Semi-algebraic colorings of complete graphs

Jacob Fox [*] János Pach [†] Andrew Suk [‡]

Abstract

In this paper, we consider edge colorings of the complete graph, where the vertices are points in $\mathbb{R}^d$, and each color class E_i is defined by a semi-algebraic relation of constant complexity on the point set. One of our main results is a multicolor regularity lemma: For any $0 < \varepsilon < 1$, the vertex set of any such edge colored complete graph with m colors can be equitably partitioned into at most $(m/\varepsilon)^c$ parts, such that all but at most an ε-fraction of the pairs of parts are monochromatic between them. Here $c > 0$ is a constant that depends on the dimension d and the complexity of the semi-algebraic relations. This generalizes a theorem of Alon, Pach, Pinchasi, Radoičić and Sharir, and Fox, Pach, and Suk. As an application, we prove the following result on generalized Ramsey numbers for such semi-algebraic edge colorings.

For fixed integers p and q with $2 \leq q \leq \binom{p}{2}$, a (p, q)-*coloring* is an edge-coloring of a complete graph in which every p vertices induce at least q distinct colors. The function $f^*(n, p, q)$ is the minimum integer m such that there is a (p, q)-coloring of K_n with at most m colors, where the vertices of K_n are points in $\mathbb{R}^d$, and each color class can be defined as a semi-algebraic relation of constant complexity. Here we show that $f^*(n, p, \lceil \log p \rceil + 1) \geq \Omega\left(n^{\frac{1}{c \log^2 p}}\right)$, and $f^*(n, p, \lceil \log p \rceil) \leq O(\log n)$, thus determining the exact value of q at which $f^*(n, p, q)$ changes from logarithmic to polynomial in n.

We also show the following result on a distinct distances problem of Erdős. Let V be an n-element planar point set such that any p members of V determine at least $\binom{p}{2} - p + 6$ distinct distances. Then V determines at least $n^{\frac{8}{7} - o(1)}$ distinct distances.

1 Introduction

Regularity lemma for multiple semi-algebraic relations. Szemerédi's regularity lemma [35] is one of the most powerful tools in modern combinatorics. It was introduced by Szemerédi in his proof [33] of the Erdős-Turán conjecture on long arithmetic progressions in dense subsets of the integers. The lemma (see [23]) has since become a central tool in extremal combinatorics with many applications in number theory, graph theory, discrete geometry, and theoretical computer science.

In its simplest version [35], the regularity lemma gives a rough structural characterization of all graphs. A partition is called *equitable* if any two parts differ in size by at most one. According to the lemma, for every $\varepsilon > 0$ there is $K = K(\varepsilon)$ such that every graph has an equitable partition of its vertex set into at most K parts such that all but at most an ε fraction of the pairs of parts

[*]Stanford University, Stanford, CA. Supported by a Packard Fellowship, by NSF CAREER award DMS 1352121, and by an Alfred P. Sloan Fellowship. Email: `jacobfox@stanford.edu`.

[†]EPFL, Lausanne and Courant Institute, New York, NY. Supported by Hungarian Science Foundation EuroGIGA Grant OTKA NN 102029, by Swiss National Science Foundation Grants 200020- 144531 and 200021-137574. Email: `pach@cims.nyu.edu`.

[‡]University of Illinois at Chicago, Chicago, IL. Supported by NSF grant DMS-1500153. Email: `suk@uic.edu`.

are ε-regular.[1] The number of parts K grows extremely fast as function of $1/\varepsilon$. It follows from the proof that $K(\varepsilon)$ may be taken to be of an exponential tower of twos of height $\varepsilon^{-O(1)}$. Gowers [19] used a probabilistic construction to show that such an enormous bound is indeed necessary. Consult [5], [27], [15] for other proofs that improve on various aspects of the result.

Alon *et al.* [1] (see also Fox *et al.* [14]) established a strengthening of the regularity lemma for point sets in $\mathbb{R}^d$ equipped with a semi-algebraic relation E. To be more precise, let V be an ordered point set in $\mathbb{R}^d$, and let $E \subset \binom{V}{2}$. We say that E is a *semi-algebraic* relation on V with *complexity* at most t if there are at most t polynomials $g_1, \ldots, g_s \in \mathbb{R}[x_1, \ldots, x_{2d}]$, $s \leq t$, of degree at most t and a Boolean formula Φ such that for vertices $u, v \in V$ such that u comes before v in the ordering,

$$(u, v) \in E \quad \Leftrightarrow \quad \Phi(g_1(u, v) \geq 0; \ldots ; g_s(u, v) \geq 0) = 1.$$

At the evaluation of $g_\ell(u, v)$, we substitute the variables $x_1, \ldots, x_d$ with the coordinates of u, the variables $x_{d+1}, \ldots, x_{2d}$ with the coordinates of v. We may assume that the semi-algebraic relation E is *symmetric*, i.e., for all points $u, v \in \mathbb{R}^d$, $(u, v) \in E$ if and only if $(v, u) \in E$. Indeed, given such an ordered point set $V \subset \mathbb{R}^d$ and a not necessary symmetric semi-algebraic relation E of complexity at most t, we can define $V^* \subset \mathbb{R}^{d+1}$ with points (v, i) where $v \in V$ and v is the ith smallest element in the given ordering of V. Then we can define a symmetric semi-algebraic relation E^* on the pairs of V^* with complexity at most $2t + 2$, by comparing the value of the last coordinates of the two points, and checking the relation E using the first d coordinates of the two points. We will therefore assume throughout this paper that all semi-algebraic relations we consider are symmetric, and the vertices are not ordered. We also assume that the dimension d and complexity t are fixed parameters, and $n = |V|$ tends to infinity.

It was shown in [1] that any point set $V \subset \mathbb{R}^d$ equipped with a semi-algebraic relation $E \subset \binom{V}{2}$ has an equitable partition into a bounded number of parts such that all but at most an ε-fraction of the pairs of parts (V_1, V_2) behave not only regularly, but *homogeneously* in the sense that either $V_1 \times V_2 \subseteq E$ or $V_1 \times V_2 \cap E = \emptyset$. Their proof is essentially qualitative: it gives a poor estimate for the number of parts in such a partition. Fox, Pach, and Suk [17] gave a much stronger quantitative form of this result, showing that the number of parts can be taken to be polynomial in $1/\varepsilon$.

In this paper, we consider points sets equipped with multiple binary relations. Let V be an n-element point set in $\mathbb{R}^d$ equipped with semi-algebraic relations $E_1, \ldots, E_m$ such that $E_1 \cup \cdots \cup E_m = \binom{n}{2}$. Then for any $\varepsilon > 0$, an m-fold repeated application of the result of Fox, Pach, and Suk [17] gives an equitable partition of V into at most $K \leq (1/\varepsilon)^{cm}$ parts such that all but an ε-fraction of the pairs of parts are complete with respect to some relation E_i. In Section 2, we strengthen this result by showing that the number of parts can be taken to be polynomial in m/ε.

Theorem 1.1. *For any positive integers $d, t \geq 1$ there exists a constant $c = c(d, t) > 0$ with the following property. Let $0 < \varepsilon < 1/2$ and let V be an n-element point set in $\mathbb{R}^d$ equipped with semi-algebraic relations $E_1, \ldots, E_m$ such that E_i has complexity at most t and $\binom{V}{2} = E_1 \cup \cdots \cup E_m$. Then V has an equitable partition $V = V_1 \cup \cdots \cup V_K$ into at most $K \leq (m/\varepsilon)^c$ parts such that all but an ε-fraction of the pairs of parts are complete with respect to some relation E_i.*

Ramsey and generalized Ramsey numbers for semi-algebraic relations. In Section 3, we give an application of Theorem 1.1 to semi-algebraic multicolor Ramsey numbers. The Ramsey

[1] For a pair (V_i, V_j) of vertex subsets, $e(V_i, V_j)$ denotes the number of edges in the graph running between V_i and V_j. The density $d(V_i, V_j)$ is defined as $\frac{e(V_i, V_j)}{|V_i||V_j|}$. The pair (V_i, V_j) is called ε-regular if for all $V_i' \subset V_i$ and $V_j' \subset V_j$ with $|V_i'| \geq \varepsilon |V_i|$ and $|V_j'| \geq \varepsilon |V_j|$, we have $|d(V_i', V_j') - d(V_i, V_j)| \leq \varepsilon$.

number $R(p;m)$ is the smallest integer n such that any m-coloring on the edges of the complete n-vertex graph contains a monochromatic copy of K_p. We define $R_{d,t}(p;m)$ analogously, where we restrict at our attention to graphs with vertices as points in $\mathbb{R}^d$ and the colorings are given by semi-algebraic binary relations of complexity at most t. Here each edge receives at least one color. Clearly, $R_{d,t}(p;m) \leq R(p;m)$.

For the case $p = 3$, in establishing a famous theorem in additive number theory, Schur [29] proved that

$$\Omega(2^m) \leq R(3;m) \leq O(m!). \tag{1}$$

While the upper bound has remained unchanged over the last 100 years, the lower bound was successively improved and the current record is $R(3;m) \geq \Omega(3.199^m)$ due to Xiaodong et al. [39].

In [32], it was shown that for fixed $d,t \geq 1$, $R_{d,t}(3;m) = 2^{O(m \log\log m)}$, improving the upper bound in (1) in the semi-algebraic setting. Our next theorem extends this result to $R_{d,t}(p;m)$, for any $p \geq 3$.

Theorem 1.2. *For fixed integers $d,t \geq 1$ and $p \geq 3$, we have*

$$R_{d,t}(p;m) = 2^{O(pm \log\log m)}.$$

Here and throughout the rest of the paper, all logarithms are in base 2.

The following natural generalization of the Ramsey function was first introduced by Erdős and Shelah (see [8]), and studied by Erdős and Gyárfás in [10].

Definition 1.3. For integers p and q with $2 \leq q \leq \binom{p}{2}$, a (p,q)-*coloring* is an edge-coloring of a complete graph in which every p vertices induce at least q distinct colors.

Let $f(n,p,q)$ be the minimum integer m such that there is a (p,q)-coloring of K_n with at most m colors. When p and q are considered fixed integers and n tends to infinity, estimating $f(n,p,q)$ can be a very challenging problem. Estimating $f(n,p,2)$ is equivalent to estimating $R(p;m)$ since $f(n,p,2)$ is the inverse of $R(p;m)$. In particular,

$$\Omega\left(\frac{\log n}{\log\log n}\right) \leq f(n,3,2) \leq O(\log n). \tag{2}$$

Erdős and Gyárfás [10] showed that

$$\Omega\left(n^{\frac{1}{p-2}}\right) \leq f(n,p,p) \leq O\left(n^{\frac{2}{p-1}}\right).$$

Surprisingly, estimating $f(n,p,p-1)$ is much more difficult. In [10], Erdős and Gyárfás asked for p fixed if $f(n,p,p-1) = n^{o(1)}$. The trivial lower bound is $f(n,p,p-1) \geq f(n,p,2) \geq \Omega\left(\frac{\log n}{\log\log\log n}\right)$. For $p = 4$, this was improved using the probabilistic technique known as dependent random choice by Kostochka and Mubayi [24] and further by Fox and Sudakov [18] to $f(n,4,3) \geq \Omega(\log n)$. This was later extended in Conlon et al. [6] to $f(n,p,p-1) \geq \Omega(\log n)$. The Erdős-Gyárfás question was answered affirmatively by Mubayi [28] for $p = 4$ who found an elegant construction which implies $f(n,4,3) \leq e^{O(\sqrt{\log n})}$. Conlon et al. [6] later answered the Erdős-Gyárfás question for all p, showing for $p \geq 4$ fixed that $f(n,p,p-1) \leq e^{(\log n)^{1-1/(p-2)+o(1)}}$. While it is now known that $f(n,p,p-1)$ does not grow as a power in n, there remains a very large gap between the upper and lower bounds on $f(n,p,p-1)$.

Here we study a variant of the function $f(n,p,q)$ for point sets $V \subset \mathbb{R}^d$ equipped with semi-algebraic relations. More precisely, we define the function $f_{d,t}(n,p,q)$ to be the minimum m such that there is a (p,q)-coloring of K_n with m colors, whose vertices can be chosen as points in $\mathbb{R}^d$, and each color class can defined by a semi-algebraic relation on the point set with complexity at most t. Unlike in the definition of $R_{d,t}(p;m)$, here we require that each edge receives *exactly* one color.

Clearly we have $f(n,p,q) \leq f_{d,t}(n,p,q)$. The following result shows the exact moment when $f_{d,t}(n,p,q)$ changes between a power of n to logarithmic in n.

Theorem 1.4. *For fixed integers $d,t \geq 1$, there is a $c = c(d,t) > 0$ such that for $p \geq 3$, we have*

$$f_{d,t}(n,p,\lceil \log p \rceil + 1) \geq \Omega\left(n^{\frac{1}{c\log^2 p}}\right).$$

Moreover, for $t \geq 4$,

$$f_{d,t}(n,p,\lceil \log p \rceil) \leq O(\log n).$$

The proof of Theorem 1.4 is given in Section 3.

Distinct distances. In his classic 1946 paper [7], Erdős asked what is the minimum number of distinct distances determined by an n-element planar point set V. He showed that a $\sqrt{n} \times \sqrt{n}$ integer lattice determines $\Theta(n/\sqrt{\log n})$ distinct distances, and conjectured that any n-element point set determines at least $n^{1-o(1)}$ distinct distances. Several authors established lower bounds for this problem, and Guth and Katz [20] answered Erdős' question by proving that any n-element planar point set determines at least $\Omega(n/\log n)$ distinct distances.

In [10], Erdős and Gyárfás studied the following generalization. For integers p and q with $p \leq q \leq \binom{p}{2}$, let $D(n,p,q)$ denote the minimum number of distinct distances determined by a planar n-element point set V with the property that any p points from V determines at at least q distinct distances. Notice by definition

$$D(n,p,q) \geq f_{d,t}(n,p,q) \geq f(n,p,q),$$

and by [20] we always have $D(n,p,q) \geq \Omega(n/\log n)$. A simple argument shows that $D(n,4,5) \geq \Omega(n)$, and clearly $D(n,4,5) \leq O(n^2)$. It is an old question of Erdős (see [9] and [3]) to determine if $D(n,4,5)$ is superlinear or $o(n^2)$. Likewise, the best known lower and upper bounds for $D(n,5,9)$ are of the form $\Omega(n)$ and $O(n^2)$.

Observe that, for every fixed integer $p > 2$, we have $D\left(n,p,\binom{p}{2} - \lfloor p/2 \rfloor + 2\right) \geq \Omega(n^2)$, since any distance can occur at most $\lfloor p/2 \rfloor - 1$ times in this setting. In [10], Erdős and Gyárfás showed that for $p \geq 7$

$$D\left(n,p,\binom{p}{2} - \lfloor p/2 \rfloor + 1\right) \geq f\left(n,p,\binom{p}{2} - \lfloor p/2 \rfloor + 1\right) \geq \Omega\left(n^{4/3}\right).$$

Sárkőzy and Selkow [31] showed that

$$D\left(n,p,\binom{p}{2} - p + \lceil \log p \rceil + 4\right) \geq f\left(n,p,\binom{p}{2} - p + \lceil \log p \rceil + 4\right) \geq \Omega\left(n^{1+\epsilon}\right),$$

where $\epsilon = \epsilon(p)$. Our next result establishes a better bound for this distinct distance problem.

Theorem 1.5. *For fixed $p \geq 6$, we have*

$$D\left(n, p, \binom{p}{2} - p + 6\right) \geq n^{\frac{8}{7} - o(1)}.$$

The proof of Theorem 1.5 appears in Section 5.

Ramsey-Turán numbers for semi-algebraic relations. The classical theorem of Turán gives the maximum number of edges in a K_p-free graph on n vertices.

Theorem 1.6 (Turán, [36]). *Let $G = (V, E)$ be a K_p-free graph with n vertices. Then*

$$|E| \leq \frac{1}{2}\left(1 - \frac{1}{p-1} + o(1)\right) n^2.$$

The only graph for which this bound is tight is the complete $(p-1)$-partite graph whose parts are of size as equal as possible. This graph can easily be realized as an intersection graph of segments in the plane, which is a semi-algebraic graph with complexity at most four. Therefore, restricting the problem to semi-algebraic graphs does not change Turán's theorem.

Let H be a fixed graph. The Ramsey-Turán number $\mathbf{RT}(n, H, \alpha)$ is the maximum number of edges an n-vertex graph G can have without containing H as a subgraph and without having an independent set of size α. Ramsey-Turán numbers were introduced by Sós and were motivated by the classical theorems of Ramsey and Turán and their connections to geometry, analysis, and number theory. One of the earliest results in Ramsey-Turán Theory a ppeared in [12], which says that for $p \geq 2$

$$\mathbf{RT}(n, K_{2p-1}, o(n)) = \frac{1}{2}\left(1 - \frac{1}{p-1}\right) n^2 + o(n^2).$$

For the even cases, a celebrated result of Szemerédi from 1972 [34] states that

$$\mathbf{RT}(n, K_4, o(n)) = \frac{1}{8} n^2 + o(n^2),$$

and a more general result of Erdős, Hajnal, Sós and Szemerédi [11] states that

$$\mathbf{RT}(n, K_{2p}, o(n)) = \frac{1}{2}\left(\frac{3p-5}{3p-2}\right) n^2 + o(n^2).$$

For more results in Ramsey-Turán theory, see a nice survey by Simonovits and Sós [30].

Here we give tight bounds on Ramsey-Turán numbers for *semi-algebraic graphs*, showing that the extremal densities for even cliques are different in the semi-algebraic setting than in the classical setting, and the densities for K_{2p} are the same as for K_{2p-1}. Let $\mathbf{RT}_{d,t}(n, K_p, o(n))$ be the maximum number of edges a graph $G = (V, E)$ can have, where V is a set of n points in $\mathbb{R}^d$ and $E \subset \binom{V}{2}$ is a semi-algebraic relation with complexity at most t, without containing K_p as a subgraph and also without having an independent set of size $o(n)$.

Theorem 1.7. *For fixed integers $d \geq 4$, $t \geq 10$ and $p \geq 2$, we have*

$$\mathbf{RT}_{d,t}(n, K_{2p-1}, o(n)) = \mathbf{RT}_{d,t}(n, K_{2p}, o(n)) = \frac{1}{2}\left(1 - \frac{1}{p-1}\right) n^2 + o(n^2).$$

In contrast to Turán's problem, the answer to the Ramsey-Turán question for K_{2p}-free graphs changes if we require the graph to be semi-algebraic with bounded complexity. We note that we do not need the full strength of Theorem 1.1 to establish Theorem 1.7, and one could use the regularity lemma established by Alon et al. [1]. The proof of Theorem 1.7 is given in Section 6.

We sometimes omit floor and ceiling signs whenever they are not crucial for the sake of clarity of presentation.

2 Multicolor semi-algebraic regularity lemma

In this section, we prove Theorem 1.1. A set $\Delta \subset \mathbb{R}^d$ is *semi-algebraic* if there are polynomials $g_1, \ldots, g_t$ and a boolean formula Φ such that

$$A = \{x \in \mathbb{R}^d : \Phi(g_1(x) \geq 0; \ldots; g_t(x) \geq 0) = 1\}.$$

We say that a semi-algebraic set in d-space has *description complexity* at most κ if the number of inequalities is at most κ, and each polynomial g_i has degree at most κ. Let $\sigma \subset \mathbb{R}^d$ be a *surface* in $\mathbb{R}^d$, that is, σ is the zero set of some polynomial $h \in \mathbb{R}[x_1, \ldots, x_d]$. The *degree* of a surface $\sigma = \{x \in \mathbb{R}^d : h(x) = 0\}$ is the degree of the polynomial h. We say that the surface $\sigma \subset \mathbb{R}^d$ *crosses* a semi-algebraic set Δ if $\sigma \cap \Delta \neq \emptyset$ and $\Delta \not\subset \sigma$. We now recall an old theorem known as the cutting lemma.

Lemma 2.1 (Chazelle et al. [4], Koltun [22])**.** *For $d \geq 1$, let Σ be a multiset of surfaces in $\mathbb{R}^d$, each of which has degree at most t, and let r be an integer parameter such that $1 \leq r \leq N$. Then there is a subdivision of $\mathbb{R}^d$ into at most $c_1 r^{2d}$ semi-algebraic sets Δ_i such that $c_1 = c_1(d,t)$, each Δ_i has complexity at most c_1, and at most $|\Sigma|/r$ surfaces from Σ cross Δ_i. In the case that $d \geq 4$, the bound on the number of semi-algebraic sets Δ_i in the subdivision of $\mathbb{R}^d$ can be improved to $c_1 r^{2d-4+\varepsilon}$.*

We note that we will use the weaker bound of $c_1 r^{2d}$ in the lemma above for this section, and use the stronger bound $c_1 r^{2d-4+\varepsilon}$ in Section 1.5.

We will first prove the following variant of Theorem 1.1, from which Theorem 1.1 quickly follows.

Theorem 2.2. *For any $\varepsilon > 0$, every n-element point set $V \subset \mathbb{R}^d$ equipped with semi-algebraic binary relations $E_1, \ldots, E_m \subset \binom{V}{2}$ with $\binom{V}{2} = E_1 \cup \cdots \cup E_m$ and each E_i has complexity at most t, can be partitioned into $K \leq c_2(\frac{m}{\varepsilon})^{5d^2}$ parts $V = V_1 \cup \cdots \cup V_K$, where $c_2 = c_2(d,t)$, such that*

$$\sum \frac{|V_i||V_j|}{n^2} \leq \varepsilon,$$

where the sum is taken over all pairs (i,j) such that (V_i, V_j) is not complete with respect to E_ℓ for all $\ell = 1, \ldots, m$.

Proof. For each relation E_i, let $g_{i,1}, \ldots, g_{i,t} \in \mathbb{R}[x_1, \ldots, x_{2d}]$ be polynomials of degree t, and let Φ_i be a boolean formula such that

$$(u,v) \in E_i \quad \Leftrightarrow \quad \Phi_i(g_{i,1}(u,v) \geq 0; \ldots; g_{i,t}(u,v) \geq 0) = 1.$$

For each point $x \in \mathbb{R}^d$, $i \in \{1, \ldots, m\}$, and $j \in \{1, \ldots, t\}$, we define the surface

$$\sigma_{i,j}(x) = \{y \in \mathbb{R}^d : g_{i,j}(x,y) = 0\}.$$

Let Σ be the family of tmn surfaces in $\mathbb{R}^d$ defined by

$$\Sigma = \{\sigma_{i,j}(u) : u \in V, 1 \leq i \leq m, 1 \leq j \leq t\}.$$

We apply Lemma 2.1 to Σ with parameter $r = tm/\varepsilon$, and subdivide $\mathbb{R}^d$ into $s \leq c_1 \left(\frac{tm}{\varepsilon}\right)^{2d}$ semi-algebraic sets Δ_ℓ, with each Δ_ℓ having complexity at most c_1, where c_1 is defined in Lemma 2.1. Moreover, at most $tmn/r = \varepsilon n$ surfaces from Σ cross Δ_ℓ for every ℓ. This implies that at most εn points in V give rise to at least one surface in Σ that cross Δ_ℓ.

Let $U_\ell = V \cap \Delta_\ell$ for each $\ell \leq s$. We now partition Δ_ℓ as follows. For $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, s\}$, define $\Delta_{\ell,i,j} \subset \mathbb{R}^d$ by

$$\Delta_{\ell,i,j} = \{x \in \Delta_\ell : \sigma_{i,1}(x) \cup \cdots \cup \sigma_{i,t}(x) \text{ crosses } \Delta_j\}.$$

Observation 2.3. *For any i, j, and ℓ, the semi-algebraic set $\Delta_{\ell,i,j}$ has complexity at most $c_3 = c_3(d,t)$.*

Proof. Set $\sigma_i(x) = \sigma_{i,1}(x) \cup \cdots \cup \sigma_{i,t}(x)$, which is a semi-algebraic set with complexity at most $c_4 = c_4(d,t)$. Then

$$\Delta_{\ell,i,j} = \left\{ x \in \Delta_\ell : \begin{array}{l} \exists y_1 \in \mathbb{R}^d \text{ s.t. } y_1 \in \sigma_i(x) \cap \Delta_j, \text{ and} \\ \exists y_2 \in \mathbb{R}^d \text{ s.t. } y_2 \in \Delta_j \setminus \sigma_i(x). \end{array} \right\}.$$

We can apply quantifier elimination (see Theorem 2.74 in [2]) to make $\Delta_{\ell,i,j}$ quantifier-free, with description complexity at most $c_3 = c_3(d,t)$.

□

Set $\mathcal{F}_\ell = \{\Delta_{\ell,i,j} : 1 \leq i \leq m, 1 \leq j \leq s\}$. We partition the points in U_ℓ into equivalence classes, where two points $u, v \in U_\ell$ are equivalent if and only if u belongs to the same members of $\mathcal{F}_\ell$ as v does. Since $\mathcal{F}_\ell$ gives rise to at most $c_3|\mathcal{F}_\ell|$ polynomials of degree at most c_3, by the Milnor-Thom theorem (see [26] Chapter 6), the number of distinct sign patterns of these $c_3|\mathcal{F}_\ell|$ polynomials is at most $(50c_3(c_3|\mathcal{F}_\ell|))^d$. Hence, there is a constant $c_5 = c_5(d,t)$ such that U_ℓ is partitioned into at most $c_5(ms)^d$ equivalence classes. After repeating this procedure to each U_ℓ, we obtain a partition of our point set $V = V_1 \cup \cdots \cup V_K$ with

$$K \leq sc_5(ms)^d = c_5 m^d s^{d+1} \leq c_5 t^{2d(d+1)} c_1^{d+1} \left(\frac{m}{\varepsilon}\right)^{5d^2} = c_2 \left(\frac{m}{\varepsilon}\right)^{5d^2},$$

where we define $c_2 = c_5 t^{2d(d+1)} c_1^{d+1}$.

For fixed i, consider the part V_i. Then there is a semi-algebraic set Δ_{ℓ_i} obtained from Lemma 2.1 such that $U_{\ell_i} = V \cap \Delta_{\ell_i}$ and $V_i \subset U_{\ell_i} \subset \Delta_{\ell_i}$. Now consider all other parts V_j such that not all of their elements are related to every element of V_i with respect to any relation E_k where $1 \leq k \leq m$. Then each point $u \in V_j$ gives rise to a surface in Σ that crosses Δ_{ℓ_i}. By Lemma 2.1, the total number of such points in V is at most εn. Therefore, we have

$$\sum_j |V_i||V_j| = |V_i| \sum_j |V_j| \leq |V_i| \varepsilon n,$$

where the sum is over all j such that $V_i \times V_j$ is not contained in the relation E_ℓ for any ℓ. Summing over all i, we have

$$\sum_{i,j} |V_i||V_j| \le \varepsilon n^2,$$

where the sum is taken over all pairs i, j such that (V_i, V_j) is not complete with respect to E_ℓ for all ℓ.

$\square$

Proof of Theorem 1.1. Apply Theorem 2.2 with approximation parameter $\varepsilon/2$. So there is a partition $\mathcal{Q} : V = U_1 \cup \cdots \cup U_{K'}$ into $K' \le (m/\varepsilon)^c$ parts with $c = c(d,t)$ and $\sum |U_i||U_j| \le (\varepsilon/2)|V|^2$, where the sum is taken over all pairs (i,j) such that (U_i, U_j) is not complete with respect to E_ℓ for all ℓ.

Let $K = 8\varepsilon^{-1}K'$. Partition each part U_i into parts of size $|V|/K$ and possibly one additional part of size less than $|V|/K$. Collect these additional parts and divide them into parts of size $|V|/K$ to obtain an equitable partition $\mathcal{P} : V = V_1 \cup \cdots \cup V_K$ into K parts. The number of vertices of V which are in parts V_i that are not contained in a part of $\mathcal{Q}$ is at most $K'|V|/K$. Hence, the fraction of pairs $V_i \times V_j$ with not all V_i, V_j are subsets of parts of $\mathcal{Q}$ is at most $2K'/K = \varepsilon/4$. As $\varepsilon/2 + \varepsilon/4 < \varepsilon$, we obtain that less than an ε-fraction of the pairs of parts of $\mathcal{P}$ are not complete with respect to any relation $E_1, \ldots, E_m$. $\square$

3 Generalized Ramsey numbers for semi-algebraic colorings

The goal of this section is to prove Theorem 1.4. Let V be a set of points in $\mathbb{R}^d$ equipped with semi-algebraic relations $E_1, \ldots, E_m$ such that each E_i has complexity at most t, $\binom{V}{2} = E_1 \cup \cdots \cup E_m$, and $E_i \cap E_j = \emptyset$ for all $i \neq j$. Let $S_1, S_2 \subset V$ be k-element subsets of V. We say that S_1 and S_2 are *isomorphic*, denoted by $S_1 \simeq S_2$, if there is a bijective function $h : S_1 \to S_2$ such that for $u, v \in S_1$ we have $(u,v) \in E_i$ if and only if $(h(u), h(v)) \in E_i$.

Let $S \subset V$ be such that $|S| = 2^s$ for some positive integer s. We say that S is *s-layered* if $s = 1$ or if there is a partition $S = S_1 \cup S_2$ such that $|S_1| = |S_2| = 2^{s-1}$, S_1 and S_2 are $(s-1)$-layered, $S_1 \simeq S_2$, and for all $u \in S_1$ and $v \in S_2$ we have $(u,v) \in E_i$ for some fixed i. Notice that given an s-layered set S, there are at most s relations $E_{i_1}, \ldots, E_{i_s}$ such that $\binom{S}{2} \subset E_{i_1} \cup \cdots \cup E_{i_s}$. Hence, the lower bound in Theorem 1.4 is a direct consequence of the following result.

Theorem 3.1. *Let $s \ge 1$ and let V be an n-element point set in $\mathbb{R}^d$ equipped with semi-algebraic relations $E_1, \ldots, E_m$ such that each E_i has complexity at most t, $E_1 \cup \cdots \cup E_m = \binom{V}{2}$, and $E_i \cap E_j = \emptyset$ for all $i \neq j$. If $m \le n^{\frac{1}{cs^2}}$, then there is a subset $S \subset V$ such that $|S| = 2^s$ and S is s-layered, where $c = c(d,t)$.*

Proof. We proceed by induction on s. The base case $s = 1$ is trivial. For the inductive step, assume that the statement holds for $s' < s$. We will specify $c = c(d,t)$ later. We start by applying Theorem 1.1 with parameter $\varepsilon = \frac{1}{m^s}$ to the point set V, which is equipped with semi-algebraic relations $E_1, \ldots, E_m$, and obtain an equitable partition $\mathcal{P} : V = V_1 \cup \cdots \cup V_K$, where

$$K \le c_2 \left(\frac{m}{\varepsilon}\right)^{5d^2} \le c_2 m^{10sd^2},$$

and $c_2 = c_2(d,t)$. Since all but an ε fraction of the pairs of parts $\mathcal{P}$ are complete with respect to E_ℓ for some ℓ, by Turán's theorem (see Theorem 1.6 in Section 1), there are $m^{s-1} + 1$ parts $V_i' \in P$

such that each pair $(V_i', V_j') \in \mathcal{P} \times \mathcal{P}$ is complete with respect to some relation E_ℓ. Since $\mathcal{P}$ is an equitable partition, we have $|V_i'| \geq \frac{n}{c_2 m^{10d^2 s}}$. By picking $c = c(d,t)$ sufficiently large, we have

$$|V_i'|^{\frac{1}{c(s-1)^2}} \geq \left(\frac{n}{c_2 m^{10d^2 s}}\right)^{\frac{1}{c(s-1)^2}} \geq m^{\frac{cs^2 - 10c_2 d^2 s}{c(s-1)^2}} \geq m.$$

By the induction hypothesis, each V_i' contains an $(s-1)$-layered set S_i for $i \in \{1, \ldots, m^{s-1}+1\}$. By the pigeonhole principle, there are two $(s-1)$-layered sets S_i, S_j such that $S_i \simeq S_j$. Since $S_i \times S_j \subset E_\ell$ for some ℓ, the set $S = S_i \cup S_j$ is an s-layered set. This completes the proof.

$\square$

To prove the upper bound for $f_{d,t}(n, p, \lceil \log p \rceil)$, when $d \geq 1$ and $t \geq 100$, it is sufficient to construct a 2^m-element point set $V \subset \mathbb{R}$ equipped with semi-algebraic relations $E_1, \ldots, E_m$ that is m-layered. More precisely, for each integer $m \geq 1$, we construct a set V_m of 2^m points in $\mathbb{R}$ equipped with semi-algebraic relations $E_1, \ldots, E_m$ such that

1. V_m with respect to relations $E_1, \ldots, E_m$ is m-layered,
2. $E_1 \cup \cdots \cup E_m = \binom{V_m}{2}$ is a partition,
3. each E_i has complexity at most four, and
4. each E_i is *shift invariant*, that is $(u, v) \in E_i$ if and only if $(u + c, y + c) \in E_i$ for $c \in \mathbb{R}$.

We start by setting $V_1 = \{1, 2\}$ and defining $E_1 = \{u, v \in V_1 : |u - v| = 1\}$. Having defined the point set V_i and relations $E_1, \ldots, E_i$, we define V_{i+1} and E_{i+1} as follows. Let $C = C(i)$ be a sufficiently large integer such that $C > 10 \max_{u \in V_i} u$. Then we have $V_{i+1} = V_i \cup (V_i + C)$, where $V_i + C$ is a translated copy of V_i. We now define the relation E_{i+1} by

$$(u, v) \in E_{i+1} \quad \Leftrightarrow \quad C/2 < |u - v| < 2C.$$

Hence, V_{i+1} with respect to relations $E_1, \ldots, E_{i+1}$ satisfies the properties stated above and is clearly $(i+1)$-layered. One can easily check that any set of p points in V_m induces at least $\lceil \log p \rceil$ distinct relations (colors).

4 Ramsey numbers with respect to semi-algebraic colorings

In this section, we prove Theorem 1.2. Consider all n-element point set V in $\mathbb{R}^d$ equipped with symmetric semi-algebraic relations $E_1, \ldots, E_m \subset \binom{V}{2}$, such that each E_i has complexity at most t and $\binom{V}{2} = E_1 \cup \cdots \cup E_m$. Note that the binary relations E_i are not necessarily disjoint. We define the Ramsey function $R_{d,t}(p_1, \ldots, p_m)$ to be the minimum integer n such that for every n-element point set with these properties, there exists an i such that we can find p_i points with every pair induced by them belonging to E_i. In particular, if $p_1 = p_2 = \cdots = p_m = p$, then $R_{d,t}(p_1, \ldots, p_m) = R_{d,t}(p; m)$. The following result thus implies Theorem 1.2.

Theorem 4.1. *For fixed integers $d, t \geq 1$, and $p_i \geq 3$ for $1 \leq i \leq m$, we have*

$$R_{d,t}(p_1, \ldots, p_m) \leq 2^{c\left(\sum_{i=1}^m p_i\right) \log\log(m+2)},$$

where $c = c(d,t)$.

Proof. Let $c = c(d,t)$ be a sufficiently large constant that will be determined later. We proceed by induction on m and $s = \sum_{i=1}^m p_i$. The base case when $m = 1$ is trivial. Now assume $m \geq 2$, and the statement holds for $m' < m$. If $s < 3m$, then there is an i such that $p_i \leq 2$. We can therefore reduce the problem to $m-1$ colors, and the statement follows by induction. Hence we can assume $s \geq 3m$, and the statement holds for $s' < s$.

Set $n = 2^{c\left(\sum_{i=1}^m p_i\right)\log\log(m+2)}$ and let $E_1, \ldots, E_m \subset \binom{P}{2}$ be semi-algebraic relations on P such that $\binom{V}{2} = E_1 \cup \cdots \cup E_m$, and each E_i has complexity at most t. Note that the relations E_i may not be disjoint. For the sake of contradiction, suppose that for every $i \in \{1, \ldots, m\}$, V does not contain p_i points such that every pair of distinct points induced by them is in E_i.

For each relation E_i, there are t polynomials $g_{i,1}, \ldots, g_{i,t}$ of degree at most t, and a Boolean function Φ_i such that

$$(u,v) \in E_i \quad \Leftrightarrow \quad \Phi_i(g_{i,1}(u,v) \geq 0, \ldots, g_{i,t}(u,v) \geq 0) = 1.$$

For $1 \leq i \leq m, 1 \leq j \leq t, v \in V$, we define the surface $\sigma_{i,j}(v) = \{x \in \mathbb{R}^d : g_{i,j}(v,x) = 0\}$, and let

$$\Sigma = \{\sigma_{i,j}(v) : 1 \leq i \leq m, 1 \leq j \leq t, v \in V\}.$$

Hence, $|\Sigma| = mtn$. We apply Lemma 2.1 to Σ with parameter $r = 2tm$, and decompose $\mathbb{R}^d$ into $c_1(2tm)^{2d}$ regions Δ_i, where $c_1 = c_1(t,d)$ is defined in Lemma 2.1, such that each region Δ_i is crossed by at most $tmn/r = n/2$ members in Σ. By the pigeonhole principle, there is a region $\Delta \subset \mathbb{R}^d$, such that $|\Delta \cap V| \geq \frac{n}{c_1(2tm)^{2d}}$, and at most $n/2$ members in Σ crosses Δ. Let V_1 be a set of exactly $\left\lfloor \frac{n}{c_1(2tm)^{2d}} \right\rfloor$ points in $V \cap \Delta$, and let V_2 be the set of points in $V \setminus V_1$ that do not give rise to a surface that crosses Δ. Hence

$$|V_2| \geq n - \frac{n}{c_1(2tm)^{2d}} - \frac{n}{2} \geq \frac{n}{4}.$$

Therefore, each point $v \in V_2$ has the property that $v \times V_1 \subset E_i$ for some fixed i. We define the function $\chi : V_2 \to \{1, \ldots, m\}$, such that $\chi(v) = i$ if and only if i is the minimum index for which $v \times V_1 \subset E_i$. Set $I = \{\chi(v) : v \in V_2\}$ and $m_0 = |I|$. That is, m_0 is the number of distinct relations (colors) between the sets V_1 and V_2. Now the proof falls into two cases.

Case 1. Suppose $m_0 > \log m$. If $i \in I$, then V_1 does not contain $p_i - 1$ elements such that every pair induced by them belongs to E_i. By the induction hypothesis, we have

$$\frac{2^{c\left(\sum_{i=1}^m p_i\right)\log\log(m+2)}}{c_1(2tm)^{2d}} \leq |V_1| \leq 2^{c\left(\sum_{i\in I}(p_i-1)+\sum_{i\notin I} p_i\right)\log\log(m+2)} \leq 2^{c\left(\sum_{i=1}^m p_i - m_0\right)\log\log(m+2)}.$$

Hence,

$$cm_0 \log\log(m+2) \leq \log(c_1(2tm)^{2d}) \leq 2d\log(c_1 2tm),$$

which implies

$$m_0 \leq \frac{2d\log(c_1 2tm)}{c\log\log(m+2)},$$

and we have a contradiction for sufficiently large $c = c(d,t)$.

Case 2. Suppose $m_0 \leq \log m$. By the pigeonhole principle, there is a subset $V_3 \subset V_2$, such that $|V_3| \geq \frac{n}{4m_0}$ and $V_1 \times V_3 \subset E_i$ for some fixed i. Hence, V_3 does not contain $p_i - 1$ points such that every pair induced by them is in E_i. By the induction hypothesis, we have

$$\frac{2^{c\left(\sum_{i=1}^m p_i\right)\log\log(m+2)}}{4m_0} \leq |V_3| \leq 2^{c\left(-1+\sum_{i=1}^m p_i\right)\log\log(m+2)}.$$

Therefore,

$$c\log\log(m+2) \leq \log(4m_0) \leq \log(4\log m),$$

which is a contradiction since c is sufficiently large. This completes the proof of Theorem 1.2.

□

5 A superlinear lower bound on distinct distances

In this section, we prove Theorem 1.5. We need the following result in extremal graph theory.

Theorem 5.1 (Kővári-Sós-Turán [25], Erdős)**.** *Let $G = (U, V, E)$ be a bipartite graph such that $|U| = m$ and $|V| = n$. If G does not contain the subgraph $K_{2,r}$ with 2 vertices in U and r vertices in V, then*

$$|E(G)| \leq O(m\sqrt{n} + n),$$

where the hidden constant depends on r.

In particular, we have that for any fixed r, all $K_{2,r}$-free graphs on n vertices have a most $O(n^{3/2})$ edges. Our next result improves this upper bound under the additional condition that the edge set E of the graph is a semi-algebraic relation with constant description complexity. A similar theorem was proved in [16] for $K_{r,r}$-free graphs $G = (V, E)$ with $V \subset \mathbb{R}^d$, where $d < r$).

Theorem 5.2. *For fixed $d \geq 4$, $r \geq 2$ and $t \geq 1$, let $V \subset \mathbb{R}^d$ be an n-element point set equipped with a semi-algebraic relation $E \subset \binom{V}{2}$ such that E has complexity at most t. If (V, E) is $K_{2,r}$-free, then $|E| \leq O\left(n^{\frac{3}{2} - \frac{1}{8d-18+\varepsilon}}\right)$, where ε is an arbitrarily small positive constant.*

Proof. Since E is semi-algebraic of description complexity at most t, there are polynomials $g_1, \ldots, g_t$ and a Boolean formula Φ such that for $u, v \in V$,

$$(u, v) \in E \quad \Leftrightarrow \quad \Phi(g_1(u, v) \geq 0, \ldots, g_t(u, v) \geq 0) = 1.$$

For each point $v \in V$, let $\sigma_{i,v} = \{x \in \mathbb{R}^2 : g_i(x, v) = 0\}$, $1 \leq i \leq t$. Set $\Sigma = \{\sigma_{i,v} : 1 \leq i \leq t, v \in V\}$. Note that $|\Sigma| = tn$.

For $r = n^{1/(4d-9+2\varepsilon)}$, we apply Lemma 2.1 to Σ, which partitions $\mathbb{R}^d$ into at most

$$c_1 r^{2d-4+\varepsilon} = c_1 n^{\frac{2d-4+\varepsilon}{4d-9+2\varepsilon}}$$

cells Δ_i, where $c_1 = c_1(t)$ and ε is arbitrarily small, such that each cell is crossed by at most $|\Sigma|/r$ surfaces from Σ. By the pigeonhole principle, there is a cell $\Delta \subset \mathbb{R}^d$ that contains at least

$$\frac{n}{c_1 n^{\frac{2d-4+\varepsilon}{4d-9+2\varepsilon}}} = (1/c_1) n^{1-\frac{2d-4+\varepsilon}{4d-9+2\varepsilon}}$$

points from V. Let $V' \subset V$ be a set of exactly $\left\lceil (1/c_1)n^{1-\frac{2d-4+\varepsilon}{4d-9+2\varepsilon}} \right\rceil$ points in $V \cap \Delta$. We can assume that n is sufficiently large so that $|V'| \geq 2$ holds. Let $U \subset V$ be the set of points in V that gives rise to at least one surface in Σ that crosses Δ. By the cutting lemma,

$$|U| \leq \frac{tn}{r} = cn^{1-\frac{1}{4d-9+2\varepsilon}}.$$

Since (V, E) is $K_{2,r}$-free, Theorem 5.1 implies that

$$|E(U, V')| \leq c'(|V'||U|^{1/2} + |U|) \leq c_3 n^{1-\frac{1}{4d-9+2\varepsilon}},$$

where $c_3 = c_3(r, t)$. Hence there is a point $v \in V'$ such that v has at most

$$\frac{|E(U, V')|}{|V'|} \leq c_4 n^{\frac{1}{2}-\frac{1}{8d-18+4\varepsilon}}$$

neighbors in U. Since G is $K_{2,r}$-free, there are at most $r-1$ points in $V \setminus (U \cup V')$ that are neighbors of v. Hence,

$$|N_G(v)| \leq |V'| + (r-1) + O\left(n^{\frac{1}{2}-\frac{1}{8d-18+4\varepsilon}}\right) \leq O\left(n^{\frac{1}{2}-\frac{1}{8d-18+4\varepsilon}}\right).$$

We remove v and repeat this argument until no vertices remain in V. Then we have that

$$|E(G)| \leq O\left(n^{\frac{3}{2}-\frac{1}{8d-18+4\varepsilon}}\right).$$

□

We will use the $d = 4$ special case of Theorem 5.2. We also need the following simple lemma.

Lemma 5.3. *Let $G = (V, E)$ be a graph with maximum degree p. Then G contains a matching of size $|E|/(2p)$.*

Proof. We proceed by induction on $|E|$. The base case $|E| = 1$ is trivial. Now assume the statement holds for all graphs with fewer than $|E|$ edges. We build our matching greedily. Let $uv \in E(G)$ be the first edge in our matching, and delete all other edges incident to u and v. Hence we have at least $|E| - 2p + 1$ edges remaining, and the statement follows by the induction hypothesis.

□

We are now ready to prove Theorem 1.5.

Proof of Theorem 1.5. Here we will show that

$$D\left(n, p, \binom{p}{2} - p + 6\right) \geq \Omega\left(n^{1+\frac{1}{7+\delta}}\right),$$

where δ is an arbitrarily small constant. Let V be a planar n-element point that determines m distinct distances $d_1, \ldots, d_m$, with the property that any p points from V determine at least $q = \binom{p}{2} - p + 6$ distinct distances, where $p \geq 6$. Let $E_i \subset \binom{V}{2}$ such that $uv \in E_i$ if and only if $|u - v| = d_i$, and let $G_i = (V, E_i)$. Notice that G_i has maximum degree smaller than p, since otherwise we would have p points determining at most $\binom{p}{2} - p + 1 < q$ distinct distances. By Lemma 5.3, there is a matching $E_i' \subset E_i$ of size $|E_i|/(2p)$ for every i.

Let $U = V \times V \subset \mathbb{R}^4$ and we define the relation $Q \subset \binom{U}{2}$ such that $(u_1u_2, v_1v_2) \in E$ if and only if u_1, u_2, v_1, v_2 are all distinct and $|u_1 - v_1| = |u_2 - v_2|$ or $|u_1 - v_2| = |u_2 - v_1|$. Clearly, Q is a semi-algebraic relation with description complexity at most four, where t is an absolute constant. By the Cauchy-Schwarz inequality, we have

$$m|Q| \geq m\sum_{i=1}^{m} 4|E_i'|^2 \geq m4\sum_{i=1}^{m}\left(\frac{|E_i|}{p}\right)^2 \geq \frac{1}{(p)^2}\left(\sum_{i=1}^{m}|E_i|\right)^2 \geq \frac{n^4 - 2n^3}{(2p)^2},$$

which implies

$$m \geq \frac{n^4 - 2n^3}{(2p)^2|Q|}. \tag{3}$$

Hence, it suffices to bound $|Q|$ from above. By a standard probabilistic argument, we can partition $U = U_1 \cup U_2$ such that at least half of the edges in Q are between U_1 and U_2 and $|U_1|, |U_2| \geq \lfloor |U|/2 \rfloor$. Therefore, it is enough to bound the number of edges between U_1 and U_2.

Fix a vertex $u_1u_2 \in U_1$, and let $N(u_1u_2) \subset U_2$ such that $N(u_1u_2) = \{v_1v_2 \in U_2 : (u_1u_2, v_1v_2) \in Q\}$. Consider the graph $G' = (V, N(u_1u_2))$. We claim that G' has maximum degree less than $p - 3$. Indeed, suppose there is a vertex v with degree $p - 3$ in G' with neighbors $w_1, \ldots, w_{p-3}$.

Observation 5.4. *The points $u_1, u_2, v, w_1, \ldots w_{p-3}$ determine at most $\binom{p}{2} - p + 3$ distinct distances.*

Proof. We proceed by induction on p. The base case $p = 4$ follows since $(u_1u_2, vw_1) \in Q$. Now assume that the statement holds up to $p - 1$. By the induction hypothesis, $u_1, u_2, v, w_1, \ldots, w_{p-1}$ determine at most

$$\binom{p-1}{2} - (p-1) + 3 = \binom{p}{2} - 2(p-1) + 3$$

distinct distances. Since $(u_1u_2, vw_{p-3}) \in Q$, we have either $|u_1 - v| = |u_2 - w_{p-3}|$ or $|u_1 - w_{p-3}| = |u_2 - v|$. Thus, adding w_{p-3} introduces at most $p - 2$ new distances. Therefore, $u_1, u_2, v, w_1, \ldots w_{p-3}$ determine at most

$$\binom{p}{2} - 2(p-1) + 3 + (p-2) = \binom{p}{2} - p + 3$$

distinct distances.

□

Since $G' = (V, N(u_1u_2))$ has maximum degree at most $p - 2$, by Lemma 5.3, there is a subset $N'(u_1u_2) \subset N(u_1u_2)$ of size $\frac{|N(u_1u_2)|}{2(p-2)}$ that forms a matching in G'. Now let $Q' \subset Q$ be the edges between U_1 and U_2 such that $(u_1u_2, v_1v_2) \in Q'$ if and only if $(u_1u_2, v_1v_2) \in N'(u_1u_2)$. Notice we have

$$|Q'| = \sum_{u_1u_2 \in U_1} |N'(u_1u_2)| \geq \sum_{u_1u_2 \in U_1} \frac{|N(u_1u_2)|}{2(p-2)} \geq \frac{|Q|}{4(p-2)}, \tag{4}$$

and therefore it suffices to bound $|Q'|$.

Observation 5.5. *If the bipartite graph (U_1, U_2, Q') contains the subgraph $K_{2,r}$ with two vertices in U_1 and r vertices in U_2, then there are $2r + 4$ points in V that determine at most $\binom{2r+4}{2} - 2r$ distinct distances.*

Proof. Let $u_1u_2, u_3u_4 \in U_1$ and $v_1v_2, \ldots, v_{2r-1}v_{2r} \in U_2$ induce a $K_{2,r}$ in Q'. Then the points $v_1, v_2, \ldots, v_{2r-1}, v_{2r}$ are all distinct in V. Now on the other side, either we have that u_1, u_2, u_3, u_4 are distinct, or we can assume without loss of generality that $u_2 = u_3$, say.

Case 1. Suppose u_1, u_2, u_3, u_4 are all distinct points. Since $(u_1u_2, v_iv_{i+1}), (u_3u_4, v_iv_{i+1}) \in Q'$, for every odd integer $i \in \{1, 3, 5, \ldots, 2r-1\}$, we get $2r$ elements of Q', and each such element gives a repeated distance. Hence, the number of repetitions is at least $2r$, so $u_1, u_2, u_3, u_4, v_1, \ldots v_{2r}$ determine at most $\binom{2r+4}{2} - 2r$ distinct distances.

Case 2. Suppose $u_2 = u_3$. In view of Observation 5.4 (with $p = 5$), the five points u_1, u_2, u_4 and v_i, v_{i+1}, for i odd, determine at most $\binom{5}{2} - 2$ distinct distances. Hence, $u_1, u_2, u_4, v_1, \ldots v_{2r}$ determine at most

$$\binom{2r+3}{2} - 2r$$

distinct distances. Now adding any point w to $u_1, u_2, u_4, v_1, \ldots v_{2r}$ gives us $2r + 4$ points that determine at most

$$\binom{2r+3}{2} - 2r + (2r+3) = \binom{2r+4}{2} - 2r$$

distinct distances.

□

Suppose that p is even and recall that $p \geq 6$. By Observation 5.5, (U_1, U_2, Q') is $K_{2,(p-4)/2}$-free, since otherwise we would have p points in V that determine at most $\binom{p}{2} - p + 4$ distinct distances, which is a contradiction. By Theorem 5.2, with $d = 4$ and for a fixed p, we have

$$|Q'| \leq O\left(|U|^{\frac{3}{2} - \frac{1}{14+\varepsilon}}\right) \leq O\left(n^{3 - \frac{2}{14+\varepsilon}}\right).$$

Together with (4) and (3), we obtain

$$m \geq \Omega\left(\frac{n^4}{|Q'|}\right) \geq \Omega\left(n^{1+\frac{2}{14+\varepsilon}}\right)$$

If p is odd then (U_1, U_2, Q') is $K_{2,(p-5)/2}$-free. Indeed, otherwise by Observation 5.5 we would have $p - 1$ points that determine at most

$$\binom{p-1}{2} - p + 5 = \binom{p}{2} - 2p + 6$$

distinct distances. Adding any point to our collection gives us p points that determine at most $\binom{p}{2} - p + 5$ distinct distances, a contradiction. Just as above, we have

$$|Q'| \leq O\left(|U|^{\frac{3}{2} - \frac{1}{14+\varepsilon}}\right) \leq O\left(n^{3 - \frac{2}{14+\varepsilon}}\right).$$

Combining this with (4) and (3), we obtain

$$m \geq \Omega\left(\frac{n^4}{|Q'|}\right) \geq \Omega\left(n^{1+\frac{2}{14+\varepsilon}}\right) = \Omega\left(n^{1+\frac{1}{7+\delta}}\right).$$

□

6 Ramsey-Turán problem for semi-algebraic graphs

In this section, we prove Theorem 6. The upper bound in Theorem 1.7 follows from

Theorem 6.1. *Let $\varepsilon > 0$ and let $G = (V, E)$ be an n-vertex graph where V is a set of n points in $\mathbb{R}^d$ and E is a semi-algebraic binary relation with complexity at most t. If G is K_{2p}-free and $|E| > \frac{1}{2}\left(1 - \frac{1}{p-1} + \varepsilon\right) n^2$, then G contains an independent set of size γn, where $\gamma = \gamma(d, t, p, \varepsilon)$.*

Proof. We apply Theorem 1.1, with $m = 1$ and parameter $\varepsilon/4$ to obtain an equitable partition $\mathcal{P} : V = V_1 \cup \cdots \cup V_K$ such that $\frac{4}{\varepsilon} \le K \le \left(\frac{4}{\varepsilon}\right)^c$, where $c = c(d, t)$, and all but an $\frac{\varepsilon}{4}$-fraction of all pairs of parts in $\mathcal{P}$ are homogeneous (complete or empty with respect to E). If $n \le 10K$, then G has an independent set of size one, and the theorem holds trivially. So we may assume $n > 10K$.

By deleting all edges inside each part, we have deleted at most

$$K\binom{\lceil n/K\rceil}{2} \le \frac{4n^2}{5K} \le \frac{n^2}{5}\varepsilon$$

edges. Deleting all edges between non-homogeneous pairs of parts, we lose an additional at most

$$\left\lceil \frac{n}{K}\right\rceil^2 \frac{\varepsilon}{4}\frac{K^2}{2} \le \varepsilon\frac{n^2}{5}$$

edges. In total, we have deleted fewer than $2\varepsilon n^2/5$ edges of G. The only edges remaining in G are edges between homogeneous pairs of parts, and we have at least $\frac{1}{2}\left(1 - \frac{1}{p-1} + \varepsilon/5\right) n^2$ edges. By Turán's theorem (Theorem 1.6), there is at least one remaining copy of K_p, and these vertices lie in p distinct parts $V_{i_1}, \ldots, V_{i_p} \in \mathcal{P}$ that form a complete p-partite subgraph. If any of these parts forms an independent set in G, say V_{i_j}, then we have an independent set of order $|V_{i_j}| \ge \lfloor n/K \rfloor \ge \gamma n$, where $\gamma = \gamma(d, t, \epsilon, p)$, and we are done. Otherwise, there is an edge in each of the p parts, and the endpoints of these p edges form a K_{2p} in G, a contradiction.

□

The lower bound on $\mathbf{RT}(n, K_{2p-1}, o(n))$ and $\mathbf{RT}(n, K_{2p}, o(n))$ in Theorem 1.7 can be seen by the following construction. First, let us recall a result of Walczak.

Lemma 6.2 ([38]). *For every $n \ge 1$, there is a collection S of $n/(p-1)$ segments in the plane whose intersection graph G_S is triangle-free and has no independent set of size $cn/\log\log n$ where c is an absolute constant.*

The construction. In the plane, take $p - 1$ dilated copies of S, label them as $S_1, \ldots, S_{p-1}$, so that S_i lies inside a ball with center $(i, 0)$ and radius $1/10$. Set $V = S_1 \cup \cdots \cup S_{p-1}$. Note that $|S_i| = n/(p-1)$ so that $|V| = n$. Let $G = (V, E)$ be the graph whose vertices are the n segments in the plane, and two vertices are connected by an edge if and only if the corresponding segments cross or their left endpoints are at least $1/2$ apart. The graph G consists of a complete $(p-1)$-partite graph, where each part induces a copy of the triangle-free graph G_S. Clearly, G is K_{2p-1}-free and does not contain an independent set of size $cn/\log\log n$. Moreover,

$$|E(G)| \ge \frac{1}{2}\left(1 - \frac{1}{p-1}\right) n^2.$$

Since every segment can be represented by a point in $\mathbb{R}^4$ and the intersection and distance relations have bounded description complexity (see [1]), E is a symmetric semi-algebraic relation with complexity at most c, where c is an absolute constant.

References

[1] N. Alon, J. Pach, R. Pinchasi, R. Radoičić, and M. Sharir, Crossing patterns of semi-algebraic sets, *J. Combin. Theory Ser. A* **111** (2005), 310–326.

[2] S. Basu, R. Pollack, and M.-R. Roy, *Algorithms in Real Algebraic Geometry*, Springer-Verlag, Berlin, 2003.

[3] P. Brass, W. Moser, and J. Pach, *Research Problems in Discrete Geometry*, Berlin, Germany, Springer-Verlag, 2005.

[4] B. Chazelle, H. Edelsbrunner, L. Guibas, and M. Sharir, A singly exponential stratification scheme for real semi-algebraic varieties and its applications, *Theor. Comput. Sci.* **84** (1991), 77–105.

[5] D. Conlon and J. Fox, Bounds for graph regularity and removal lemmas, *Geom. Funct. Anal.* **22** (2012), 1191–1256.

[6] D. Conlon, J. Fox, C. Lee, and B. Sudakov, The Erdős-Gyárfás problem on generalized Ramsey numbers, *Proceedings of the London Mathematical Society* **110** (2015), 1–15.

[7] P. Erdős, On sets of distances of n points, *Amer. Math. Monthly* **53** (1946), 248–150.

[8] P. Erdős, Solved and unsolved problems in combinatorics and combinatorial number theory, in: *Proc. 12th Southeastern Conf. on Combinatorics, Graph Theory and Computing, Vol. I (Baton Rouge, La., 1981), Congr. Numer.* **32** (1981), 49–62.

[9] P. Erdős, Some of my recent problems in combinatorial number theory, geometry and combinatorics, in: *Graph Theory, Combinatorics, Algorithms and Applications, vol. 1* (Y. Alavi et al., eds.), Wiley 1995, 335–349.

[10] P. Erdős and A. Gyárfás, A variant of the classical Ramsey problem, *Combinatorica* **17** (1997), 459–467.

[11] P. Erdős, A. Hajnal, V. Sós, and E. Szemerédi, More results on Ramsey-Turán type problems, *Combinatorica* **3** (1983), 69–81.

[12] P. Erdős and V. Sós, Some remarks on Ramsey's and Turan's theorem, *Coll. Math. Soc. J. Bolyai* **4**, *Comb. Theory and its Appl.*, North-Holland (1969), 395–404.

[13] P. Erdős and G. Szekeres, A combinatorial problem in geometry, *Compos. Math.* **2** (1935), 463–470.

[14] J. Fox, M. Gromov, V. Lafforgue, A. Naor, and J. Pach, Overlap properties of geometric expanders, *J. Reine Angew. Math.* (Crelle's Journal), **671** (2012), pp. 49–83.

[15] J. Fox and L. M. Lovász, A tight lower bound for Szemerédi's regularity lemma, arXiv preprint arXiv:1403.1768, 2014.

[16] J. Fox, J. Pach, A. Sheffer, A. Suk, and J. Zahl, A semi-algebraic version of Zarankiewicz's problem, arXiv preprint arXiv:1407.5705, 2014.

[17] J. Fox, J. Pach, and A. Suk, Density and regularity theorems for semi-algebraic hypergraphs, in: *Proc. 26th Ann. ACM-SIAM Symp. on Discrete Algorithms(SODA '15). San Diego CA,* 2015, 1517–1530.

[18] J. Fox and B. Sudakov, Ramsey-type problem for an almost monochromatic K_4, *SIAM Journal of Discrete Mathematics* **23** (2008), 155–162.

[19] W. T. Gowers, Lower bounds of tower type for Szemerédi's uniformity lemma, *Geom. Funct. Anal.* **7** (1997), 322–337.

[20] L. Guth, N. Katz, On the Erdős distinct distances problem in the plane, *Annals of Math.* **181** (2015), 155–190.

[21] P. Keevash, Hypergraph Turán Problems, *Surveys in Combinatorics*, Cambridge University Press, 2011, 83–140.

[22] V. Koltun, Almost tight upper bounds for vertical decompositions in four dimensions, *J. ACM* **51** (2004), 699–730.

[23] J. Komlós and M. Simonovits, Szemerédi's regularity lemma and its applications in graph theory, in: *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), Bolyai Soc. Math. Stud.* **2**, J. Bolyai Math. Soc., Budapest, 1996, 295–352.

[24] A. V. Kostochka and D. Mubayi, When is an almost monochromatic K_4 guaranteed?, *Combin. Probab. Comput.* **17** (2008), 823–830.

[25] P. Kővári, V. Sós, and P. Turán, On a problem of Zarankiewicz, *Colloq. Math.* **3** (1954), 50–57.

[26] J. Matoušek, *Lectures on Discrete Geometry*, Springer-Verlag, New York, 2002.

[27] G. Moshkovitz and A. Shapira, A short proof of Gowers' lower bound for the regularity lemma, *Combinatorica*, DOI: 10.1007/s00493-014-3166-4.

[28] D. Mubayi, Edge-coloring cliques with three colors on all 4-cliques, *Combinatorica* **18** (1998), 293–296.

[29] I. Schur, Über die Kongruenz $x^m + y^m = z^m \mod p$, *Jahresber. Deutch. Math. Verein.* **25** (1916), 114–117.

[30] M. Simonovits and V.T. Sós, Ramsey-Turán theory, *Discrete Math.* **229** (2001), 293–340.

[31] G. Sárkőzy and S. Selkow, On edge colorings with at least q colors in every subset of p vertices, *Electron. J. Combin.* **8** (2001), no. 1.

[32] A. Suk, Semi-algebraic Ramsey numbers, arXiv preprint arXiv:1406.6550, 2014.

[33] E. Szemerédi, Integer sets containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 299–345.

[34] E. Szemerédi, On graphs containing no complete subgraph with 4 vertices (Hungarian), *Mat. Lapok* **23** (1972), 113–116.

[35] E. Szemerédi, Regular partitions of graphs, *Colloques Internationaux CNRS* **260** *Problémes Combinatoires et Théorie des Graphes, Orsay (1976)*, CNRS, 1976, 399–401.

[36] P. Turán, On an extremal problem in graph theory, *Matematikai és Fizikai Lapok* (in Hungarian) 48 (1941), 436–452.

[37] P. Varnavides, On certain sets of positive density, *J. London Math. Soc.* **34** (1959), 358–360.

[38] B. Walczak, Triangle-free geometric intersection graphs with no large independent sets, *Discrete Comput. Geom.* **53** (2015), 221–225.

[39] X. Xiaodong, X. Zheng, G. Exoo, S.P. Radziszowski, Constructive lower bounds on classical multicolor Ramsey numbers, *Electronic Journal of Combinatorics* **11(1)** (2004).