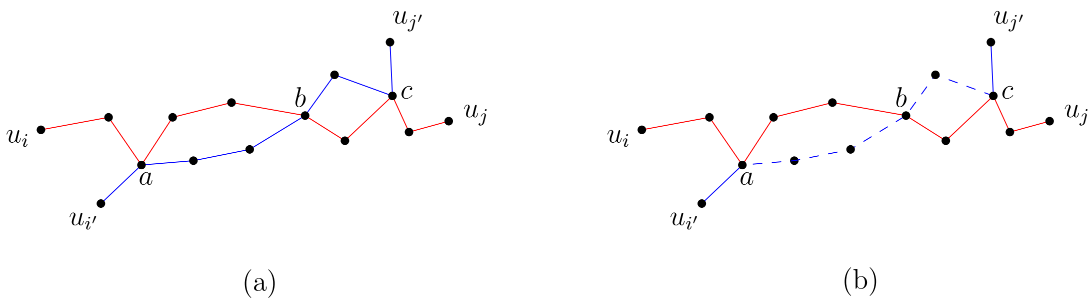

Figure 3: (a) $\delta(u_i, u_j)$ (red path) and $\delta(u_{i'}, u_{j'})$ (blue path) intersect at a, b, and c. (b) Replacing subpaths of $\delta(u_{i'}, u_{j'})$ by subpaths of $\delta(u_i, u_j)$ between the intersection points.

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 2.3, H' has at least 6 pairwise crossing edges. Let $P = \{\delta(u_1, u_{1'}), \delta(u_2, u_{2'}), \ldots, \delta(u_6, u_{6'})\}$ be the set of the corresponding 6 pairwise crossing paths in H'. For each $1 \leq i \leq 6$, let $u_{i,i'}$ be a point on $\delta(u_i, u_{i'})$ that belongs to both $B(r, u_i)$ and $B(r, u_{i'})$ (such a point exists since $d(u_i, u_{i'}) \leq 2r$).

Lemma 3.3. *Let $\delta(u_i, u_{i'})$ and $\delta(u_j, u_{j'})$ be two crossing paths from P and let x be their intersection point. Assume, w.l.o.g., that x is between $u_{i,i'}$ and u_i in $\delta(u_i, u_{i'})$, and between $u_{j,j'}$ and u_j in $\delta(u_j, u_{j'})$; see Figure 4. Then, either $u_{j,j'} \in B(r, u_i)$ or $u_{i,i'} \in B(r, u_j)$.*

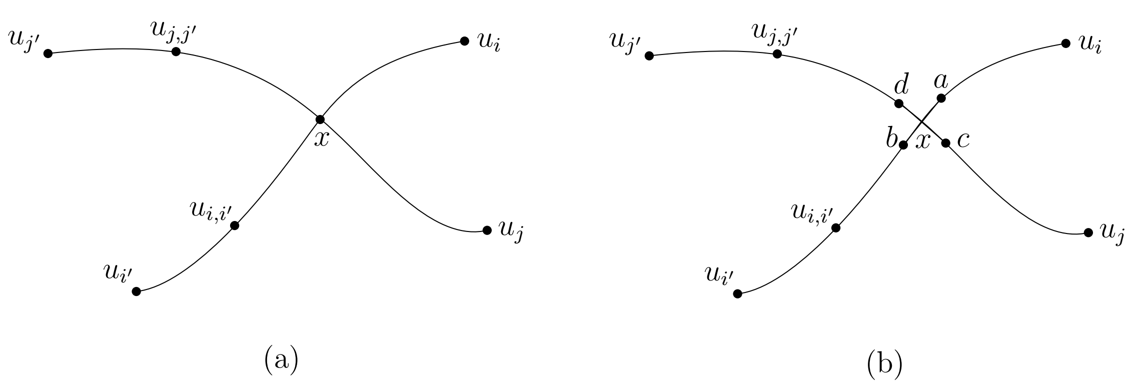

Figure 4: $\delta(u_i, u_{i'})$ and $\delta(u_j, u_{j'})$ intersect at x. (a) x is a point of V, and (b) x is an intersection point of the edges (a, b) and (c, d).

Proof. We distinguish between two cases.
Case 1: x is a point of V; 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 $B(r, u_i)$.
Case 2: x is not a point of V. 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_i and c is between x and u_j; see Figure 4(b).

- If $d(u_i, a) = d(u_j, c)$, then, since G_{opt} is a unit disk graph, by Lemma 2.1, 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 $B(r, u_i)$.
  - If (b, c) is in G_{opt}, then $d(a, u_{i,i'}) = d(c, u_{i,i'})$, and hence,

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

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

- Otherwise, assume, w.l.o.g., that $d(u_i, a) < d(u_j, c)$. Since G_{opt} is a unit disk graph, by Lemma 2.1, 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 $B(r, u_i)$.

□

Lemma 3.4. *$(V, \mathcal{R})$ satisfies the $(m, 5)$-property.*

Proof. By Lemma 3.3, for every two paths $\delta(u_i, u_{i'})$ and $\delta(u_j, u_{j'})$ in P, either $u_{j,j'} \in B(r, u_i) \cup B(r, u_{i'})$ or $u_{i,i'} \in B(r, u_j) \cup B(r, u_{j'})$. We construct a directed graph on the vertices $\{u_1, u_2, \ldots, u_6\}$, such that there is a directed edge from u_i to u_j if and only if $u_{j,j'} \in B(r, u_i) \cup B(r, u_{i'})$. 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 contained in at least 3 other balls, in addition to the balls $B(r, u_l)$ and $B(r, u_{l'})$. Thus, $u_{l,l'}$ is contained in at least 5 balls of A. Therefore, $(V, \mathcal{R})$ satisfies the $(m, 5)$-property. □

Now, by Theorem 4.1, the VC-dimension of the dual range space of $(V, \mathcal{R})$ is 4, and, by Lemma 3.4, $(V, \mathcal{R})$ satisfies the $(m, 5)$-property. Therefore, by Theorem 2.2, there exists a set of at most t balls of $\mathcal{B}_r$ that cover all vertices of V. □

Upper bound on c

We show, in Figure 5, a unit disk graph G of diameter $2r$ on n vertices for which every ball of radius r does not contain more than $\frac{n}{3}$. G contains $n = 16r$ points and its diameter is $d = 2r$. Each ball 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 ball of radius r can cover at most 2 cliques. Therefore, for sufficiently large n, we have $c \leq \frac{1}{3}$.

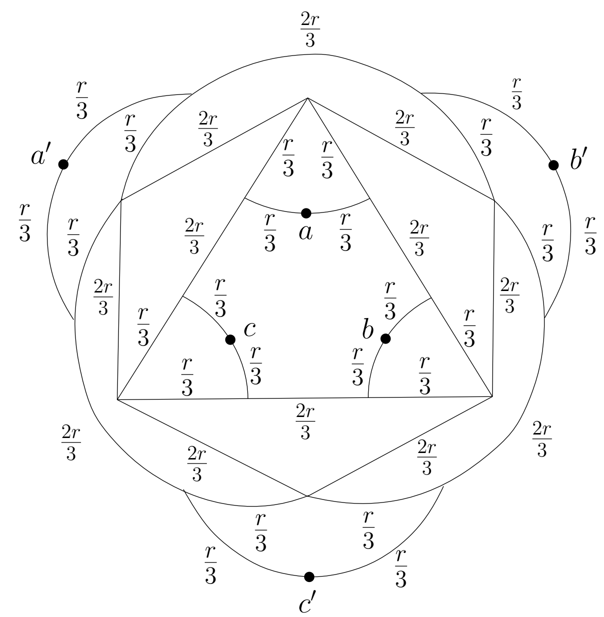

Figure 5: G contains $16r$ points and its diameter is $d = 2r$. Each ball of radius r covers at most $6r$ points.

Generalization for odd d

In this section, we extend our result 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 ball $B(r,u)$ of size at least $cn_{d+1} \geq cn_d$. Let $G_r(u)$ be the subgraph induced by $B(r,u)$, and notice that $G_r(u)$ is a $(d+1)$-club but may not be a d-club. In the following lemma, we show that there is a subgraph of $G_r(u)$ of diameter $d-1$ that contains at least $1/12$ of the vertices of $B(r,u)$.

Lemma 3.5. *The vertices of $G_r(u)$ can be covered by at most 12 balls of radius $r-1$.*

Proof. Let $D_2(u) = \{v \in B(r,u) : d(u,v) = 2\}$, i.e., the set of all vertices of $B(r,u)$ of distance two from u. Let I be a maximal independent set of $D_2(u)$. By the packing argument in unit disk graphs [14], we have $|I| \leq 12$. Let v be a vertex in $B(r,u)$, and let $\delta(u,v)$ be a shortest path between u and v in $G_r(u)$. 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 $B(r,u)$ is contained in $B(r-1,x)$, for some $x \in I$, and therefore, $B(r,u)$ is covered by $\bigcup_{x \in I} B(r-1,x)$. □

By Lemma 3.5, we can find a ball $B^*(r-1,x)$ that contains at least 1/12 of the vertices of $B(r,u)$. Since $r = \frac{d+1}{2}$, the graph induced by $B^*(r-1,x)$ is a d-club of G and has a diameter at most $2(r-1) = d-1$.

The following theorem summarizes the result of this section.

Theorem 3.6. *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 $c \in (0, \frac{1}{3}]$.*

4 The VC-Dimension of $(V, \mathcal{R})$

Recall that G_{opt} is a maximum d-club of G, V is the set of vertices of G_{opt}, and $\mathcal{B}_r = \{B(r,u) : u \in V\}$ is the set of all balls of radius r centered at vertices of V. Recall also that $(V, \mathcal{R})$ is the range space, in which $\mathcal{R} = \mathcal{B}_r$. In this section, we prove the following theorem.

Theorem 4.1. *The range space $(V, \mathcal{R})$ has VC-dimension 4.*

Proof. The proof is similar to the proof of Proposition 1 in [12]. We first prove that the VC-dimension of $(V, \mathcal{R})$ is at most 4. For the sake of contradiction, suppose that there exists a subset $S = \{u_1, u_2, u_3, u_4, u_5\}$ of V, such that S is shattered by $\mathcal{R}$. Thus, for each $1 \leq i < j \leq 5$, there is a ball $B(r, c_{i,j})$ in $\mathcal{B}_r$, such that $B(r, c_{i,j}) \cap S = \{u_i, u_j\}$. Let $T_r(c_{i,j})$ be a BFS-tree of radius r rooted at $c_{i,j}$ in G_{opt}, and 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.

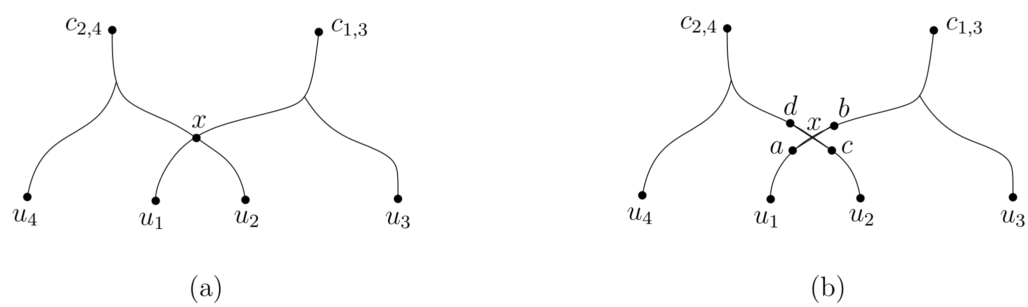

Figure 6: $P_{1,3}$ and $P_{2,4}$ intersect at x. (a) x is a point of V, and (b) x is an intersection point of the edges (a,b) and (c,d).

Lemma 4.2. *Either $u_1 \in B(r, c_{2,4})$ or $u_2 \in B(r, c_{1,3})$.*

Proof. The proof is similar to the proof of Lemma 3.3. We distinguish between two cases.
Case 1: x is a point of V; see Figure 6(a). Assume, w.l.o.g., that $d(x,u_1) \leq d(x,u_2)$. Thus,

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

Therefore, u_1 is of distance at most r from $c_{2,4}$ and, hence, is contained in $B(r, c_{2,4})$.
Case 2: x is not a point of V. 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, since G_{opt} is a unit disk graph, by Lemma 2.1, 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 $B(r, c_{2,4})$. 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 $B(r, c_{1,3})$.

- Otherwise, assume, w.l.o.g., that $d(a, u_1) < d(c, u_2)$. Since G_{opt} is a unit disk graph, by Lemma 2.1, 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 $B(r, c_{2,4})$.

$\square$

Since $B(r, c_{2,4}) \cap S = \{u_2, u_4\}$ and $B(r, c_{1,3}) \cap S = \{u_1, u_3\}$, we have a contradiction. Therefore, the VC-dimension of $(V, \mathcal{R})$ is at most 4.

To prove that the VC-dimension of $(V, \mathcal{R})$ 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 $\mathcal{R} = \{B(r, u) : u \in V\}$. The distance between every two points of S is r. For each subset $S' \subset S$, $S \cap B(r, v_{S'}) = S'$, and $S \cap B(r, a) = S$. $\square$

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 $d \geq 2$. Our algorithm is simple and efficient, however, its analysis is not trivial and based on interesting 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, this constant does not depend on the diameter d. 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 $d \geq 2$.

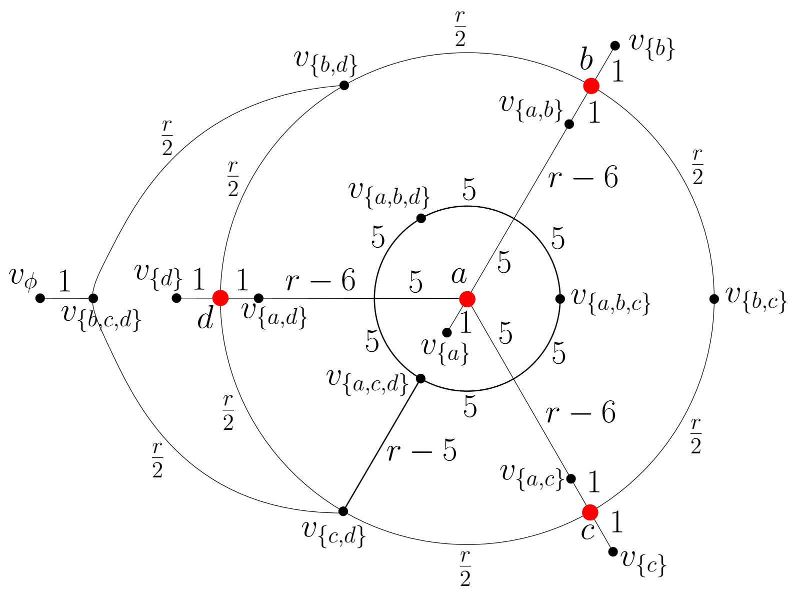

Figure 7: Shattering the points a, b, c, and d. $S \cap B(r, v_{S'}) = S'$, for each $S' \subset S$, and $S \cap B(r, a) = S$.

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. As mentioned in Section 1.2, in general graphs, every d-club is a d-clique, but not vice versa. This holds also for unit disk graphs (the graph in Figure 1 can be easily realized as a unit disk graph). 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.

6 Acknowledgments

The authors would like to thank the Fields Institute for hosting the workshop in Ottawa and for their financial support. The authors would also like to thank the anonymous reviewers for their helpful comments and suggestions.

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, Y. Doi, E. Miyano, K. Samizo, and H. Shimizu. Optimal approximation algorithms for maximum distance-bounded subgraph problems. *Algorithmica*, 80:1834–1856, 2018.

[6] Y. Asahiro, E. Miyano, and K. Samizo. Approximating maximum diameter-bounded subgraphs. In *LATIN, LNCS 6034*, pages 615–626, 2010.

[7] B. Balasundaram, S. Butenko, and Trukhanov S. Novel approaches for analyzing biological networks. *J. Combin. Optim.*, 10:23–39, 2005.

[8] 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.

[9] 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.

[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] G. D. da Fonseca, and C. M. H. de Figueiredo, and V. G. Pereira de Sá, and R. C. S. Machado. Efficient sub-5 approximations for minimum dominating sets in unit disk graphs. *Theor. Comput. Sci.*, 540–541:70–81, 2014.

[15] A. Gräf, M. Stumpf, and G. Weißenfels. On coloring unit disk graphs. *Algorithmica*, 20(3):277–293, 1998.

[16] S. Har-Peled. *Geometric Approximation Algorithms.* American Mathematical Society, Boston, USA, 2011.

[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.

[22] D. Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory Comput.*, 3:103–128, 2007.