The Schur-Erdős problem for semi-algebraic colorings

Jacob Fox* János Pach† Andrew Suk‡

Abstract

We consider m-colorings of the edges of a complete graph, where each color class is defined semi-algebraically with bounded complexity. The case $m = 2$ was first studied by Alon et al., who applied this framework to obtain surprisingly strong Ramsey-type results for intersection graphs of geometric objects and for other graphs arising in computational geometry. Considering larger values of m is relevant, e.g., to problems concerning the number of distinct distances determined by a point set.

For $p \geq 3$ and $m \geq 2$, the classical Ramsey number $R(p; m)$ is the smallest positive integer n such that any m-coloring of the edges of K_n, the complete graph on n vertices, contains a monochromatic K_p. It is a longstanding open problem that goes back to Schur (1916) to decide whether $R(p; m) \leq 2^{cm}$, where $c = c(p)$. We prove that this is true if each color class is defined semi-algebraically with bounded complexity, and that the order of magnitude of this bound is tight. Our proof is based on the Cutting Lemma of Chazelle *et al.*, and on a Szemerédi-type regularity lemma for multicolored semi-algebraic graphs, which is of independent interest. The same technique is used to address the semi-algebraic variant of a more general Ramsey-type problem of Erdős and Shelah.

1 Introduction

The Ramsey 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. The existence of $R(p; m)$ follows from the celebrated theorem of Ramsey [20] from 1930, and for the special case when $p = 3$, Issai Schur proved the existence of $R(3; m)$ in 1916 in his work related to Fermat's Last Theorem [21]. He showed that

$$\Omega(2^m) \leq R(3; m) \leq O(m!).$$

While the upper bound has remained unchanged over the last 100 years, the lower bound was successively improved and the current record is $R(3; m) = \Omega(3.199^m)$ due to Xiaodong *et al.* [25]. It is a major open problem in Ramsey theory, for which Erdős offered some price money, to close the gap between the lower and upper bounds for $R(3; m)$.

In this paper, we study edge-colorings of complete graphs where each color class is defined algebraically with bounded complexity. Over the last decade, several researchers have shown that

*Stanford University, Stanford, CA. Supported by a Packard Fellowship and by NSF grant DMS-1855635. Email: jacobfox@stanford.edu.

†EPFL, Lausanne and Rényi Institute, Budapest. Supported by Swiss National Science Foundation Grants 200020-162884 and 200021-175977. Email: pach@cims.nyu.edu.

‡Department of Mathematics, University of California at San Diego, La Jolla, CA, 92093 USA. Supported by an NSF CAREER award and by an Alfred Sloan Fellowship. Email: asuk@ucsd.edu.

some of the classical theorems in extremal combinatorics can be significantly improved if the underlying graphs are intersection graphs of geometric objects of bounded "description complexity" or bounded VC-dimension, graphs of incidences between points and hyperplanes, distance graphs, or, more generally, semi-algebraic graphs [1, 11, 5, 10, 22]. To make this statement more precise, we need to introduce some terminology. 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 necessarily 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. Hence, all edges are unordered and we denote $uv = \{u, v\}$. We also assume that the dimension d and complexity t are fixed parameters, and $n = |V|$ tends to infinity.

Let $R_{d,t}(p; m)$ be the minimum n such that every n-element point set V in $\mathbb{R}^d$ equipped with m semi-algebraic binary relations (edge-colorings) $E_1, \ldots, E_m \subset \binom{V}{2}$, each of complexity at most t, where $E_1 \cup \cdots \cup E_m = \binom{V}{2}$, contains a subset $S \subset V$ of size p such that $\binom{S}{2} \subset E_k$ for some k. Note that the relations E_i are not necessarily disjoint, that is, an edge uv may have several colors. Clearly $R_{d,t}(p; m) \leq R(p; m)$. It was known that for fixed $d, t \geq 1$, $R_{d,t}(3; m) = 2^{O(m \log\log m)}$, which is much smaller than Schur's bound $R(3; m) = O(m!)$ mentioned in the first paragraph of the Introduction; see [22]. In this paper, we completely settle Schur's problem for semi-algebraic graphs, by showing that in this setting Schur's lower bound (which is semi-algebraic with bounded complexity) is tight in the sense that these Ramsey numbers grow exponentially in the number of colors. In fact, we prove this in a more general form, for any $p \geq 3$.

Theorem 1.1. *For integers $d, t \geq 1$ and $p \geq 3$, there is a constant $c = c(d, t, p)$ such that*

$$R_{d,t}(p; m) \leq 2^{cm}.$$

Our proof uses geometric techniques and is based on the Cutting Lemma of Chazelle, Edelsbrunner, Guibas, and Sharir [3] described in Section 2.

Edge-colorings of semi-algebraic graphs with m colors can be used, e.g., for studying problems concerning the number of distinct distances determined by a point set; see [12]. One can explore the fact that multicolored semi-algebraic graphs have a very nice structural characterization, reminiscent of Szemerédi's classic regularity lemma for general graphs [24], but possessing much stronger homogeneity properties. Our next theorem provides such a characterization, which is of independent interest. To state our result, we need some notation and terminology.

A partition is called *equitable* if any two parts differ in size by at most one. According to Szemerédi's lemma, for every $\varepsilon > 0$ there is a $K = K(\varepsilon)$ such that the vertex set of every graph has an equitable partition into at most K parts such that all but at most an ε-fraction of the pairs of parts are ε-regular.[1] It follows from Szemerédi's proof that $K(\varepsilon)$ may be taken to be an exponential tower of 2s of height $\varepsilon^{-O(1)}$. Gowers [14] used a probabilistic construction to show that such an enormous bound is indeed necessary.

Alon *et al.* [1] (see also Fox, Gromov *et al.* [9]) established a strengthening of the regularity lemma for point sets in $\mathbb{R}^d$ equipped with a semi-algebraic relation E. It was shown in [1] that for any semi-algebraic graph of bounded complexity defined on the vertex set $V \subset \mathbb{R}^d$ (that is, for any semi-algebraic binary relation $E \subset \binom{V}{2}$), V 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 $V_1 \times V_2 \cap E = \emptyset$. The first proof of this theorem was essentially qualitative: it gave a poor estimate for the number of parts in such a partition. Fox, Pach, and Suk [11] gave a stronger quantitative form of this result, showing that the number of parts can be taken to be polynomial in $1/\varepsilon$.

Let V be an n-element point set in $\mathbb{R}^d$ equipped with m semi-algebraic relations $E_1, \ldots, E_m$ such that $E_1 \cup \cdots \cup E_m = \binom{V}{2}$ of bounded complexity. In other words, suppose that the edges of the complete graph on V are colored with m colors, where each color class is semi-algebraic. Then, for any $\varepsilon > 0$, an m-fold repeated application of the result of Fox, Pach, and Suk [11] 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_k, i.e., all edges between the two parts are of color k, for some k. In Section 4, we strengthen this result by showing that the number of parts can be taken to be *polynomial* in m/ε.

Theorem 1.2. *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 each E_k 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 $4/\varepsilon \leq K \leq (m/\varepsilon)^c$ parts such that all but an ε-fraction of the pairs of parts are complete with respect to some relation E_k.*

In Section 5, we apply this result to solve a problem of Erdős and Shelah [6] in the semi-algebraic setting. Let d, t, p, q, n be positive integers, $p \geq 3$, and $2 \leq q \leq \binom{p}{2}$. Let $f_{d,t}(n,p,q)$ be the minimum m such that there exists a semi-algebraic m-coloring of the edges of the complete graph of n vertices (with parameters d and t, as above) with the property that each edge has exactly one color and any set of p vertices induces at least q distinct colors. Notice that here we must assume that the color classes E_i are disjoint, since otherwise $f_{d,t}(n,p,q) = q$ by assigning all q colors to every edge. Our next theorem precisely determines the smallest q for a given p, where $f_{d,t}(n,p,q)$ changes from a $\log n$ to a power of n.

Theorem 1.3. *For fixed integers $d, t \geq 1$, there is a $c = c(d,t) > 0$ such that for $p \geq 3$, we have*

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

Moreover, for $t \geq 4$,

[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 $|d(V_i', V_j') - d(V_i, V_j)| \leq \varepsilon$.

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

In [12], we studied a geometric instance of this problem, where every set of p points induces at least q *distinct distances*.

Our paper is organized as follows. In the next section, we describe the Cutting Lemma of Chazelle *et al.*, which is the main geometric tool used in all proofs. In Section 3, we establish Theorem 1.1. Section 4 contains the proof of our multicolored semi-algebraic regularity lemma, Theorem 1.2, which is applied in the following section to deduce Theorem 1.3. We end this paper with some concluding remarks.

2 The cutting lemma

The main tool we use to prove Theorems 1.1 and 1.2 is commonly referred to as the *cutting lemma*, which we now recall. 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

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

We say that a semi-algebraic set in d-space has *description complexity* at most t if the number of inequalities is at most t, and each polynomial g_i has degree at most t. 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 $\Delta \not\subset \sigma$.

Let Σ be a collection of surfaces in $\mathbb{R}^d$, each having bounded degree. A $(1/r)$-*cutting* for Σ is a family Ψ of disjoint (possibly unbounded) semi-algebraic sets of bounded complexity such that

1. each $\Delta \in \Psi$ is crossed by at most $|\Sigma|/r$ surfaces from Σ, and
2. the union of all $\Delta \in \Psi$ is $\mathbb{R}^d$.

In [3], Chazelle *et al.* (see also [15]) proved the following.

Lemma 2.1 (Cutting lemma)**.** *Let Σ be a multiset of N surfaces in $\mathbb{R}^d$, each surface having degree at most t, and let r be an integer parameter such that $1 \leq r \leq N$. Then there is a constant $c_1 = c_1(d, t)$ such that Σ admits a $(1/r)$-cutting Ψ, where $|\Psi| \leq c_1 r^{2d}$, and each semi-algebraic set $\Delta \in \Psi$ has complexity at most c_1.*

We note that the original statement of Chazelle *et al.* [3] and Koltun [15] is stronger. Namely, they also guarantee that the number of cells in the cutting Ψ is at most $r^{2d-4+\epsilon}$ for $d \geq 4$. Here, for simplicity, we use the weaker bound of $c_1 r^{2d}$, as stated above.

3 Multicolor Ramsey numbers for small cliques –Proof of Theorem 1.1

Theorem 1.1 will easily follow from Theorem 3.1 below. For integers $p_1, \ldots, p_m \geq 2$, $d, t \geq 1$, let $R_{d,t}(p_1, \ldots, p_m)$ be the minimum integer n with the following property. Every complete graph K_n, whose n vertices lie in $\mathbb{R}^d$ and whose edges are colored with m colors such that each color class is defined by a semi-algebraic relation of description complexity t, contains a monochromatic copy of K_{p_k} in color k for some $1 \leq k \leq m$.

Theorem 3.1. *For any $d, t \geq 1$ and $p \geq 3$, there exists a constant $c = c(d,t,p)$ satisfying the following condition. For any m integers $p_1, \ldots, p_m \leq p$, we have*

$$R_{d,t}(p_1, \ldots, p_m) \leq 2^{c \sum_{k=1}^{m} p_k}.$$

Proof. Fix $d, t \geq 1$, $p \geq 3$ and set $c = c(d,t,p)$ to be a large constant that will be determined later. We will show that $R_{d,t}(p_1, \ldots, p_m) \leq 2^{c \sum_{k=1}^{m} p_k}$ by induction on $s = \sum_{k=1}^{m} p_k$. The base case $s \leq 10 \cdot 2^{10dtp}$ follows for c sufficiently large.

Now assume that the statement holds for $s' < s$. Set $n = 2^{cs}$ and let V be an n-element point set in $\mathbb{R}^d$ equipped with semi-algebraic relations $E_1, \ldots, E_m \subset \binom{V}{2}$ such that $\binom{V}{2} = E_1 \cup \cdots \cup E_m$ and each E_k has complexity at most t. Recall that an edge uv may have several colors. We will show that there is a subset $S \subset V$ of size p_k such that $\binom{S}{2} \subset E_k$ for some $k \in [m]$, in other words, we will find a monochromatic copy of K_{p_k} in color k for some $k \in [m]$. Throughout the proof, we will let c_1 be as defined in Lemma 2.1.

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

$$uv \in E_k \quad \Leftrightarrow \quad \Phi_k(g_{k,1}(u,v) \geq 0, \ldots, g_{k,t}(u,v) \geq 0) = 1.$$

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

Before we continue, let us briefly sketch the idea of the proof. We start by applying Lemma 2.1 (the cutting lemma) to $\Sigma = \{\sigma_{k,\ell}(v) : k \in [m], \ell \in [t], v \in V\}$, the set of surfaces which determines the neighborhoods of each vertex, and obtain a space partition which induces a partition of the vertex set $V = V_1 \cup \cdots \cup V_K$. If there is a "large" part V_j with many distinct colors appearing in $V_j \times (V \setminus V_j)$, then we show that V_j induces few distinct colors, and by induction we can find a monochromatic copy of K_{p_k} for some $k \in [m]$. If none of the "large" parts has the above property, the colors of nearly all edges can be defined by much fewer polynomial inequalities, i.e., by a much smaller set of surfaces $\Sigma' \subset \Sigma$. Now we can repeat.

In what follows, we spell out these ideas in full detail. Set $m_0 = m$ and define $m_i = 4d \log(c_1 m_{i-1} t)$ for $i > 0$. We will establish the following claim.

Claim 3.2. *Let V and $E_1, \ldots, E_m \subset \binom{V}{2}$ be defined as above. Then for $i \geq 0$ we will recursively find either*

1. *a monochromatic copy of K_{p_k} in color k for some $k \in [m]$, or*
2. *a function $\chi_i : V \to 2^{[m]}$ such that $|\chi_i(v)| \leq m_i$, and the number of edges $uv \in \binom{V}{2}$ with the property that for one of its endpoints, say u, no color assigned to uv belongs to $\chi_i(u)$, is at most $\frac{4n^2}{tm_{i-1}}$. We will refer to these edges as bad at stage i. All edges that are not bad are called good at this stage, meaning that, there is a color k appearing on uv such that $k \in \chi_i(u)$, and there is a color k' appearing on uv such that $k' \in \chi_i(v)$.*

Proof. We start by setting $\chi_0(v) = [m]$ for all $v \in V$, and $m_0 = m$. Having found χ_i with the properties above, we will produce χ_{i+1} as follows. We have $m_{i+1} = 4d \log(c_1 m_i t)$, and let us assume that $m_i > (8c_1dtp)^2$. Hence, there are at most $\frac{4n^2}{tm_{i-1}}$ bad edges. Let Σ be the set of surfaces $\sigma_{k,\ell}(v)$, where $v \in V$, $k \in \chi_i(v)$, and $1 \leq \ell \leq t$. This implies that $|\Sigma| \leq nm_it$.

We apply Lemma 2.1 to Σ with parameter $r = (tm_i)^2$ to obtain a $(1/(tm_i)^2)$-cutting $\Psi = \{\Delta_1, \Delta_2, \ldots, \Delta_{K_0}\}$, such that $K_0 \leq c_1(tm_i)^{4d}$. Hence, we have a partition $\mathcal{P}_0 : V = V_1 \cup \cdots \cup V_{K_0}$, where $V_j = V \cap \Delta_j$ for $\Delta_j \in \Psi$. For each part V_j of size greater than $2n/(tm_i)$, we (arbitrarily) partition V_j into parts of size $\lfloor 2n/(tm_i) \rfloor$ and possibly one additional part of size less than $2n/(tm_i)$. Let $\mathcal{P} : V = V_1 \cup \cdots \cup V_K$ be the resulting partition, where $K \leq 2c_1(tm_i)^{4d}$ and $|V_j| \leq 2n/(tm_i)$ for all j.

Now we define $\chi_{i+1}(v)$ for all $v \in V$.

Case 1. If $v \in V_j$ for some V_j with $|V_j| < \frac{n}{2c_1(tm_i)^{4d+1}}$, we set $\chi_{i+1}(v) = \emptyset$.

Case 2. Suppose $v \in V_j$ such that $|V_j| \geq \frac{n}{2c_1(tm_i)^{4d+1}}$. In order to define $\chi_{i+1}(v)$, we need some preparation. Let $\Delta_j \in \Psi$ such that $V_j \subset \Delta_j$. We define $X_j \subset V \setminus V_j$ to be the set of vertices from $V \setminus V_j$ that gives rise to a surface in Σ that crosses Δ_j. By Lemma 2.1, the cutting lemma, we have $|X_j| \leq n/(tm_i)$.

Fix a vertex $v \in V \setminus \{V_j, X_j\}$. Since none of the surfaces of the form $\sigma_{k,\ell}(v)$, where $k \in \chi_i(v)$ and $\ell \in \{1, \ldots, t\}$, cross Δ_j, either $v \times V_j$ is monochromatic with color k for some $k \in \chi_i(v)$, or none of the colors in $\chi_i(v)$ appear in $v \times V_j$. Let S_j be the set of vertices $v \in V \setminus \{V_j, X_j\}$ satisfying the former condition and let T_j denote a set of vertices $v \in V \setminus \{V_j, X_j\}$ satisfying the latter one. Since there are at most $4n^2/(tm_{i-1})$ bad edges, we have

$$|T_j| \frac{n}{2c_1(tm_i)^{4d+1}} \leq \frac{4n^2}{tm_{i-1}},$$

which implies

$$|T_j| \leq \frac{8nc_1(tm_i)^{4d+1}}{tm_{i-1}} \leq \frac{n}{tm_i},$$

where the last inequality follows from the fact that $m_{i-1} = 2^{m_i/4d}/(c_1t)$, and the assumption $m_i > (8c_1dtp)^2$. Now, suppose there are at least $m_{i+1} = 4d\log(c_1tm_i)$ distinct colors between V_j and S_j. Let $I = \{k_1, \ldots, k_{m_{i+1}}\} \subset [m]$ be the set of these m_{i+1} distinct colors. Then there are m_{i+1} vertices $v_1, \ldots, v_{m_{i+1}} \in S_j$, possibly with repetition, such that $v_w \times V_j$ is monochromatic with color $k_w \in I$, for each $w \in \{1, \ldots, m_{i+1}\}$. Hence, if V_j contains a monochromatic copy of K_{p_k-1} in color $k \in I$, we would have a monochromatic copy of K_{p_k} in color k. On the other hand, if V_j does not contain a monochromatic copy of K_{p_k-1} in color k for no $k \in I$, then, using that

$$|V_j| \geq \frac{n}{2c_1(tm_i)^{4d+1}} > 2^{cs-8d\log(c_1m_it)} > 2^{c(s-m_{i+1})} = 2^{c\left(\sum_{k\in I}(p_k-1)+\sum_{k\notin I}p_k\right)}$$

for a sufficiently large c, we obtain by induction that there is a monochromatic copy of K_{p_k} in color k where $k \notin I$.

Therefore, we can assume that the number of distinct colors between V_j and S_j is less than $m_{i+1} = 4d\log(c_1m_it)$. For every vertex $v \in V_j$, define $\chi_{i+1}(v)$ as the set of all colors that appear on the edges belonging to $v \times S_j$.

Now that we have defined m_{i+1} and χ_{i+1} such that $|\chi_{i+1}(v)| \leq m_{i+1}$ for all $v \in V$, it remains to show that the number of edges $uv \in \binom{V}{2}$ with the property that for one of its endpoints, say

u, no color assigned to uv belongs to $\chi(u)$, is at most $\frac{4n^2}{tm_i}$. Let $B \subset \binom{V}{2}$ be the collection of such edges. Notice that if $uv \in B$, then either

1. both u and v lie inside the same part in the partition $\mathcal{P}$, or
2. u or v lies inside a part V_j such that $|V_j| < \frac{n}{2c_1(tm_i)^{4d+1}}$, or
3. $u \in V_j$ with $|V_j| \geq \frac{n}{2c_1(tm_i)^{4d+1}}$ and $v \in X_j \cup T_j$, or
4. $v \in V_j$ with $|V_j| \geq \frac{n}{2c_1(tm_i)^{4d+1}}$ and $u \in X_j \cup T_j$.

Since each part in $\mathcal{P}$ has at most $2n/(tm_i)$ vertices, the number of edges of type 1 is at most $\frac{tm_i}{2}\left(\frac{2n}{tm_i}\right)^2/2 = n^2/(tm_i)$. The number of edges of type 2 is at most

$$\left(\frac{n}{2c_1(tm_i)^{4d+1}}\right) K \cdot n \leq \frac{n^2}{tm_i}.$$

Since $|X_j|, |T_j| \leq n/(tm_i)$, the number of edges of types 3 and 4 is at most

$$\sum_j |V_j| \frac{2n}{tm_i} \leq \frac{2n^2}{tm_i}.$$

Hence, $|B| \leq \frac{4n^2}{tm_i}$. Therefore, either we have found a monochromatic copy of K_{p_k} in color k for some $k \in [m]$, or we have found m_{i+1} and χ_{i+1} with the desired properties. □

Let w be the minimum integer such that $m_w \leq (8c_1dtp)^2$. Then either we have found a monochromatic copy of K_{p_k} in color k for some $k \in [m]$, or we have obtained m_w and χ_w with the desired properties. Since there are at most $4n^2/(tm_{w-1}) < n^2/8$ bad edges, there is a vertex $v \in V$ incident to at least $n/2$ good edges. Moreover, since $|\chi_w(v)| \leq m_w \leq (8c_1dtp)^2$, at least $\frac{n}{2(8c_1dtp)^2}$ of these edges incident to v have color k' for some color $k' \in \chi_w(v)$. Let $S \subset V$ be the set of endpoints of these edges. If S contains a monochromatic copy of $K_{p_{k'}-1}$ in color k', then we are done. On the other hand, if S does not contain a monochromatic copy of $K_{p_{k'}-1}$ in color k', and using the lower bound

$$|S| \geq \frac{n}{2(8c_1dtp)^2} = \frac{2^{cs}}{2(8c_1dtp)^2} \geq 2^{c(\sum_{k\neq k'} p_k + (p_{k'}-1))},$$

for $c = c(d,t,p)$ sufficiently large, we conclude by induction that S contains a monochromatic copy of K_{p_k} for some $k \neq k'$. This completes the proof of Theorem 3.1. □

4 Multicolor semi-algebraic regularity lemma –Proof of Theorem 1.2

First, we prove the following variant of Theorem 1.2, which easily implies Theorem 1.2.

Theorem 4.1. *For any $\varepsilon > 0$, every n-element point set $V \subset \mathbb{R}^d$ equipped with semi-algebraic binary relations $E_1, \ldots, E_m \subset \binom{V}{2}$ such that $\binom{V}{2} = E_1 \cup \cdots \cup E_m$ and each E_k 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_k for all $k = 1, \dots, m$.

Proof. For each relation E_k, let $g_{k,1}, \dots, g_{k,t} \in \mathbb{R}[x_1, \dots, x_{2d}]$ be polynomials of degree at most t, and let Φ_k be a boolean formula such that

$$uv \in E_k \quad \Leftrightarrow \quad \Phi_k(g_{k,1}(u,v) \geq 0; \dots ; g_{k,t}(u,v) \geq 0) = 1.$$

For each point $x \in \mathbb{R}^d$, $k \in \{1, \dots, m\}$, and $\ell \in \{1, \dots, t\}$, we define the surface

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

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

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

We apply Lemma 2.1 to Σ with parameter $r = tm/\varepsilon$ to obtain a $(1/r)$-cutting Ψ, where $|\Psi| = s \leq c_1 \left(\frac{tm}{\varepsilon}\right)^{2d}$, such that each semi-algebraic set $\Delta_i \in \Psi$ has complexity at most c_1, where c_1 is defined in Lemma 2.1. Hence, at most $tmn/r = \varepsilon n$ surfaces from Σ cross Δ_i for every i. This implies that at most εn points in V give rise to at least one surface in Σ that crosses Δ_i.

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

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

That is, $\Delta_{i,j,k}$ will correspond to the vertices in Δ_i that may not be homogeneous to the set of vertices in Δ_j with respect to color k.

Observation 4.2. *For any i, j, and k, the semi-algebraic set $\Delta_{i,j,k}$ has complexity at most $c_3 = c_3(d,t)$.*

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

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

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

Set $\mathcal{F}_i = \{\Delta_{i,j,k} : 1 \leq k \leq m, 1 \leq j \leq s\}$. We partition the points in U_i into equivalence classes, where two points $u, v \in U_i$ are equivalent if and only if u belongs to the same members of $\mathcal{F}_i$ as v does. Since $\mathcal{F}_i$ gives rise to at most $c_3|\mathcal{F}_i|$ polynomials of degree at most c_3, by the Milnor-Thom theorem (see [18] Chapter 6), the number of distinct sign patterns of these $c_3|\mathcal{F}_i|$ polynomials is at most $(50c_3(c_3|\mathcal{F}_i|))^d$. Hence, there is a constant $c_5 = c_5(d,t)$ such that U_i is partitioned into at most $c_5(ms)^d$ equivalence classes. After repeating this procedure to each U_i, we obtain a partition of our point set $V = V_1 \cup \dots \cup V_K$ 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_{w_i} obtained from Lemma 2.1 such that $U_{w_i} = V \cap \Delta_{w_i}$ and $V_i \subset U_{w_i} \subset \Delta_{w_i}$. 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_{w_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_k for any k. Summing over all i, we have

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

where the sum is taken over all pairs i, j such that (V_i, V_j) is not complete with respect to E_k for all k. $\square$

Proof of Theorem 1.2. Apply Theorem 4.1 with approximation parameter $\varepsilon/2$. Hence, there is a partition $\mathcal{Q} : V = U_1 \cup \cdots \cup U_{K'}$ into $K' \leq (m/\varepsilon)^c$ parts with $c = c(d,t)$ and $\sum |U_i||U_j| \leq (\varepsilon/2)|V|^2$, where the sum is taken over all pairs (i,j) such that (U_i, U_j) is not complete with respect to E_k for all k.

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 $E_1, \ldots, E_m$. $\square$

5 Generalized Ramsey numbers for semi-algebraic colorings –Proof of Theorem 1.3

Due to the lack of understanding of the classical Ramsey number $R(p;m)$, Erdős and Shelah (see [6]) introduced the following generalization, which was studied by Erdős and Gyárfás in [7].

Definition 5.1. 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. Here, both p and q are considered fixed integers, where $p \geq 3$, $2 \leq q \leq \binom{p}{2}$, and n tends to infinity. Trivially, we have $f(n,p,\binom{p}{2}) = \binom{n}{2}$, and at the other end, 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,

$$\Omega\left(\frac{\log n}{\log\log n}\right) \leq f(n,3,2) \leq O(\log n). \tag{1}$$

Erdős and Gyárfás [7] determined certain ranges for $q \in \{2,3,\ldots,\binom{p}{2}\}$ for which $f(n,p,q)$ is quadratic, linear, and subpolynomial in n. In particular, they showed that

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

which implies that $f(n,p,q)$ is polynomial in n for $q \geq p$. Surprisingly, estimating $f(n,p,p-1)$ is much more difficult. They [7] asked for p fixed if $f(n,p,p-1) = n^{o(1)}$. The trivial lower bound is $f(n,p,p-1) \geq f(n,p,2) \geq \Omega\left(\frac{\log n}{\log\log n}\right)$, which was improved by several authors [17, 13], and it is now known [4] that $f(n,p,p-1) \geq \Omega(\log n)$. In the other direction, Mubayi [19] found an elegant construction which implies $f(n,4,3) \leq e^{O(\sqrt{\log n})}$, and later, Conlon *et al.* [4] gave another example which implies $f(n,p,p-1) \leq e^{(\log n)^{1-1/(p-2)+o(1)}}$. Hence, it is now known that $f(n,p,p-1)$ does not grow as a power in n.

Here, we study the variant of the function $f(n,p,q)$ for point sets $V \subset \mathbb{R}^d$ equipped with semi-algebraic relations. Let $f_{d,t}(n,p,q)$ 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. We note that here we require that each edge receives *exactly* one color. Clearly, we have $f(n,p,q) \leq f_{d,t}(n,p,q)$. Theorem 1.3 stated in the Introduction shows the exact value of q for which $f_{d,t}(n,p,q)$ changes from $\log n$ to a power of n.

In the rest of this section, we prove Theorem 1.3. 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_k has complexity at most t, $\binom{V}{2} = E_1\cup\cdots\cup E_m$, and $E_k \cap E_\ell = \emptyset$ for all $k \neq \ell$. Let $S_1, S_2 \subset V$ be q-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 $uv \in E_k$ if and only if $h(u)h(v) \in E_k$.

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 $uv \in E_k$ for some fixed k. Notice that given an s-layered set S, there are at most s relations $E_{k_1},\ldots,E_{k_s}$ such that $\binom{S}{2} \subset E_{k_1} \cup\cdots\cup E_{k_s}$. Hence, the lower bound in Theorem 1.3 is a direct consequence of the following result.

Theorem 5.2. *Let $s \geq 1$ 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_k has complexity at most t, $E_1\cup\cdots\cup E_m = \binom{V}{2}$, and $E_k\cap E_\ell = \emptyset$ for all $k \neq \ell$. If $m \leq n^{\frac{1}{cs^2}}$, 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 1.2 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

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

and $c_2 = c_2(d,t)$. Since all but an ε fraction of the pairs of parts in $\mathcal{P}$ are complete with respect to E_k for some k, by Turán's theorem, there are $m^{s-1}+1$ parts $V_i' \in \mathcal{P}$ such that each pair $(V_i', V_j') \in \mathcal{P}\times\mathcal{P}$ is complete with respect to some relation E_k. 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

$$|V_i'|^{\frac{1}{c(s-1)^2}} \geq \left(\frac{n}{c_2 m^{10d^2 s}}\right)^{\frac{1}{c(s-1)^2}} \geq m^{\frac{cs^2-10c_2d^2s}{c(s-1)^2}} \geq m.$$

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 $S_i \times S_j \subset E_k$ for some k, 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 m distinct 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 $uv \in E_i$ if and only if $(u+c, v+c) \in E_i$ 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

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

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).

Let us remark that the arguments above hold for semi-algebraic relations $E_1,\ldots,E_m$ that are not necessarily disjoint if one defines a (p,q)-coloring as follows. Given a coloring $\chi : \binom{V(K_n)}{2} \to 2^{[m]}$ on the edges of K_n, where each edge receives *at least* one color among $[m]$, χ is a (p,q)-*coloring* if for every set $S \subset V$ of size p, no matter how you choose one color in $\chi(uv)$ for each edge $uv \in \binom{S}{2}$, S will induce at least q distinct colors.

6 Concluding remarks

In [22], it was shown that $R_{1,t}(3;m) > (1681)^{m/7}$ for $t > 5$, thus implying that the upper bound in Theorem 1.1 is tight up to a constant factor in the exponent. This can be improved as follows. Let $C(p) = \lim_{m\to\infty} R(p;m)^{1/m}$. Note that this limit exists by considering product colorings, but may be finite or infinite. Then for each $C < C(p)$, there is a $t = t(C,p)$, such that for all m sufficiently large we have

$$R_{1,t}(p;m) > C^m.$$

Indeed, take a fixed coloring of the edges of K_N which realizes $R(p;m_0) > C^{m_0}$, and recursively blow up this graph by introducing m_0 new colors at each stage. Then this coloring can be realized semi-algebraically in $\mathbb{R}$ with $t = O(m_0^2)$ linear constraints for each color class based on distances.

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] 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.

[4] D. Conlon, J. Fox, C. Lee, and B. Sudakov, The Erdős-Gyárfás problem on generalized Ramsey numbers, *Proc. Lond. Math. Soc.* **110** (2015), 1–15.

[5] D. Conlon, J. Fox, J. Pach, B. Sudakov, A. Suk, Ramsey-type results for semi-algebraic relations, *Trans. Amer. Math. Soc.* **366** (2014), 5043–5065.

[6] 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.

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

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

[9] 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.

[10] J. Fox, J. Pach, A. Sheffer, A. Suk, J. Zahl, A semi-algebraic version of Zarankiewicz's problem, *J. Eur. Math. Soc.* **19** (2017), 1785–1810.

[11] J. Fox, J. Pach, and A. Suk, A polynomial regularity lemma for semi-algebraic hypergraphs and its applications in geometry and property testing, *SIAM J. Comput.* **45** (2016), 2199–2223.

[12] J. Fox, J. Pach, and A. Suk, More distinct distances under local conditions, *Combinatorica* **3**8 (2018), no. 2, 501–509.

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

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

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

[16] 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.

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

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

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

[20] F. Ramsey, On a problem in formal logic, *Proc. London Math. Soc.* **30** (1930), 264–286.

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

[22] A. Suk, Semi-algebraic Ramsey numbers, *J. Combin. Theory Ser. B* **116** (2016), 465–483.

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

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

[25] X. Xiaodong, X. Zheng, G. Exoo, S.P. Radziszowski, Constructive lower bounds on classical multicolor Ramsey numbers, *Electron. J. Combin.* **11(1)** (2004).