Uniformity of point samples in metric spaces using gap ratio

Arijit Bishnu[1], Sameer Desai[1], Arijit Ghosh[2], Mayank Goswami[2], and Subhabrata Paul[1]

1 **ACM Unit, Indian Statistical Institute, India**
2 **Department: Algorithm and complexity, Max-Planck-Institut für Informatik, Saarbrücken, Germany.**

Abstract

Teramoto et al. [23] defined a new measure of uniformity of point distribution called the *gap ratio* that measures the uniformity of a finite point set sampled from $\mathcal{S}$, a bounded subset of $\mathbb{R}^2$. We attempt to generalize the definition of this measure over all metric spaces. While they look at online algorithms minimizing the measure at every instance, wherein the final size of the sampled set may not be known a priori, we look at instances in which the final size is known and we wish to minimize the final gap ratio. We solve optimization related questions about selecting uniform point samples from metric spaces; the uniformity is measured using gap ratio. We give lower bounds for specific as well as general instances, prove hardness results on specific metric spaces, and a general approximation algorithm framework giving different approximation ratios for different metric spaces.

Keywords and phrases Gap ratio, dispersion, discrepancy, uniformity measure, metric space, lower bounds, hardness, approximation

1 Introduction

Generating uniformly distributed points over a specific domain has applications in digital halftoning; see [2, 23, 25] and the references therein, numerical integration [10, 18], computer graphics [10], etc. Meshing also requires uniform distribution of points over a region of interest [5]. There are different measures of uniformity of points that we discuss below.

One such notion is the discrepancy [10, 18] of a point set. The intuitive notion of discrepancy of a point sequence P (lying inside a unit square $\mathcal{S}$) is captured by the ratio of the points lying inside an arbitrary set B to the measure (area) of B. B can be any geometric shape, e.g. square, circle in $\mathbb{R}^2$, hyperrectangle, hypersphere in any $\mathbb{R}^d$. For all possible scales and placements of B that encloses points of P, we would get different values of discrepancy; usually the supremum value defines discrepancy. For a formalization of this notion, an interested reader is referred to [10, 18]. Let $|P| = n$ and $\mathrm{vol}(B)$ denote the area of B. The expected number of points that would lie inside B if P is distributed uniformly and independently at random is $n \cdot \mathrm{vol}(B)$. Let $D(P, B)$ denote the deviation of P from uniform distribution inside a particular B, i.e. $D(P, B) = n \cdot \mathrm{vol}(B) - |P \cap B|$. Let $\mathcal{R}$ denote the set of all shapes similar to B. The quantity $D(P, \mathcal{R}) = \sup_{R \in \mathcal{R}} |D(P, R)|$ is the *discrepancy* of P for shapes similar to B. The above definition is for a fixed P. Now, we extend the definition for any n. The function $D(n, \mathcal{R}) = \inf_{P \subset S \text{ \& } |P|=n} D(P, \mathcal{R})$ captures the notion of the least possible discrepancy of a point set sized n. To compute uniformity using the above measure, the quantity is to be computed for all possible scales and positions of B.

Another notion of uniformity has been captured by the idea of maximizing the minimum distance among points inside $\mathcal{S}$. This is equivalent to packing equal radius circles inside $\mathcal{S}$

[11, 19, 20, 21]. Packing equal radius circles has remained a difficult problem [17]. This measure does not take into effect large empty areas inside $\mathcal{S}$.

We observe that both of the above measures are hard to compute. Motivated by problems in digital halftoning, Teramoto et al. [23] defined a new measure of uniformity called the *gap ratio* that measures uniformity in $\mathbb{R}^2$. The basic notion of this uniformity measure is a ratio between the maximum and minimum gaps among points. The minimum gap is the distance between the closest pair of points of P. The maximum gap is the radius of the maximum empty circles among points in P and is linked to the Voronoi diagram of P.

Definition of gap ratio

Teramoto et al. [23], who introduced the problem, were interested in the online version of the gap ratio problem. They wanted to insert k points one by one in such a way that uniformity is achieved at every point insertion. We first recall their definition.

Let $\mathbb{S}^d = [0,1]^d$ be the unit cube in the d-dimensional space $\mathbb{R}^d$ and $P = (p_1, \ldots, p_k)$ be a sequence of k points contained in $\mathbb{S}^d$, and inserted in the order $p_1, p_2, \ldots, p_k$. For each $i = 1, \ldots, k$, let a subsequence P_i of P be the first i points of P. Let $\mathcal{CH}(P_i)$ denote the convex hull of P_i. For each point sequence P_i, define the current point set as $S_i = P_i \cup S_0$, where S_0 is the set of the 2^d extremum (corner) points of $\mathbb{S}^d$. Using the smallest among all pairwise distances in S_i, the minimum gap r_i is defined as $r_i := \min_{p,q\in S_i, p\neq q} \delta(p,q)/2$, where $\delta(p,q)$ is the Euclidean distance between two points p and q. The maximum gap is defined as $R_i := \max_{p\in\mathbb{S}^d} \min_{q\in S_i} \delta(p,q)$. The minimum and maximum gaps are analogous to the radius of the smallest and largest empty circles respectively of points in P_i. An empty circle is a circle whose centre is located inside $\mathcal{CH}(P_i)$ and contains no point of P_i. The gap ratio for P_i is denoted as $GR_i := R_i/r_i$. For a point sequence P, the maximum gap ratio is defined as $GR_P := \max_{i=1,\ldots,k} GR_i$. For a fixed integer k, the optimal gap ratio GR_k for any k-point sequence is defined as $GR_k := \min\{GR_P \mid P \text{ is any } k\text{-point sequence in } \mathbb{S}^d\}$. Given d and k, we want to find a point sequence P of k points in $\mathbb{S}^d$ that achieves the optimal gap ratio GR_k. We extend the above definition of Teramoto et al. [23] without considering the sequence of insertion. Also, notice a fact that $\mathbb{S}^d$ can be replaced by any metric space.

▶ Definition 1. Let $(\mathcal{M}, \delta)$ be a metric space and let P be a set of k points sampled from $\mathcal{M}$. Define the minimum gap as $r_P := \min_{p,q\in P, p\neq q} \delta(p,q)/2$. The maximum gap brings into play the interrelation between the metric space $\mathcal{M}$ and $P(\subset \mathcal{M})$, the set sampled from $\mathcal{M}$, and is defined as $R_P := \sup_{q\in\mathcal{M}} \delta(q,P)$, where $\delta(q,P) := \min_{p\in P} \delta(q,p)$ is the distance of q from the set P. The gap ratio for the point set P is defined as $GR_P := R_P/r_P$. In the rest of the paper, we would mostly not use the subscript P.

The *maximum gap* and the *minimum gap* are not actually maximum and minimum of the same quantity, so the ratio need not be greater than 1. Consider as an example, $\mathcal{M}$ to be two unit (diameter=1) balls in the Euclidean plane with centres distance 10 apart, and P to be two points, one in each ball. In this case, the gap ratio $GR = \frac{R}{r} \leqslant \frac{2}{9}$.

The maximum gap calculates how far a point in $\mathcal{M}$ can be from P and the minimum gap calculates how close two points in P can be to each other. In a geometric sense, this means that the maximum gap is analogous to the minimum radius required to cover $\mathcal{M}$ with equally sized balls (i.e., *covering balls*) around each point of P, and the minimum gap is the maximum radius of equally sized balls centred around each point in P (i.e., *packing balls*) having pairwise disjoint interiors. In a uniformly distributed point set, we expect the covering balls to have minimum overlap among themselves (*thin covering*) and the packing balls to be as close as possible (*tight packing*). So, we expect the maximum gap to be minimised and the minimum gap to be maximised to measure uniformity. Thus the gap ratio can be a good measure of estimating uniformity of point samples.

Teramoto et al.'s [23] definition deals with a continuous metric space, but $\mathcal{M}$, as in Definition 1 can be both continuous and discrete. Using this generalized definition, we can pose combinatorial optimization questions where $\mathcal{M}$, for example, can be a set S of N points, and we would like to choose a subset $P \subset S$ of n points from S, such that the gap ratio is minimized. Asano [2] in his work opened this area of research, where he asked discrepancy like questions in a discrete setting. Asano opined that the discrete version of this discrepany-like problem will make it amenable to ask combinatorial optimization related questions. We precisely do that in this paper for different metric spaces.

We summarize the results of the paper in the following table.

Metric Space		Lower Bounds	Hardness	Approximation
General		none	yes	2-approx. hard
Discrete	Graph (connected)	$\frac{2}{3}$	yes	approx. factor: 3 $\frac{3}{2}$-approx. hard
Continuous	Path-Connected	1	yes	approx. factor: 2
	Unit Square in $\mathbb{R}^2$	$\frac{2}{\sqrt{3}} - o(1)$	-	approx. factor: $\frac{2}{\frac{2}{\sqrt{3}} - o(1)}$

In Section 2, we review the previous results related to gap ratio. We deal with lower bounds for gap ratio in Section 3. We show NP-hardness results for some variants in Sections 4 and approximation hardness results in Section 5. We discuss some constant-ratio approximation algorithms in Section 6. Our approximation algorithms work for both continuous and discrete metric spaces. Finally, Section 7 concludes the paper.

2 Previous results

Teramoto et al. [23] introduced this problem motivated by combinatorial approaches and applications in digital halftoning [1, 3, 4, 22]. They proved a lower bound of $2^{\lfloor k/2 \rfloor/(\lfloor k/2 \rfloor+1)}$ for the gap ratio in the one dimensional case where k points are inserted in the interval $[0, 1]$. They also proposed an algorithm that achieves the optimal gap ratio in the one dimensional case in linear time. But the problem was difficult to solve in two or higher dimensions. They got a gap ratio of 2 in 2-dimension using ideas of Voronoi insertion where the new point was inserted in the centre of a maximum empty circle [7]. They also proposed a local search based heuristic for the problem and provided experimental results in support.

Asano [1] discretized the above problem and showed a gap ratio of at most 2 where k integral points are inserted in the interval $[0, n]$ where n is also a positive integer and $0 < k < n$. He also showed that a point sequence may not always exist if a gap ratio of strictly less than 2 is needed, but a tight upper bound on the length of the sequence for given values of k and n can be proved. No optimal solution for the two or higher dimensions is known.

Zhang et al. [25] focused on the discrete version of the problem and proposed an insertion strategy that achieved a gap ratio of at most $2\sqrt{2}$ in a bounded two dimensional grid. They also showed that no online algorithm can achieve a gap ratio strictly less than 2.5 for a 3×3 grid.

The above shows that both continuous and discrete versions of gap ratio problem have been looked at and some efforts have been made at proving lower bounds. In this paper, we initiate a generalised study on combinatorial optimization problems related to gap ratio for different metric spaces. We also show some lower bounds, that lead to some approximation guarantees.

3 Lower bounds

Here we study the lower bounds for the gap ratio problem in both continuous and discrete metric spaces. We first point out that there exist metric spaces, both continuous and discrete, having point samples with arbitrarily small gap ratio. See Appendix-A for details.

3.1 Discrete metric space

In coming up with a discrete metric space that can have arbitrarily small gap ratio, we considered a graph that is not connected. Next we study the lower bound of gap ratio on a metric space $\mathcal{M}$ which is the vertex set V of an undirected connected graph $G = (V, E)$. The distance between a pair of vertices is the length of the shortest path between them.

▶ Lemma 2. The lower bound for the gap ratio is $\frac{2}{3}$ when the metric space $\mathcal{M}$ is a connected undirected graph and the lower bound is achieved only when $R = 1$ and $r = \frac{3}{2}$.

Proof. Suppose a set of vertices $P \subset \mathcal{M}$ is sampled. Let the closest pair of vertices in P be distance q apart. Thus $r = \frac{q}{2}$. Now between these two vertices, there is a path of $q - 1$ vertices in $\mathcal{M} \setminus P$. Among these $q - 1$ vertices, the vertex farthest from P is at a distance $\lfloor \frac{q}{2} \rfloor$ from P. Thus $R \geqslant \lfloor \frac{q}{2} \rfloor$ and $GR = \frac{R}{r} \geqslant \frac{2}{q} \lfloor \frac{q}{2} \rfloor$. Note that, when $q = 1$, clearly we have a gap ratio greater or equal to 2. Now, we analyze this expression for even and odd values of q.

- **q is even:** In this case, $GR \geqslant \frac{2}{q} \lfloor \frac{q}{2} \rfloor = \frac{2}{q}\frac{q}{2} = 1$.
- **q is odd and $q \geqslant 3$:** In this case, $GR \geqslant \frac{2}{q} \lfloor \frac{q}{2} \rfloor = \frac{2}{q} \left(\frac{q}{2} - \frac{1}{2} \right) = \frac{q-1}{q}$. It is simple to see that this function is monotonically increasing. This means the lowest value of the gap ratio occurs at $q = 3$ where it becomes $\frac{2}{3}$.

It is important to see here that $\frac{q-1}{q}$ is a lower bound to the gap ratio for odd q and 1 is a lower bound to the gap ratio for even q. Thus when $q > 3$, either q is odd and $q \geqslant 5$, in which case we have $\frac{q-1}{q} \geqslant \frac{4}{5} > \frac{2}{3}$, or q is even, in which case the lower bound is $1 > \frac{2}{3}$. Thus, the gap ratio $GR = \frac{2}{3}$ implies $q = 3$ which means $r = \frac{3}{2}$. Therefore, $R = GR \times r = 1$. Hence, $GR = \frac{2}{3}$ only when $R = 1$ and $r = \frac{3}{2}$. For example, a path on $3k+1$ vertices achieves a gap ratio of $2/3$ if we select every third vertex starting from the first vertex.

Hence, the lower bound for the gap ratio is $\frac{2}{3}$. ◀

3.2 Continuous metric space

We can have arbitrarily small gap ratio for continuous spaces that are disconnected. However, if we consider path connected spaces we have the following bound.

▶ Lemma 3. The lower bound of gap ratio is 1 when $\mathcal{M}$ is a path connected metric space.

Proof. In a connected metric space $(\mathcal{M}, \delta)$, consider a sampled point set P. Suppose the closest pair of points $x, y \in P$ is distance $2r$ apart. Consider disks of radius r around each point of P. This set of disks must have pairwise disjoint interiors as x and y are the closest pair of points in P. Consider a point $z \in \mathcal{M}$ on the boundary of the disk around x. There must be such a point, else, we have a contradiction to path-connectedness of $\mathcal{M}$. Note that z must be at distance r from P. Hence, $R \geqslant r$. ◀

Next we consider the continuous metric of unit square S in $\mathbb{R}^2$ as in Teramoto et al.'s problem [23]. To prove the lower bound on gap ratio, we appeal to packing and covering. To find a possible lower bound on the gap ratio, we would want to increase r and reduce R, as much as possible. To this end we define packing and covering densities.

▶ Definition 4 (Packing and covering densities [16] [24]). The density of a family $\mathcal{S}$ of sets with respect to a set C of finite positive Lebesgue measure is defined as $d(\mathcal{S},C) = \left(\sum_{S\in\mathcal{S},S\cap C\neq\emptyset}\mu(S)\right)/\mu(C)$, where μ is the Lebesgue measure. If C is the plane then we define the density as follows. Let $C(r)$ denote the disk of radius r centred at the origin. Then we have $d(\mathcal{S},C) = \lim_{r\to\infty} d(\mathcal{S},C(r))$. If the limit on the right hand side does not exist then we consider lower density defined by $d_-(\mathcal{S}) = \lim_{r\to\infty}\inf d(\mathcal{S},C(r))$, and the upper density defined by $d_+(\mathcal{S}) = \lim_{r\to\infty}\sup d(\mathcal{S},C(r))$. The *packing density* $d_p(K)$ of a convex body K is defined to be the least upper bound of the upper densities of all packings of the plane with copies of K, and, analogously, the *covering density* $d_c(K)$ of K is the greatest lower bound of the lower densities of all coverings of the plane with copies of K.

▶ Lemma 5. Let the metric space $\mathcal{M}$ be the unit square in a Euclidean plane. Then the lower bound for gap ratio is $\left(\frac{2}{\sqrt{3}} - o(1)\right)$, where k is the cardinality of $P \subset \mathcal{M}$.

Proof. Let $2r$ be the minimum pairwise distance between the point of P. Consider a circle of radius r around each point of P. This forms a packing of k circles of radius r in a square of side length $(1+2r)$. Suppose the density of such a packing is d_1. Now, we can tile the plane with such squares packed with circles. Thus we have a packing of the plane of density d_1. It is known that the density of the densest packing of equal circles in a plane is $\pi/\sqrt{12}$ [16]. Then obviously $d_1 \leqslant k\pi/\sqrt{12}$ as we have packed the plane with density d_1. Hence, $d_1 = k\pi r^2/(1+2r)^2 \leqslant \pi/\sqrt{12}$. Consequently we have, $r \leqslant \left(\sqrt{k\sqrt{12}} - 2\right)^{-1}$.

On the other hand, let $R = \sup_{x\in\mathcal{M}} \delta(x,P)$. Consider a circle of radius R around each point of P. We claim that these circles form a covering of the unit square. We prove by contradiction. Assume that there exists at least one point x in the unit square that is not covered by any circle. Among all points of P, let p be the closest to x. Consider the circle around p. This circle does not cover x. Thus, $R < \sup_{x\in\mathcal{M}} \delta(x,P)$. Hence, we have a contradiction. Thus, the circles of radius R around each point of P form a covering of the unit square with k circles of radius R. Suppose the density of such a covering is D_1. Now, we can tile the plane with this unit square. Thus we have a covering of the plane with density D_1. It is known that the density of the thinnest covering of the plane by equal circle is $2\pi/\sqrt{27}$ [16]. Then obviously $D_1 \geqslant 2\pi/\sqrt{27}$ as we have covered the plane with density D_1. Thus we have, $D_1 = k\pi R^2/1 \geqslant 2\pi/\sqrt{27}$, giving us $R \geqslant \sqrt{2}/\sqrt{k\sqrt{27}}$. Hence, the gap ratio is $\frac{R}{r} \geqslant \left(\sqrt{k\sqrt{12}} - 2\right)\sqrt{2}/\sqrt{k\sqrt{27}} = \frac{2}{\sqrt{3}} - o(1)$. ◀

This lower bound has a bearing on the problem posed by Teramoto et al. [23]. They had obtained a gap ratio of 2 in the online version, whereas, the lower bound for the problem is asymptotically 1.1547.

4 NP-hardness results

In this section, we show that finding minimum gap ratio is NP-hard for both discrete and continuous metric spaces.

▶ Definition 6 (The gap ratio problem). Given a metric space $(\mathcal{M},\delta)$, an integer k and a parameter g, we need to find a set $P \subset \mathcal{M}$ such that $|P| = k$ and $GR_P \leqslant g$.

4.1 Discrete case: graph metric space

In this subsection, we show that the problem of finding minimum gap ratio is NP-complete even for graph metric space. To this end, we need the concept of a variation of domination problem, called efficient domination problem. A subset $D \subseteq V$ is called an *efficient dominating set* of $G = (V, E)$ if $|N_G[v] \cap D| = 1$ for every $v \in V$, where $N_G[v] = \{v\} \cup \{x | vx \in E\}$. An efficient dominating set is also known as *independent perfect dominating set* [6]. Given a graph $G = (V, E)$ and a positive integer k, the efficient domination problem is to find an efficient dominating set of cardinality at most k. The efficient domination problem is known to be NP-complete [9].

▶ Theorem 7. In graph metric space, gap ratio problem is NP-complete.

Proof. First note that, the gap ratio problem in graph metric space is in NP. To prove the hardness, we use a reduction from efficient domination problem, to the gap ratio problem. Given an instance of efficient domination problem $G = (V, E)$ and k, set $\mathcal{M} = V$ as the metric space and the shortest path distance between two vertices as the metric δ.

▶ Claim 8. $G = (V, E)$ has an efficient dominating set of cardinality k if and only if there exists a sampled set P of k points (vertices) whose gap ratio is $2/3$.

Proof. Let D be an efficient dominating set of G of cardinality k. Set the sampled set $P = D$. Since D is a dominating set, $R = 1$. Again, note that, there cannot be a pair of vertices x and y in D such that $\delta(x, y) < 3$. This is because, if there exists a pair of vertices $x, y \in D$ with $\delta(x, y) \leqslant 2$, then there exists a vertex $v \in V$ such that $x, y \in (N_G[v] \cap D)$. On the other hand, if the closest pair of vertices $x, y \in D$ has $\delta(x, y) \geqslant 4$, then there exists a vertex $v \in V$ such that $(N_G[v] \cap D) = \emptyset$. Hence, there must be a pair of vertices x and y in D such that $\delta(x, y) = 3$. Therefore, $r = 3/2$ and the gap ratio becomes $2/3$.

Conversely, let P be the sampled set having k points with gap ratio $2/3$. From Lemma 2, we know that gap ratio $2/3$ is achievable only when $R = 1$ and $r = 3/2$. So, clearly $D = P$ is an efficient dominating set of G of cardinality k. ◀

Thus the gap ratio problem is NP-complete for graph metric space. ◀

4.2 Continuous case

In this subsection, we show that gap ratio is hard for a continuous metric space. To show this hardness, we reduce from the problem of system of distant representatives (SDR) in unit disks [12]. We first define the problem.

▶ Definition 9 ($S(q, l)$-DR). [12] Given a parameter $q > 0$ and a family $\mathcal{F} = \{F_i | i \in I, F_i \subseteq X\}$ of subsets of X, a mapping $f : I \to X$ is called a *System of q-Distant Representatives* (shortly an *Sq-DR*) if

(1) $f(i) \in F_i$ for all $i \in I$ and
(2) distance between $f(i)$ and $f(j)$ is at least q, for $i, j \in I$ and $i \neq j$.

When the family $\mathcal{F}$ is a set of unit diameter disks with centres that are at least l distance apart, we denote the mapping by $S(q, l)$-DR.

Fiala et al. proved that $S(1, l)$-DR is NP-hard [12]. For the general version $S(q, l)$-DR, we give a proof sketch using Fiala et al.'s technique. Note that for $q \leqslant l$, the centres of the disks suffice as our representatives. So assume that $q > l$. We restate a generalised version of their result below. The proof of the following theorem is in Appendix-B.

▶ Theorem 10. $S(q, l)$-DR is NP-hard for $q > l$ on the Euclidean plane.

▶ Lemma 11. $S(q,l)$-DR-1 is NP-complete for $q > l$, where $S(q,l)$-DR-1 denotes $S(q,l)$-DR with the constraint that one representative point should be on the boundary of one of the disks.

Proof. Clearly, a solution to $S(q,l)$-DR-1 is a solution to $S(q,l)$-DR. Conversely, if a solution of $S(q,l)$-DR is given, we can translate the entire solution point set until one point hits the boundary to obtain a solution to $S(q,l)$-DR-1.

It is easy to see that $S(q,l)$-DR-1 is in NP. Hence, it is NP-complete for $q > l$. ◀

▶ Theorem 12. Let $\mathcal{M}$ be a continuous metric space. Then, it is NP-hard to find a finite set $P \subset \mathcal{M}$ of cardinality k such that the gap ratio of P is at most $\frac{2}{q}$ for some $q > 2$.

Proof. We show that if there is a polynomial algorithm to find a finite set $P \subset \mathcal{M}$ of cardinality k such that the gap ratio of P is at most $\frac{2}{q}$ for some $q > 2$, then there is also a polynomial algorithm for $S(q,l)$-DR-1.

Consider an instance of $S(q,l)$-DR-1, a family $\mathcal{F} = \{F_1, F_2, \ldots, F_k\}$ of k disks of unit diameter such that their centres are at least distance l apart, where $q > l > 2$ (even with this restriction the proof of Theorem 10 goes through).

We run the algorithm for the gap ratio problem k times, each time on a separate instance. The instance for the ith iteration would have the disks $\{F_j | j \neq i\}$ and a circle of unit diameter with its centre being the same as the centre of F_i.

▶ Claim 13. If a single iteration of the above process results "yes", then we have a solution to the $S(q,l)$-DR-1 instance.

Proof. Suppose that the gap ratio of a given point set is at most $\frac{2}{q}$ for the ith instance. If it so happens that two points are within the same disk, then $r \leqslant \frac{1}{2}$. Thus for the gap ratio to fall below $\frac{2}{q}$ we need $R \leqslant \frac{2r}{q} \leqslant 1/q < 1$. But considering the number of points that we are choosing, we must have an empty disk, which would contain a point x such that $R \geqslant d(P,x) \geqslant l - \frac{1}{2} > 1$, giving us a contradiction. Thus we have that each disk contains exactly one point from P. Since, $l > 2$ and F_i is a circle, $R = 1$. Thus, we get $r = \frac{1}{GR} \geqslant \frac{q}{2}$, making the closest pair to be at least a distance q apart. ◀

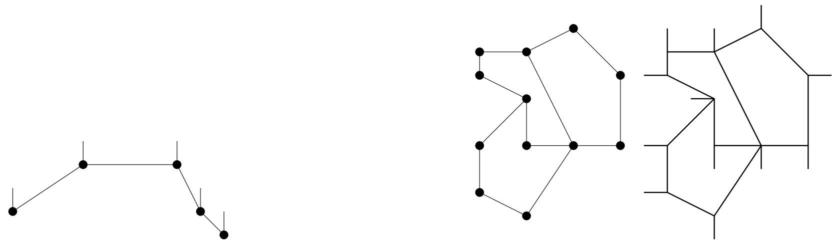
Since $S(q,l)$-DR-1 is NP-hard, the gap ratio problem must also be NP-hard. ◀

4.3 Continuous Case: Path Connected Space

In this section we show that it is NP-hard to find k points in a path connected space such that $R = r$. To prove this, we start by proving that in a path connected space it is NP-Hard to find k points such that $R = r = \frac{3}{2}$ by reducing from the efficient dominating set problem. Later we extend the result for all positive real values of r.

▶ Theorem 14. It is NP-hard to find a set P of k points in a path connected space $\mathcal{M}$ such that $R_P = r_P = \frac{3}{2}$.

Proof. Let us consider an instance of the efficient domination problem, an undirected graph $G(V,E)$, and a parameter k. From this graph we form a metric space $(\mathcal{M}, \delta)$ as follows. In $\mathcal{M}$ each edge of E corresponds to a unit length path. We place at each vertex of V an ϵ-path, where $0 < \epsilon < \frac{1}{4}$, which is merely an ϵ long curve protruding from the vertex as shown in Figure 1(a). The vertices merely become points on a path formed by consecutive edges as shown in Figure 1(b). If there are edge-crossings, we do not consider the crossing to be an intersection but rather consider it as an embedding in $\mathbb{R}^3$. This ensures that different paths only intersect at vertices of the graph (this makes sure that there is direct correspondence between the the path lengths in the graph and the path lengths of the metric space). The distance, δ, between two points in this space is defined by the length of the shortest curve joining the two points.



(a) (b)

Figure 1 (a) ϵ-paths; (b) The graph and the metric space. The open ended lines are the ϵ-paths

We show that finding a set P of k points in $\mathcal{M}$ such that $R_P = r_P = \frac{3}{2}$ is equivalent to finding an efficient dominating set of size k in G.

▶ Claim 15. Suppose $I \subset V$ is an efficient dominating set in G. Then we have a set $P \subset \mathcal{M}$ with $|I| = |P|$ such that $R_P = r_P = \frac{3}{2}$.

This is easy to see using the ideas in the proof of Claim 8.

Conversely, given a set P' of k points in $\mathcal{M}$ such that $R_{P'} = r_{P'} = \frac{3}{2}$, we want to find an efficient dominating set in G. If $P' \subset V$, then we are done as P' is an efficient dominating set in G (refer to Lemma 2). Otherwise, if $P' \not\subset V$, then from P' we construct another set $P \subset V$ such that $R_P = r_P = \frac{3}{2}$. We form P by appropriately moving points of P' to the points corresponding to V. The proof of the following claim is given in Appendix-B.

▶ Claim 16. $P' \subset V$ or $P' \cap V = \emptyset$.

By Claim 16, if $P' \not\subset V$, then $P' \cap V = \emptyset$. Note that in this case P' cannot have midpoints of the graph edges as between any two midpoints at distance 3 from each other, there is a vertex with an ϵ-path which is distance $\frac{3}{2}$ from both points. Thus the other end of this ϵ-path must be distance $\frac{3}{2} + \epsilon$ from both points contradicting the fact that $R_{P'} = \frac{3}{2}$. Thus each point in P' must have a closest vertex. We form the set P by moving each point of P' to its closest vertex. The proof of the following claim is given in Appendix-B.

▶ Claim 17. $R_P = r_P = \frac{3}{2}$.

By Claim 17, without loss of generality, we can assume that the sampled set is a subset of V. Now by Lemma 2 it follows that if we can find a set P of k points in $\mathcal{M}$ such that $R_P = r_P = \frac{3}{2}$, then we can find an efficient dominating set of k vertices in G.

Hence, it is NP-hard to find a set P of k points in a path connected space such that $R_P = r_P = \frac{3}{2}$. ◀

In the above reduction, taking the edge lengths to be $\frac{2x}{3}$ instead of 1 and $\frac{2x\epsilon}{3}$-paths instead of ϵ-paths we have that it is NP-hard to find a set of k points in a path connected space such that $R_P = r_P = \frac{3}{2} \times \frac{2x}{3} = x$. Since this can be done for any positive x, we have the following corollary to Theorem 14.

▶ Corollary 18. It is NP-hard to find a set P of k points in a path connected space such that $R_P = r_P = x$ for all $x > 0$.

By Corollary 18 we have the following result.

▶ Theorem 19. It is NP-hard to find a set of k points in a path connected space such that gap ratio is 1.

5 Approximation hardness

Knowing that finding minimum gap ratio is NP-hard, we show hardness of approximation results in this section.

5.1 General metric space

In this subsection, we show that it is NP-hard to approximate the gap ratio better than a factor of 2. To show the hardness of approximation, we need the concept of independent dominating set. A subset $D \subseteq V$ is called an *independent dominating set* if $|N_G[v] \cap D| \geqslant 1$ for every $v \in V$ and $G[D]$, the induced subgraph of G on D, is an independent set. Given a graph $G = (V, E)$ and a positive integer k, the independent domination problem is to find an independent dominating set of cardinality $\leqslant k$.

▶ Theorem 20. In a general metric space, it is NP-hard to approximate the gap ratio better than a factor of 2.

Proof. To show this hardness, we make a reduction from independent dominating set problem, which is known to be NP-hard [13]. Let $G = (V, E)$ and k be an instance of independent domination problem. We make a weighted complete graph over V such that all edges present in G have weight 1 and all other edges have weight 2. Now the metric space $\mathcal{M}$ is given by the vertex set of the complete graph and the metric is defined by the edge weights.

▶ Claim 21. $G = (V, E)$ has an independent dominating set of cardinality k if and only if there exists a sampled set P in $\mathcal{M}$ of k points with gap ratio 1.

Proof. Let D be an independent dominating set of G of cardinality k. Let the sampled set P in $\mathcal{M}$ is given by $P = D$. Now since D is independent in G, any two points in P are at a distance of 2 in the metric space $\mathcal{M}$. Thus $r = 1$. Since D is a dominating set of G, every point in $\mathcal{M}$ has an edge of weight 1 with some point in P. Thus $R = 1$. Hence, the gap ratio equals 1.

Conversely, suppose a point set P of cardinality k has been selected from $\mathcal{M}$ such that the gap ratio is 1. Note that in the metric space, the value of R is either 1 or 2 and the value of r is either $\frac{1}{2}$ or 1. Consequently, the minimum gap ratio is 1 and the only way that can happen is if R and r both take value 1. Now $R = 1$ means the farthest point in $\mathcal{M}$ is at a distance of 1 from the set P. Thus every point in $\mathcal{M}$ has an edge of weight 1 with some point in P or is in P. Thus P forms a dominating set in G. Also, as $r = 1$, we have the closest pair in P is distance 2 apart. Thus, the set P is an independent dominating set in G of cardinality k. ◀

Suppose, for a contradiction, there is an α-approximation algorithm for the gap ratio problem, where $\alpha < 2$. Since the minimum value of the gap ratio is 1, we can have a set P of cardinality k in $\mathcal{M}$ with a gap ratio $\alpha \times 1 < 2$. Also note that, the possible values of gap ratio are 1, 2, and 4. Hence, P must have gap ratio 1. From the above claim, we have an independent dominating set of cardinality k of the underlying graph G. ◀

5.2 Graph

In Section 4.3, we reduced the problem of finding a set of k points in a graph such that the gap ratio is $\frac{2}{3}$ to the problem of finding a set of k points in a path-connected space such that the gap ratio is 1. We use this hardness of gap ratio being 1 on instances similar to the one created in the reduction to prove $\frac{3}{2}$ approximation hardness on graphs.

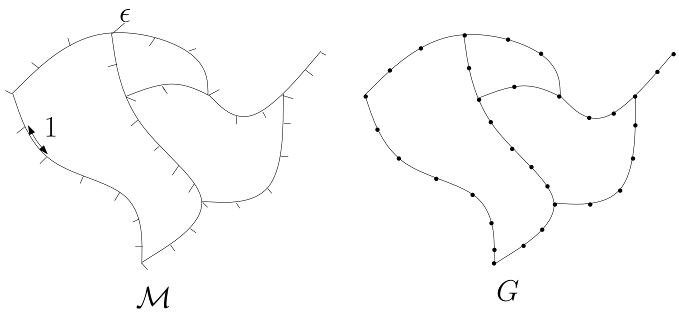


Figure 2 Illustration of the reduction

Although we mentioned gap ratio 1, the specific result that we will be using is Corollary 18. So our starting instance is a space formed by joining integer length curves at their ends (so that points that divide these curves into unit length curves form a connected graph with the unit length curves as edges). Also for some $0 < \epsilon < \frac{1}{4}$ we join curves of length ϵ (at one end) at points such that the integer length curves are divided into unit length curves. Let us call this path connected space $\mathcal{M}$. Note that $\mathcal{M}$ is similar to the path connected space formed in Section 4.3, but, the general shape of the space may vary. The reduction is illustrated in Figure 2. The metric on this space is defined by the length of the shortest path between pairs of points. We form the graph $G = (V, E)$ by putting vertices at the place where the ϵ-length curves are joined to the integer length curves. The ϵ-length protrusions are discarded and the unit length curves between the vertices form the edge set.

▶ Theorem 22. There exists a polynomial time algorithm to find $P \subset \mathcal{M}$ such that $|P| = k$ and $R_P = r_P = \frac{2t+1}{2}$ for some $t \in \{1, 2, ..., \}$ if and only if there exists a polynomial time algorithm to find a set of k vertices in G such that the gap ratio of the set is strictly less than 1.

Proof. Suppose we have a set of k vertices in G with gap ratio strictly less than 1. Let q be the minimum distance of a pair of points in this set. Then by proof of Lemma 2, we have gap ratio is at least $\frac{2}{q}\lfloor\frac{q}{2}\rfloor$ and $r = \frac{q}{2}$. Now unless $R = \lfloor\frac{q}{2}\rfloor$, we have gap ratio greater than 1. If q is even, then the gap ratio will be atleast 1. Hence, q must be odd. Thus the corresponding point set (viz., P) in $\mathcal{M}$ has $R_P = r_P = \frac{q}{2}$.

Conversely suppose we have a set $P \subset \mathcal{M}$ such that $|P| = k$ and $R_P = r_P = \frac{2t+1}{2}$ for some $k \in \{1, 2, ..., \}$. Then using the ideas in Theorem 14 one can verify that the points can be shifted to vertices the graph vertices to get a gap ratio of $\frac{2t}{2t+1} < 1$ in the G. ◀

This gives us that it is NP-hard to find a set with gap ratio less than 1 in graphs, i.e it is NP-hard to find an algorithm which approximates gap ratio within a factor of $\frac{3}{2}$.

Note here that if we could have proven Theorem 22 for $|P| = k$ and $R_P = r_P = \frac{t}{2}$ for some $t \in \{2, 3, ..., \}$, then we wouldn't need to say strictly less than 1 in the statement.

6 Approximation algorithms

6.1 Farthest point insertion

Here we will show that Gonzalez's *farthest point insertion method* [14] for k-centre clustering gives constant approximation factors for gap ratio. The method is outlined in Appendix-C.

Let $(\mathcal{M}, \delta)$ be a metric space of n points. Without loss of generality, let $P = \{p_1, \ldots, p_k\}$ be the set with optimal gap ratio, and let $GR = \alpha$.

▶ Lemma 23. In Algorithm 1 (given in Appendix-C), $R_{S_i} \leqslant R_{S_{i-1}}$ for each $i \in \{2, \ldots, k\}$ and the gap ratio GR_{S_i} is at most 2 at each iteration.

Proof. We prove this by induction on i. Clearly $R_{S_2} \leqslant R_{S_1}$ as that is how q_2 is chosen. Suppose for some t, $R_{S_j} \leqslant R_{S_{j-1}}$ for $j = 2, \ldots, t$. Now by our scheme, q_{t+1} is chosen at a distance R_t from q_t. Thus there exists a point x at a distance $R_{S_{t+1}}$ from q_{t+1}. Hence by definition of R_{S_i}, we have $R_{S_{t+1}} = \delta(x, q_{t+1}) \leqslant \delta(x, S_t) \leqslant R_{S_t}$.

Note that at each insertion, we have chosen q_i at a distance of $R_{S_{i-1}}$ from S_{i-1}. So, $r_{S_i} = \frac{R_{S_{i-1}}}{2}$. Thus, for all $i \in \{2, \ldots, k\}$, we get $GR_{S_i} = R_{S_i}/\frac{R_{S_{i-1}}}{2} = 2 \times \frac{R_{S_i}}{R_{S_{i-1}}} \leqslant 2$. ◀

The main theorem of this section is as follows.

▶ Theorem 24. Farthest point insertion gives the following approximation guarantees: (i) if $\alpha \geqslant 1$, then the approximation ratio is $\frac{2}{\alpha} \leqslant 2$, (ii) if $\frac{2}{3} \leqslant \alpha < 1$, the approximation ratio is $\frac{2}{\alpha} \leqslant 3$, and (iii) if $\alpha < \frac{2}{3}$, the approximation ratio is $\frac{4}{2-\alpha} < 3$.

Proof. Case (i) and (ii) follow directly from Lemma 23. We deal with Case (iii).

Let us define closed balls centred at p_i's as follows: $B_i = \{x \in P : \delta(p_i, x) \leqslant r_P\}$ and $B'_i = \{x \in P : \delta(p_i, x) \leqslant \alpha r_P\}$.

▶ Claim 25. For all $i \in \{2, \ldots, k\}$, $2r_{S_i} \geqslant (2-\alpha)r_P$.

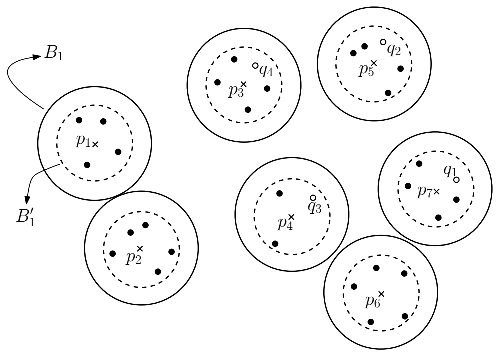


Figure 3 The cross points denote the set P and the hollow points denotes the set S_i

Proof. Note that B'_j 's cover whole of P. The case of $i = 2$ follows from the fact that $\alpha < 2/3$. Assume the result is true for $i \geqslant 2$. We will now show the result is true for S_{i+1} if $i \leqslant k-1$. We claim that q_{i+1} belong to a closed ball B'_j that does not contain any other q_t for $t \leqslant i$. To reach a contradiction, let us assume q_{i+1} falls into a ball B'_j that contains another q_t for some $t \leqslant i$. This would imply $2r_{S_{i+1}} \leqslant \delta(q_t, q_{i+1}) \leqslant 2\alpha r_P$. Note that as $\alpha < 2/3$, we have $2\alpha r_P < (2-\alpha)r_P$. But since, $i \leqslant k-1$, there exists $p_{t'}$ such that $B'_{t'}$ is empty. That implies we could have selected $p_{t'}$ instead of q_{i+1} to get $2r_{S_{i+1}} = \min\{2r_{S_i}, \delta(p_{t'}, S_i)\} \geqslant (2-\alpha)r_P$. Note that last inequality follows from the fact that $2r_{S_i} \geqslant (2-\alpha)r_P$ (by induction) and $\delta(p_{t'}, S_i) \geqslant (2-\alpha)r_P$.

Now that we know q_{i+1} falls into a separate ball B'_j, it is easy to see that $2r_{S_{i+1}} \geqslant \min\{2r_{S_i}, \delta(p_j, S_i)\} \geqslant (2-\alpha)r_P$. ◀

From the proof of Claim 25 we have for all $j \in \{1, \ldots, k\}$, $|B'_j \cap S_k| = 1$. Thus we have $R_{S_k} \leqslant 2\alpha r_P$, since B'_j cover $\mathcal{M}$. Combining this with the fact that $2r_{S_k} \geqslant (2-\alpha)r_P$ (Claim 25), we have $GR_{S_k} \leqslant \frac{4\alpha}{2-\alpha}$ and consequently $\frac{GR_{S_k}}{GR_P} \leqslant \frac{4}{2-\alpha} < 3$. ◀

From the lower bound results in Section 3 we have the following corollaries to Theorem 24.

▶ Corollary 26. The approximation algorithm gives an approximation ratio of 2 when the metric space is continuous, compact and path connected.

▶ Corollary 27. The approximation algorithm gives an approximation ratio of 3 when the metric space is restricted to graph metric space.

▶ Corollary 28. The approximation algorithm gives an approximation ratio of $\rho(k)$, when the metric space is restricted to a unit square in the Euclidean plane, where $\rho(k) = \frac{\sqrt[4]{27}\sqrt{k}}{\sqrt[4]{3}\sqrt{k}-\sqrt{2}} = \frac{2}{\frac{2}{\sqrt{3}}-o(1)}$.

6.2 Doubling algorithm for the online setting

Here we show that the widely used *doubling algorithm* [8] gives a gap ratio of 4 in online setting. In this setting, let $(\mathcal{M}, \delta)$ be a discrete metric space and our goal is to sample a set P of k points such that the gap ratio is minimized. The outline of the algorithm is given in the Appendix-C.

▶ Lemma 29. Algorithm 2 (given in Appendix-C) outputs a set P of k points such that gap ratio is 4.

Proof. Suppose that at each iteration, $2r$ is minimum interpoint distance between the points in P at line 13 of Algorithm 2. Note that, each visited data point is within distance $2r$ from P at line 8 and at line 12 each point of P is within distance $2r$ of P'. Since at line 13 we set $P = P'$, using triangle inequality, we can say that all the data points visited so far are within distance $4r$ from P. Hence, the gap ratio becomes at most $4r/r = 4$. ◀

We can use the ideas of Section 6.1 to get constant approximation factors for different metric spaces for the online case.

7 Conclusion

In this work, we generalize the definition of gap ratio given by Teramoto et al. [23], for general metric spaces. We show non-existence of lower bound for specific metric spaces and it seems that metric spaces that are not connected does not admit a lower bound. On the other side, we show constant lower bounds for gap ratio for connected undirected graphs and metric spaces of unit squares in $\mathbb{R}^2$. We also show that the problem is NP-hard for discrete and continuous metric spaces and design relevant approximation algorithms. Thus we feel that we have been able to make some progress in the direction pointed to by Asano [2]. We would also want to see the effect of random process of point generation on gap ratio in the limiting case.

Acknowledgements

Arijit Bishnu wishes to thank Professor Tetsuo Asano for introducing him to the problem and for helpful discussions. Arijit Ghosh wants to thank Geevarghese Philip for helpful discussions on the problem.

References

1 T. Asano. Computational geometric and combinatorial approaches to digital halftoning. In *CATS*, page 3, 2006.

2 T. Asano. Online uniformity of integer points on a line. *Inf. Process. Lett.*, 109(1):57–60, 2008.

3 T. Asano, N. Katoh, K. Obokata, and T. Tokuyama. Combinatorial and geometric problems related to digital halftoning. In *Theoretical Foundations of Computer Vision*, pages 58–71, 2002.

4 T. Asano, N. Katoh, K. Obokata, and T. Tokuyama. Matrix rounding under the l^p-discrepancy measure and its application to digital halftoning. *SIAM J. Comput.*, 32(6):1423–1435, 2003.

5 T. Asano and S. Teramoto. On-line uniformity of points. In *Book of Abstracts for 8th Hellenic-European Conference on Computer Mathematics and its Applications, Athens, Greece*, pages 21–22, 2007.

6 D. Bange, A. Barkauskas, L. Host, and P. Slater. Generalized domination and efficient domination in graphs. *Discrete Mathematics*, 159(1–3):1 – 11, 1996.

7 M. d. Berg, O. Cheong, M. v. Kreveld, and M. Overmars. *Computational Geometry: Algorithms and Applications.* Springer-Verlag TELOS, Santa Clara, CA, USA, 3rd ed. edition, 2008.

8 N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. E. Schapire, and M. K. Warmuth. How to use expert advice. *J. ACM*, 44(3):427–485, May 1997.

9 Y. Chain-Chin and R. Lee. The weighted perfect domination problem and its variants. *Discrete Applied Mathematics*, 66(2):147 – 160, 1996.

10 B. Chazelle. *The discrepancy method - randomness and complexity.* Cambridge University Press, 2001.

11 C. R. Collins and K. Stephenson. A circle packing algorithm. *Comput. Geom.*, 25(3):233–256, 2003.

12 J. Fiala, J. Kratochvíl, and A. Proskurowski. Systems of distant representatives. *Discrete Applied Mathematics*, 145(2):306–316, 2005.

13 M. R. Garey and D. S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness.* W. H. Freeman & Co., New York, NY, USA, 1990.

14 T. F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theor. Comput. Sci.*, 38:293–306, 1985.

15 J. Kratochvíl. A special planar satisfiability problem and a consequence of its np-completeness. *Discrete Applied Mathematics*, 52(3):233 – 252, 1994.

16 W. Kuperberg. An inequality linking packing and covering densities of plane convex bodies. *Geometriae Dedicata*, 23(1):59–66, 1987.

17 M. Locatelli and U. Raber. Packing equal circles in a square: a deterministic global optimization approach. *Discrete Applied Mathematics*, 122(1–3):139 – 166, 2002.

18 J. Matoušek. *Geometric Discrepancy: An Illustrated Guide.* Algorithms and Combinatorics. Springer, 1999.

19 K. J. Nurmela and P. R. J. Östergård. Packing up to 50 equal circles in a square. *Discrete & Computational Geometry*, 18(1):111–120, 1997.

20 K. J. Nurmela and P. R. J. Östergård. More optimal packings of equal circles in a square. *Discrete & Computational Geometry*, 22(3):439–457, 1999.

21 K. J. Nurmela, P. R. J. Östergård, and R. aus dem Spring. Asymptotic behavior of optimal circle packings in a square. *Canadian Mathematical Bulletin*, 42(3):380–385, 1999.

22 K. Sadakane, N. Takki-Chebihi, and T. Tokuyama. Discrepancy-based digital halftoning: Automatic evaluation and optimization. In *Theoretical Foundations of Computer Vision*, pages 301–319, 2002.

23 S. Teramoto, T. Asano, N. Katoh, and B. Doerr. Inserting points uniformly at every instance. *IEICE Transactions*, 89-D(8):2348–2356, 2006.

24 G. Tóth. New results in the theory of packing and covering. In P. Gruber and J. Wills, editors, *Convexity and Its Applications*, pages 318–359. Birkhäuser Basel, 1983.

25 Y. Zhang, Z. Chang, F. Y. L. Chin, H.-F. Ting, and Y. H. Tsin. Uniformly inserting points on square grid. *Inf. Process. Lett.*, 111(16):773–779, 2011.

APPENDIX-A: Lower Bound Results

Example 1: Given an $\epsilon > 0$, we construct an example of a discrete metric space and a sampled set admitting a gap ratio ϵ.

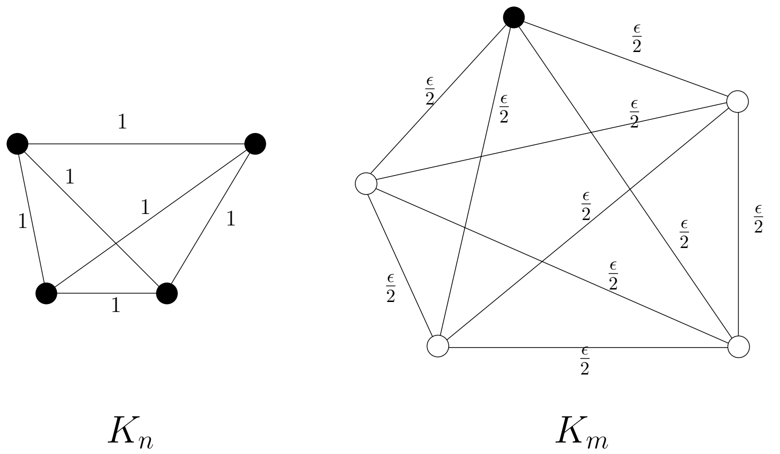


Figure 4 Lower bound for the discrete case. Filled in vertices form the set P.

Consider the complete graph K_n for some $n \in \mathbb{N}$ with each edge having unit weight and the complete graph K_m for some $m \in \mathbb{N}$ with each edge having weight $\frac{\epsilon}{2}$. Let $V[G]$ denote the vertex set of the graph G. Now suppose that the metric space $(\mathcal{M}, \delta)$ is $V[K_n] \cup V[K_m]$ with the metric δ being the edge weights when there are edges between vertices of $V[K_n] \cup V[K_m]$ and ∞, otherwise. Let the sampled set $P = V[K_n] \cup \{v\}$ for some $v \in V[K_m]$. We have $\min_{q \in P} \delta(p, q) = \frac{\epsilon}{2}$ for all $p \in \mathcal{M} \setminus P$. Thus $R = \frac{\epsilon}{2}$. By the definition of P, $r = \frac{1}{2}$. Thus $GR = \frac{R}{r} = \epsilon$.

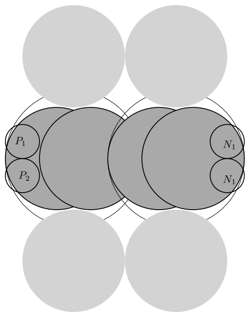
Example 2: Given an $\epsilon > 0$, we construct an example of a discrete metric space and a sampled set admitting a gap ratio ϵ.

Let us consider two balls (say A and B) in $\mathbb{R}^d$ of diameter 1 with distance between their centres being $\frac{2}{\epsilon} + 1$, where $0 < \epsilon < 1$. The metric space $\mathcal{M}$ is defined as $A \cup B$ and the set P is defined by two points, one each in A and B. In this case, the distance between the two points in P must be at least $\frac{2}{\epsilon}$. Hence $r \geqslant \frac{1}{\epsilon}$ and $R \leqslant 1$. Thus the gap ratio becomes less or equal to ϵ.

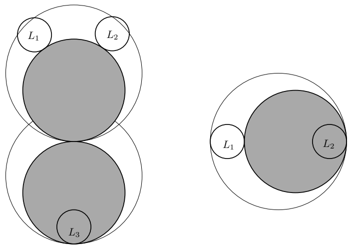
APPENDIX-B: NP-hardness

Proof of Theorem 10 As mentioned in [12], the problem $S(q, l)$-DR is equivalent to considering disks of diameter $q+1$ around the centres of the unit disks and asking whether we can fit one disk of diameter q inside each of the disks of diameter $q+1$ as a representative such that these representatives are pairwise disjoint. For this hardness proof, we use a reduction from planar 3-SAT problem, which is known to be NP-hard [15].

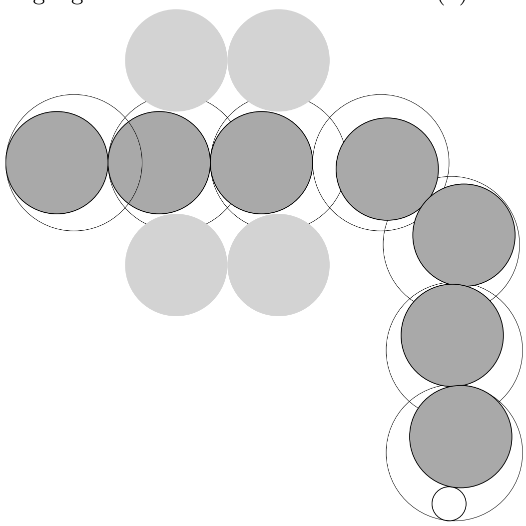
Let Φ be a CNF formula, where each variable has one positive and two negative occurrences and each clause consists of two or three literals. Let G_Φ be the bipartite graph of vertex set $V \cup C$, where V is the variable set and C is the clause set and the edge set is defined by $E = \{xc | x \text{ or } \bar{x} \text{ occurs in clause } c\}$. By definition of planar 3-SAT problem, G_Φ is planar. So, we take a planar embedding of G_Φ and form an instance of $S(q, l)$-DR as follows.



(a) Variable gadget



(b) Clause Gadget



(c) Connector Gadget

Figure 5 The Gadgets

Each variable, clause and edge is replaced by some gadgets as shown in Figure 5. Figure 5(b) denotes the clause gadgets for clause of 2 and 3 literals. In Figure 5 the small disks are auxiliary disks of unit diameter, the lightly shaded disks (diameter q) are part of the reserved area (the same as in [12]), the darkly shaded disks are sample representatives (diameter q) of the big disks (diameter $q+1$). The big disks are placed so that the minimum distance between the centers of any pair is at least l.

The variables are replaced by variable gadgets with three auxiliary disks P_1, N_1 and N_2 such that their centers have points from $\mathbb{Z}^2$ in the direction of the edges towards the gadgets representing the clauses involving this variable (one positive and two negative occurrences). The clause gadgets are placed similarly on the location of replacing the vertices representing clauses. The edges are replaced with connector gadgets so that one of its auxiliary disks is identified by one of the variable gadgets and the other by one of the clause gadget's auxiliary disks. This forms an instance of $S(q,l)$-DR problem.

Suppose, a solution of $S(q,l)$-DR exists. Note that any arbitrary placement of representatives in the clause gadgets must intersect at least one of the auxiliary disks. We interpret this as the clause being satisfied by that particular literal. Note that whenever a solution of $S(q,l)$-DR exists, at least one of the two auxiliary disks in the connector gadgets must intersect one of the representative disks. Also, the representative of the last disk on the other end of the connector associated to the intersected auxiliary disk (L_1, L_2 or L_3) must engulf the auxiliary disk of that disk (due to the reserve area limiting the kind of representatives allowed). Thus if a clause gadget representative has selected a literal with positive occurrence in it then the auxiliary disk P_1 must be engulfed by the representative of a connector gadget. Thus, if a solution of $S(q,l)$-DR exists, a representative of the disks in the variable gadget cannot intersect P_1. Now, we assign a variable $x := true$ if in the corresponding variable gadget the auxiliary disk P_1 is not intersected by the representatives of the variable gadget for x, otherwise we set $x := false$.

Conversely, given a solution of planar 3-SAT instance, we can construct a solution of $S(q,l)$-DR by using the above rule.

Hence, $S(q,l)$-DR is NP-hard. ◀

Proof of Claim 16: In the proof whenever we say the path from x to y, we mean the geodesic path. Suppose that $P' \cap V \neq \emptyset$ and $P' \cap V^c \neq \emptyset$. Let $x, y \in P'$ be the closest pair such that $x \in V$ and $y \notin V$. Then obviously $\delta(x,y) > 3$ as the minimum distance can be 3 and the only points at distance 3 from x must belong to V. Call the first three vertices after x on the path from x to y as x_1, x_2 and x_3. Consider a ball of radius $\frac{3}{2}$ centred at x_2. It must contain some point of P' as $R_{P'} = \frac{3}{2}$. Let us call this point z. Thus $\delta(z, x_2) \leqslant \frac{3}{2}$. Then $\delta(z,y) \leqslant \delta(z,x_2) + \delta(x_2,y) \leqslant \frac{3}{2} + \delta(x_2,y) < 2 + (x_2,y) = \delta(x,y)$. If z is a vertex then this is a clear contradiction, but if it is not a vertex then $\delta(x,z) \leqslant \delta(x,x_2) + \delta(x_2,z) \leqslant 2 + \frac{3}{2} = 3.5$. Thus $\delta(x,y) \leqslant 3.5$ (because y must be closer than z to x) and $\delta(x_2,y) \leqslant \frac{3}{2}$ as shown in Fig. 6. Consider the point at a distance $\frac{3}{2}$ from x on the path from x to y. A ball of radius $\frac{3}{2}$ centred at this point contains x but not y as $\delta(x,y) > 3$. Between this point and x_2 on the path from x to y there must be a point q such that a ball of radius $\frac{3}{2}$ around q contains neither x nor y. Again the ball must contain at least one point $p \in P'$ (see Fig. 6) .

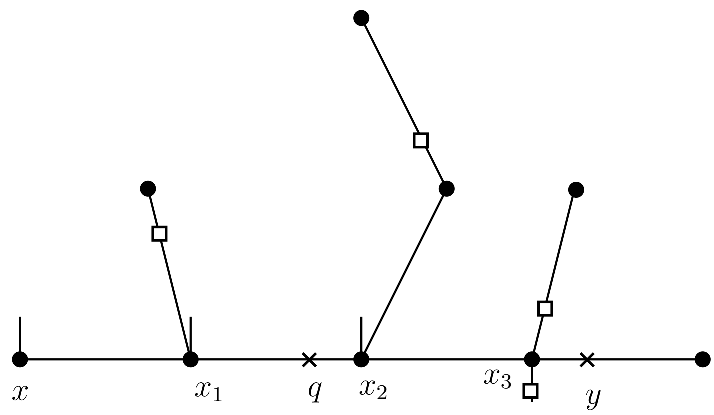


Figure 6 Possible positions of p

- **p is on an edge of x_1:** Then $\delta(x,p) < 3$ which is a contradiction.
- **p is on an edge of a neighbour of x_2 other than x_1 and x_3:** Then we have $\delta(p,q) = \delta(p,x_2) + \delta(x_2,q) \leqslant \frac{3}{2}$ i.e., $\delta(p,x_2) < \frac{3}{2}$. Thus $\delta(p,y) \leqslant \delta(p,x_2) + \delta(x_2,y) < 3$ which is a contradiction.
- **p is on an edge of x_3:** Then $\delta(p,y) < 3$ which is a contradiction.

Thus such a point q cannot exist. Then it means such a pair x and y cannot exist either. Hence we have $P' \subset V$ or $P' \cap V = \emptyset$.

This proves the claim. ◀

Proof of Claim 17: We have $r_{P'} = \frac{3}{2}$.

Suppose $r_P < \frac{3}{2}$. Now note that we are moving our points to the closest vertex to get P. Suppose we obtain the pair of vertices $u, v \in P$ from $x, y \in P'$ such that $\delta(u,v) < 3$

i.e., $\delta(u,v) \leqslant 2$. Then $\delta(u,x) < 0.5$ and $\delta(y,v) < 0.5$. Thus $\delta(x,y) \leqslant \delta(u,x) + \delta(u,v) + \delta(y,v) < 0.5 + 2 + 0.5 = 3$. This is a contradiction.

Suppose $r_P > \frac{3}{2}$. Suppose we obtain the pair of vertices $u, v \in P$ from $x, y \in P'$ such that $\delta(u,v) > 3$ i.e., $\delta(u,v) \geqslant 4$. Then $\delta(u,x) < 0.5$ and $\delta(y,v) < 0.5$.Thus $\delta(u,v) \leqslant \delta(u,x) + \delta(x,y) + \delta(y,v) < 0.5 + 3 + 0.5 = 4$. Again, we have contradiction.

Thus we have $r_P = \frac{3}{2}$.

We must have $R_P \geqslant \frac{3}{2}$ by Lemma 3.

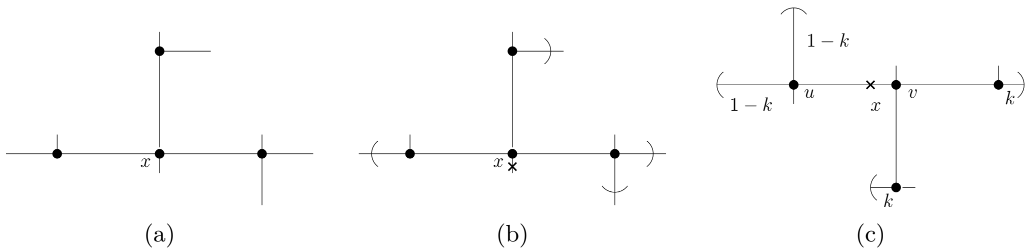


Figure 7 (a) Case 1: x is a vertex, (b) Case 2: x is on an ϵ-path. We have taken the ball from case 1 and the brackets denote the boundary of the ball in this case, (c) Case 3: x is on a full edge

Suppose we have $R_P > \frac{3}{2}$. Then there is a point x such that a ball of radius $\frac{3}{2}$ doesn't contain any point of P. But it must contain a point of P'. Again let us consider cases.

1) **x is a vertex:** Then the ball around x must contain only full edges and edges of length half (see Fig. 7(a)). It also contains a point (say y) of P'. The closest vertex of any such point must be inside this ball. This gives us a contradiction.
2) **x is on an ϵ-path** : This case is similar to the previous case as the ball in this case would clearly be a subset of the ball in the previous case (see Fig. 7(b)).
3) **x is on a full edge** : In this case it is important to note that every point in the ball around x lies on a path (geodesic) that goes through x and lies completely within the ball. Each path is of length 3 or less and x is at the centre of the path (see Fig.7(c)). Let v be the nearest vertex of x and u be the other vertex of the edge on which x lies. Thus all neighbours of v are in the ball.Thus each path of length 3 inside the ball is formed by one edge of length l (where $l = \frac{1}{2} - \delta(x,v)$), two edges of length 1 (including uv) and one edge of length $1-l$ (see Fig. 7). Note that edges of length $1-l$ are incident on u and edges of length l are incident on neighbours v (excluding u).Let us say a point $y \in P'$ lies in this ball (without loss of generality we may assume that y is on a full edge as an ϵ-path cannot intersect with the ball without the corresponding vertex being in the ball). Thus if y lies on one of the edges of length 1 its closest vertex will be u, v or a neighbour of v all of which are in the ball. If y lies on one of the edges of length l then its closest vertex will be a neighbour of v which is in the ball. So assume y lies on an the edge of length $1-l$. Thus if y is within distance $\frac{1}{2}$ of u then the closest vertex for y is u which is also a contradiction. Thus let us assume that y is more than distance $\frac{1}{2}$ of u.
   Then there is a point w between x and v such that a ball of radius $\frac{3}{2}$ has y on its boundary. Again this ball will contain paths of length at most 3. And the paths of length 3 can be characterised by one edge of length l_1 ($l_1 = \frac{1}{2} - \delta(w,v)$), two edges of length 1 (including uv) and one edge of length $1-l_1$. The edges of length $1-l_1$ are subsets of the edges of length $1-l$ (the difference is $\delta(w,x)$). Now if the only point in P' on the boundary of this ball is y then between w and v we must

have a point such that a ball of radius $\frac{3}{2}$ centred around it does not intersect P' at all which is not possible. Hence there must be another point $p \in P'$ at distance of 3 from y such that x, w and v are on the path from y to p (because $\delta(w, p) = \frac{3}{2}$ and as mentioned earlier there are only two such kind of points and if p is on an edge of length $1 - l_1$ then $\delta(y, p) = 2(1 - l_1) < 3$). Then p is on an edge of length l_1 in which case the closest vertex to p is a neighbour of v which was in the ball around x. Thus again we have a contradiction.

This proves the claim. ◀

APPENDIX-C: Approximation

Algorithm 1 Pseudocode of *farthest point insertion*

Input: metric space $(\mathcal{M}, \delta)$ and k;
// $\mathcal{M} = \{p_1, \ldots, p_n\}$
Initialize: q_1 arbitrary point from $\mathcal{M}$ and $S_1 = \{q_1\}$;
for $i = 1$ to $k - 1$ **do**
  $q_{i+1} \leftarrow \mathrm{argmax}_{p_j \in \mathcal{M}}\, \delta(p_j, S_i)$;
  // q_{i+1} is the point farthest from S_i in $\mathcal{M}$
  $S_{i+1} \leftarrow S_i \cup \{q_{i+1}\}$;
end for
Output: S_k and $GR_{S_k} = \frac{R_{S_k}}{r_{S_k}}$;

Algorithm 2 Pseudocode of *doubling algorithm*

Initialize: $P \leftarrow$ {first k data points }
$r \leftarrow$ smallest interpoint distance in P
repeat forever:
while $|P| \leqslant k$ **do**
  get a new point x
  if $\delta(x, P) > 2r$ **then**
    $P \leftarrow P \cup \{x\}$
  end if
end while
$P' \leftarrow \emptyset$
while there exists $y \in P$ such that $\delta(y, P') > 2r$ **do**
  $P' \leftarrow P' \cup \{y\}$
end while
$P \leftarrow P'$
$r \leftarrow 2r$