Applications of random algebraic constructions to hardness of approximation

Boris Bukh, Karthik C. S., and Bhargav Narayanan

Abstract. In this paper, we show how one may (efficiently) construct the following two types of extremal combinatorial objects whose existence was previously conjectural.

- Panchromatic graphs: For fixed $k \in \mathbb{N}$, a k-panchromatic graph is, roughly speaking, a balanced bipartite graph with one partition class equipartitioned into k colour classes in which the common neighbourhoods of panchromatic k-sets of vertices are much larger than those of k-sets that repeat a colour. The question of their existence was raised by Karthik and Manurangsi in 2020.
- Threshold graphs: For fixed $k \in \mathbb{N}$, a k-threshold graph is, roughly speaking, a balanced bipartite graph in which the common neighbourhoods of k-sets of vertices on one side are much larger than those of $(k+1)$-sets. The question of their existence was raised by Lin in 2018.

Our constructions utilise varieties cut out by (carefully chosen) random polynomials, and the analysis of these constructions relies on machinery from algebraic geometry; the technical tools developed to accomplish this may be of independent interest. As applications of our constructions, we show the following conditional time lower bounds on the parameterised set intersection problem where, given a collection of n sets over universe $[n]$ and a parameter k, the goal is to find k sets with the largest intersection.

- Assuming ETH, for any computable function $F\colon \mathbb{N} \to \mathbb{N}$, no $n^{o(k)}$-time algorithm can approximate the parameterised set intersection problem up to factor $F(k)$. This improves considerably on the previously best-known result under ETH due to Lin, who ruled out any $n^{o(\sqrt{k})}$ time approximation algorithm for this problem.
- Assuming SETH, for every $\varepsilon > 0$ and any computable function $F\colon \mathbb{N} \to \mathbb{N}$, no $n^{k-\varepsilon}$-time algorithm can approximate the parameterised set intersection problem up to factor $F(k)$. No result of comparable strength was previously known under SETH, even for solving this problem exactly.

1. Introduction

Over the last five decades, a symbiotic relationship has developed between the areas of extremal combinatorics and complexity theory (broadly construed); see the wonderful book of Jukna [33] or one of the surveys of Alon [2, 3, 4] for various applications of extremal combinatorial objects to proving lower bounds in theoretical computer

Date: 1 July, 2021.

2010 *Mathematics Subject Classification.* Primary 05C35; Secondary 68Q27, 68R05.

science. In particular, this synergistic exchange with extremal combinatorics can be explicitly seen in subareas such as circuit/formula lower bounds [8, 35], communication complexity [15, 40, 28], error correcting codes [49, 7, 30], and derandomization [5, 46, 20, 16].

In this paper, our first goal is to prove the existence of certain extremal bipartite graphs, namely threshold graphs and panchromatic graphs. The question of their existence was motivated by applications in hardness of approximation, and our second goal is to prove, using these graphs, conditional time lower bounds on the parameterised set intersection problem. Our constructions will rely crucially on random polynomials, and our third goal here is to prove various results, likely of independent interest, about the common zeroes of random polynomials over finite fields. Before we can state our results, it will help to have some background, to which we now turn.

Over the last few years, a new area in theoretical computer science, namely *hardness of approximation in* P, has benefited significantly from some of the deep results in extremal combinatorics. Hardness of approximation in P, roughly speaking, maybe treated as the union of two subareas, namely, hardness of approximation in *parameterised complexity* and hardness of approximation in *fine-grained complexity*.

In parameterised complexity, one studies the computational complexity of problems with respect to multiple parameters of the input or output. For example, in the k-SetIntersection problem, we are given a collection of n sets over the universe $[n]$ and a parameter k as input, and the goal is to find k sets in the collection which maximize the intersection size. A problem with inputs of size n along with a parameter k is said to be fixed parameter tractable if it can be solved by an algorithm running in time $T(k) \cdot \text{poly}(n)$ for some computable function T. In many interesting cases, including for the k-SetIntersection problem, assuming the $\mathsf{W}[1]{\neq}\mathsf{FPT}$ hypothesis, it is possible to show that no such algorithm exists i.e., that the problem is not fixed parameter tractable. In light of this, one could then ask for approximation algorithms. In the case of k-SetIntersection, the task would then be to design an approximation algorithm running in time $T(k) \cdot \text{poly}(n)$ that can find k sets in the collection whose intersection size is at least $1/F(k)$ of the intersection size of the optimal solution for some pair of computable functions T and F. Inapproximability results in parameterised complexity aim to typically rule out such algorithms (under the $\mathsf{W}[1]{\neq}\mathsf{FPT}$ hypothesis) for various classes of functions F; a notion particularly relevant to this paper is that of *total* FPT *inapproximability*, in which we rule out $F(k)$-approximation algorithms running in $T(k) \cdot \text{poly}(n)$ time for all computable functions T and F. We refer the reader to the textbooks [24, 22] for an excellent introduction to the area.

In fine-grained complexity, one aims to refine the Cobham–Edmonds thesis [25, 19] by trying to understand the exact time required to solve problems in P, by basing their conditional time lower bounds on several plausible (and popular) conjectures such as

SETH and ETH. For example, k-SetIntersection can be naïvely solved by exhaustive search, i.e., by computing the intersection sizes of all k-tuples of sets from the given collection of n sets. Can we do any better? For instance, is there an algorithm running in time $n^{o(k)}$ that can solve k-SetIntersection? Or even less ambitiously, is there an algorithm running in time $n^{k-0.1}$ that can solve k-SetIntersection? The theory of fine-grained complexity aims to rule out such algorithms, and inapproximability results in this area aim to prove the same conditional time lower bounds, but now against approximation algorithms.

A major difficulty addressed by results in hardness of approximation in P is that of generating a gap, i.e., one must start with a hard problem with no gap (for which the time lower bound is only against exact algorithms) and reduce it to a problem of interest while generating a non-trivial gap in the process. One of the main approaches to generate the aforementioned gap, and the motivation behind our construction of threshold graphs, is the *threshold graph composition* (TGC) framework introduced in the breakthrough work of Lin [42] to show the total FPT inapproximability of the k-SetIntersection problem. This technique was later used to prove the first non-trivial inapproximability result for the k-SetCover problem [17], and in the proof of the current state-of-the-art inapproximability result for the same [43]. Moreover, the result on the k-SetIntersection problem in [42] was used in [9] as the starting point to prove inapproximability results for problems in coding theory such as the k-Minimum Distance problem and the k-Nearest Codeword problem, and for lattice problems such as the k-Shortest Vector problem and the k-Nearest Vector problem.

At a very high level, in TGC, we compose an instance of the input problem that has no gap, with an extremal combinatorial object called a *threshold graph*, to produce a gap instance of the desired problem. The two main challenges in using this framework are to construct the requisite threshold graph, and to find the right way to compose the input and the threshold graph. Our construction of threshold graphs will address the first of these challenges.

Another key issue that often arises in proving conditional time lower bounds for problems in P is the following. When trying to prove time lower bounds for a particular problem, it is often natural (and sometimes seemingly necessary) to first prove the lower bound for a coloured version of the same problem, and then reduce it to the uncoloured version of the problem. For instance, if we would like to prove lower bounds based on SETH for a problem Ψ, then it is almost always the case that we first divide the variable set of size n (of the SAT formula arising from the SETH assumption) into k equal parts and reduce the problem of deciding SAT to a problem in P where, given as input k collections each containing $2^{n/k}$ partial assignments to the subset of n/k variables in that part, we would like to find one partial assignment from each collection that, when stitched together, forms a full satisfying assignment to the original SAT instance.

From this problem (in P), if we would like to reduce to Ψ, it is often convenient (and sometimes imperative) to first reduce to a k-coloured version of Ψ, and then reduce this coloured version to Ψ itself. This final task is sometimes easy, such as for problems like k-SetCover or k-OrthogonalVectors, but often non-trivial, such as for k-SetIntersection or closest pair in a point-set. It is worth reiterating here that in the other direction, reducing the uncoloured problem to its coloured version is almost always easy; typically, one can reduce the uncoloured variant to its coloured counterpart via the celebrated colour coding technique of Alon, Yuster and Zwick [6].

In [23, 37], the authors proposed the *panchromatic graph composition* (PGC) framework to address this issue, and this serves as the motivation behind our construction of *panchromatic graphs*. In particular, they outlined how these panchromatic graphs, assuming they exist, can be composed with the coloured version of a problem to reduce it to the uncoloured version of the same problem. Also, it is worth noting that the same issue arises in proving time lower bounds against approximation algorithms as well, i.e., it is often easier to prove hardness of approximation results for coloured versions of problems than for their uncoloured counterparts. With this in mind, it is desirable to have panchromatic graphs with certain additional gap properties so that we can design gap preserving reductions between problems. Our construction of panchromatic graphs will address all of these challenges.

In summary, the role of extremal combinatorial objects in the existing literature on hardness of approximation in P is twofold: threshold graphs are used in the TGC framework to generate gaps in hard problem instances, and panchromatic graphs are used in the PGC framework to reduce hard instances of coloured variants of various computational problems to their uncoloured (computationally easier) counterparts.

1.1. **Our contributions.** Our contributions are primarily twofold. First, in Section 1.1.1, we show how to efficiently construct threshold graphs and panchromatic graphs; even the existence of such graphs was previously conjectural. Second, in Section 1.1.2, we demonstrate some applications of these graphs (with panchromatic graphs featuring more prominently) to prove *tight* conditional time lower bounds, under ETH and SETH, for approximating k-SetIntersection.

1.1.1. *Constructions of panchromatic and threshold graphs.* Here, we describe our main combinatorial results that demonstrate the existence of the aforementioned extremal bipartite graphs.

We start with panchromatic graphs.

Definition 1.1 (Panchromatic graphs; informal version of Definition 3.1)**.** *An (n, k, t, s)-panchromatic graph is a bipartite graph $G(A, B)$ where A is partitioned into k parts, say $A_1, \ldots, A_k$, with $|A_1| = \cdots = |A_k| = |B| = n$ satisfying the following pair of conditions.*

Completeness: *For every k-set $X \subset A$ for which $A_i \cap X \neq \emptyset$ for all $i \in [k]$, the number of common neighbours of X in B is at most t, and a positive fraction (depending only on k) of such k-sets have exactly t common neighbours in B.*

Soundness: *For every k-set $X \subset A$ for which $A_i \cap X$ is empty for some $i \in [k]$, the number of common neighbours of X in B is at most s.*

In [37], the authors studied panchromatic graphs when $k = 2$. Using (non-trivial) density properties of Reed–Solomon codes and Algebraic-Geometric codes, they were able to show that $(n, 2, t, t^{o(1)})$-panchromatic graphs exist for $t = 2^{(\log n)^{1-o(1)}}$, and can be constructed efficiently. They then raised the natural question of existence for general k, indicating that if such graphs exist, they could then potentially be used to prove improved hardness and inapproximability results for k-SetIntersection. We resolve this open problem from [37] and prove the following result.

Theorem 1.2 (Informal restatement of Theorem 3.3)**.** *For each $k \in \mathbb{N}$ and any integer $\lambda > 1$, there exist $(n, k, t, t/\lambda)$-panchromatic graphs for infinitely many $n \in \mathbb{N}$, where $t = t(k, \lambda)$ depends only on k and λ.*

In [37], the authors note that their technique to construct panchromatic graphs is limited to the case of $k = 2$, and remark that one needs to construct objects with more structure than just *maximum distance separable* codes in a certain sense. Our construction, detailed in Section 1.2.1, does just this, introducing new ideas that go beyond standard coding-theoretic properties. On a different note, it is natural to ask if the requirement in the completeness condition that a positive fraction (depending on k) of k-sets have exactly t-sized common neighbourhoods can be strengthened to demand the same of *every* such k-set. It turns out that our result is in fact best-possible in the following sense: as $n \to \infty$ and for any $t = t(k)$, there do not exist $(n, k, t, t-1)$-panchromatic graphs in which a $(1 - 1/t)$-fraction of the panchromatic k-sets have exactly t-sized common neighbourhoods; this may be shown using the Kövári—Sós—Turán theorem and Hölder's inequality, but we omit the details here.

Next, we turn our attention to threshold graphs.

Definition 1.3 (Threshold graphs; informal version of Definition 3.2)**.** *An (n, k, t, s)-threshold graph is a bipartite graph $G(A, B)$ with $|A| = |B| = n$ satisfying the following pair of conditions.*

Completeness: *For every k-set of vertices $X \subset A$, the number of common neighbours of X in B is at least t.*

Soundness: *For every $(k+1)$-set of vertices $X \subset A$, the number of common neighbours of X in B is at most s.*

These graphs are closely related to constructions for Turán-type problems in extremal graph theory. Indeed, if the completeness condition above is weakened to only require

that a positive fraction (depending on k) of k-sets $X \subset A$ have at least t common neighbours in B, then the celebrated norm-graphs of [39, 8] achieve these weakened requirements.

Lin [42] raised the question of the existence of threshold graphs, and noted that if threshold graphs exist, then there is a very short proof showing the total FPT inapproximability of k-SetIntersection as follows. Starting with an instance $H(V, E)$ of the canonical W[1]-hard k-clique problem on n vertices, we combine it with a (n, k, t, s)-threshold graph $G(V, B)$ to yield an instance of $\binom{k}{2}$-SetIntersection with $|E|$ sets on the universe B, where for every edge $e = (u, v) \in E$, we include the element $b \in B$ in the set associated with this edge if and only if b is a common neighbour of u and v in G. It then follows that if there is a k-clique in H, then there are $\binom{k}{2}$ sets whose intersection size is at least t, and if there is no k-clique in H, then every $\binom{k}{2}$ sets have intersection size at most s.

However, since the existence of threshold graphs was previously unknown, the argument showing total FPT inapproximability of k-SetIntersection in [42] is rather delicate. We resolve this open problem from [42] and show that threshold graphs exist, obtaining a very short proof of the total FPT inapproximability of k-SetIntersection as a byproduct.

Theorem 1.4 (Informal restatement of Theorem 3.4)**.** *For each $k \in \mathbb{N}$ and for infinitely many $n \in \mathbb{N}$, there exist $(n, k, n^{\Omega(1/k)}, k^{O(k)})$-threshold graphs.*

The parameters in this result match the parameters obtainable via norm-graphs, but crucially, our construction also achieves the stronger completeness property discussed earlier. In fact, it is possible to improve the $k^{O(k)}$ above to $2^{O(k)}$ using more involved algebraic-geometry, but we avoid the extra complexity of that approach here.

1.1.2. *Applications to the parameterise set intersection problem.* Here, we describe our conditional time lower bounds for the k-SetIntersection problem. In order to set the context for the complexity of this problem, we briefly recall its complexity in the world of NP.

In the world of complexity, SetIntersection is well-known as a notorious problem to prove any kind of hardness of approximation result for; that said, there is a general belief that it is a hard problem as no non-trivial polynomial time approximation algorithms for this problem are known. However, to this date, even ruling out a PTAS under the standard P $\neq$NP hypothesis remains open! The best inapproximability result for this problem is based on assuming that SAT problems of size n cannot be solved by randomised algorithms in time 2^{n^ε}, under which Xavier [51] shows that there is no polynomial time algorithm which can approximate SetIntersection up to factor n^δ for some $\delta = \delta(\varepsilon) > 0$. It is worth noting that this inapproximability result relies on the highly non-trivial and celebrated quasi-random PCP construction of Khot [38].

Given this context, it was truly a breakthrough when Lin [42], introducing some novel techniques, proved the total FPT inapproximability of k-SetIntersection (under W[1] $\neq$ FPT hypothesis). Of course, using our construction of threshold graphs, we now have a very short proof of this powerful result.

Lin [42] further refined his inapproximability result and showed, assuming ETH, that for sufficiently large $k \in \mathbb{N}$, no randomised $n^{o(\sqrt{k})}$-time algorithm can approximate k-SetIntersection to a factor $n^{1/\Omega(\sqrt{k})}$. Clearly, this result is stronger than ruling out $F(k)$ approximation algorithms (for some function F), but the running time lower bound is far from tight. The following result, the first application of our constructions, shows that we can improve on Lin's result and obtain tight running time lower bounds under ETH (albeit for weaker approximation factors).

Theorem 1.5 (Informal restatement of Theorem 6.4)**.** *Let $F : \mathbb{N} \to \mathbb{N}$ be any computable function. Assuming* ETH*, for sufficiently large $k \in \mathbb{N}$, no randomised $n^{o(k)}$-time algorithm can approximate k-*SetIntersection *to a factor $F(k)$.*

In the world of fine-grained complexity, it is also of interest to prove, under stronger assumptions than ETH, even tighter running time lower bounds than the $n^{o(k)}$ bound above. In particular, one would like to rule out $n^{k-0.1}$-time algorithms for k-SetIntersection under SETH, essentially showing that the naïve algorithm for k-SetIntersection is optimal. To the best of our knowledge, prior to our work, it was not known if one could even rule out *exact* algorithms for k-SetIntersection running in $n^{k-0.1}$-time! We remedy this situation; the following strong inapproximability result under SETH is the second application of our constructions.

Theorem 1.6 (Informal restatement of Theorem 6.2)**.** *Let $F : \mathbb{N} \to \mathbb{N}$ be any computable function. Assuming* SETH*, for every $\varepsilon > 0$ and integer $k > 1$, no randomised $n^{k(1-\varepsilon)}$-time algorithm can approximate k-*SetIntersection *to a factor $F(k)$.*

Both of these results are crucially reliant on our construction of panchromatic graphs. It is worth noting that for the coloured variant of k-SetIntersection, one can easily show tight running time lower bounds under ETH and SETH against exact algorithms, and by using non-trivial gap creating techniques, these tight running time lower bounds were extended to near polynomial factor approximation algorithms for the coloured variant in [36]. The situation (for the coloured variant) is similar in the world of NP as well; see [18]. Finally, we remark that by using the hardness of approximation results in [36] under the k-SUM hypothesis, we can use the PGC framework to rule out randomised $n^{k(1/2-\varepsilon)}$-time approximation algorithms for k-SetIntersection, under the k-SUM hypothesis, up to any factor $F(k)$.

1.1.3. *Reverse colour coding.* We conclude this discussion of our results by briefly highlighting a broader implication. For many computational problems, it is often

natural to define and study a coloured variant. For some problems, the coloured variant turns out to be even more natural; for example, any k-CSP (i.e., constraint satisfaction problems of arity k) on k variables can be seen as a coloured version of the maximum edge biclique problem. Establishing computational equivalences between coloured and non-coloured variants of problems is thus a basic question worthy of exploration. As noted earlier, for some problems, there is a straightforward equivalence between the two versions. However, there are many important problems for which this equivalence is nontrivial (and potentially not true). The celebrated colour coding technique of Alon, Yuster and Zwick [6] provides an efficient way for a problem to be reduced to its coloured variant. Our construction of panchromatic graphs (when combined with PGC, as will be described in Section 1.2.2) now gives us a rather general method to reverse the colour coding technique.

1.2. **Our techniques.** Our main technical contribution is the construction of panchromatic graphs and threshold graphs which we describe in Section 1.2.1. We also provide an overview of how these are used to prove Theorems 1.5 and 1.6 in Section 1.2.2

1.2.1. *Constructions of Panchromatic and Threshold Graphs.* To motivate our approach, we start by explaining, briefly, why a natural first attempt at constructing threshold graphs fails. It is natural to consider a random bipartite graph where each edge is included independently with an appropriately chosen probability p. Indeed, it is easy to see that such a construction can ensure that *most* k-sets of vertices on one side have fewer common neighbours than *most* $(k+1)$-sets. However, it is essentially impossible to avoid some *exceptional* k-sets and $(k+1)$-sets at the relevant edge density p. Without getting into the details, the reason for this is simple: the size of the common neighbourhoods in this probability space have long, smoothly-decaying tails, and since there are many sets to consider, it is overwhelmingly likely that exceptional sets exist. For more on this issue, we refer the reader to [10].

When it comes to panchromatic graphs, while there is no immediate natural candidate construction, it seems clear that assuming one wishes to construct such objects randomly, one needs to introduce some level of correlation between different edges, while simultaneously preserving enough independence to allow us to analyse the resulting random graph, a delicate task from a purely probabilistic perspective.

It turns out that there is a natural way to circumvent all the obstacles outlined above, namely, by considering random graphs in which adjacency is determined by a randomly chosen algebraic variety. Concretely, our approach, which works over the finite field $\mathbb{F}_q$ for any prime power $q \in \mathbb{N}$, is as follows.

(1) We construct threshold graphs as follows. We build A by independently sampling q^{k+1} random polynomials of degree d from $\mathbb{F}_q[X_1, \ldots, X_{k+1}]$ for a suitable $d =$

$d(k)$. Then, with $B = \mathbb{F}_q^{k+1}$, we define a bipartite graph G between A and B by joining $f \in A$ to $x \in B$ if $f(x) = 0$.

(2) To construct panchromatic graphs, we proceed as follows. First, we independently choose random polynomials $w_1, \dots, w_k$ of degree D from $\mathbb{F}_q[X_1, \dots, X_k]$ for a suitable $D = D(k)$. Next, for $i \in [k]$, we take A_i to be a set of q^k random polynomials of the form $w_i + p$, where each such p is an independently sampled random polynomial of degree d from $\mathbb{F}_q[X_1, \dots, X_k]$ for a suitable $d = d(k)$. Finally, with $B = \mathbb{F}_q^k$, we define a bipartite graph G between A and B by joining $f \in A$ to $x \in B$ if $f(x) = 0$.

While the random algebraic graphs above are quite easy to describe, their analysis is far from simple; in particular, to prove our main results, we shall rely on the Lang–Weil estimate [41], which is a consequence of the Riemann hypothesis for function fields (but see [47] for a relatively elementary proof). Along the way, we shall prove a several results about the zero sets of random polynomials over finite fields that may be of independent interest. An illustrative example is the following probabilistic analogue of Bézout's theorem over finite fields.

Theorem 1.7 (Informal restatement of Theorems 4.3 and 4.4)**.** *For $k, d \in \mathbb{N}$ and a prime power $q \in \mathbb{N}$, let Z be the (random) number of common roots over $\mathbb{F}_q^k$ of k independently chosen k-variate random $\mathbb{F}_q$-polynomials of degree d. Then, as $q \to \infty$, we have*

$$\mathbb{P}[Z = d^k] \geq \frac{1 - o(1)}{(d^k)!},$$

as well as

$$\mathbb{P}[Z > d^k] = O(q^{-d}).$$

To place these techniques in context, it is worth mentioning that the first traces of this random algebraic method go back some way, to work of Matoušek [44] in discrepancy theory, but it is the variant originating in [10] and developed further in [11, 21] that we shall build upon in this paper.

1.2.2. *Hardness of Approximating* k-SetIntersection. The common starting point for Theorems 1.5 and 1.6 is the Unique k-MaxCover problem defined in [36]. We refrain from defining it here, but it is immediate from its definition that it can be easily reformulated as the coloured version of k-SetIntersection, hereafter panchromatic k-SetIntersection. In panchromatic k-SetIntersection, we are given k collections, each consisting of n subsets of the universe $[n]$, and the goal is to choose one set from each collection such that their intersection size is maximise. From [36], it follows that assuming SETH (respectively ETH), there is no $n^{k-\varepsilon}$-time (respectively $n^{o(k)}$-time) algorithm that can approximate panchromatic k-SetIntersection to an $F(k)$ factor for any computable function F.

It is easier to describe the PGC technique in terms of graphs, so we reformulate the panchromatic k-SetIntersection problem as follows: given a bipartite graph $H(X, Y)$ where $X = X_1 \sqcup \cdots \sqcup X_k$ corresponds to the k collections of sets and Y corresponds to the universe (so $|X_1| = \cdots = |X_k| = |Y| = n$), the goal is to find $(x_1, \ldots, x_k) \in X_1 \times \cdots \times X_k$ which has the largest sized common neighbourhood in Y. We also consider a $(n, k, t, t/\lambda)$-panchromatic graph $G(X, B)$ as guaranteed by our Theorem 1.2. Now, given G and H as above, the PGC technique, roughly speaking, boils down to analyzing the graph $H^*(X, Y \times B)$ where if $(x, b) \in X_i \times B$ is an edge in G and $(x, y) \in X_i \times Y$ is an edge in H, then we have the edge $(x, (y, b)) \in X_i \times Y \times B$ in H^*.

In the completeness case, if the maximum panchromatic common neighbourhood size in H was r, then the same set of vertices would have a common neighbourhood of size $t \cdot r$ in H^*, whereas in the soundness case, if the maximum panchromatic common neighbourhood size in H was s, then the maximum common neighbourhood size is at most $t \cdot s$ in H^*. From the soundness of the panchromatic graph, we know that if we pick k vertices in X not all from different colour classes, then their common neighbourhood is of size at most $(t/\lambda) \cdot |Y|$. The results we desire then follow by setting λ appropriately, and importantly noting that $|Y| = O(r)$ in the hard instances given by [36]; recall that the common neighbourhood problem on H^* where we ignore the colour classes is the k-SetIntersection problem.

Our composition technique using panchromatic graphs strictly improves on the techniques introduced in [23, 37]. The PGC technique described above also improves the inapproximability results of [37], albeit only in the lower order terms, and also simplifies their hardness of approximation proof for the *monochromatic maximum inner product* problem.

1.3. **Organization of Paper.** In Section 2, we formally define the problems and hypotheses of interest in this paper. In Section 3, we carefully define panchromatic and threshold graphs and state our main results about them. In Section 4, we prove some important intermediate results that will be used to analyze our constructions of panchromatic and threshold graphs. In Section 5, we give the construction of panchromatic graphs and threshold graphs. In Section 6, we prove our fine-grained inapproximability results for k-SetIntersection. Finally, in Section 7 we highlight a few important open problems and research directions.

2. Preliminaries

Here, we formally define all the problems and hypotheses used in the paper. First, we define the k-SAT problem and then define the two popular fine-grained hypotheses concerning this problem.

k-SAT problem. In the k-SAT problem, we are given a CNF formula φ over n variables $x_1, \ldots, x_n$ where each clause contains at most k literals. Our goal is to decide if there exist an assignment to $x_1, \ldots, x_n$ which satisfies φ.

In this paper, we require a fine-grained notion (of algorithms) in the complexity class RP and a fine-grained notion of *reverse unfaithful random* (RUR) *reductions* [1, 45]. An FPT notion of such algorithms and reductions was introduced in [9] and the notion of randomised fine-grained reduction was introduced in [13]. A promise problem Π is a pair of languages $(\Pi_{\text{YES}}, \Pi_{\text{NO}})$ such that $\Pi_{\text{YES}} \cap \Pi_{\text{NO}} = \emptyset$. A Monte Carlo algorithm $\mathcal{A}$ is said to be a (one-sided) randomised algorithm for a (promise) problem Π if the following hold.

Yes: For all $x \in \Pi_{\text{YES}}$, $\Pr[\mathcal{A}(x) = 1] \geq 1/2$.
No: For all $x \in \Pi_{\text{NO}}$, $\Pr[\mathcal{A}(x) = 0] = 1$.

Moreover, we say that $\mathcal{A}$ runs in time T if the running time of $\mathcal{A}$ on every randomness is upper bounded by T.

We start by recalling the exponential time hypothesis (ETH); see [31, 32, 50], for example.

Hypothesis 2.1 (Exponential time hypothesis (ETH))**.** *There exists $\epsilon > 0$ such that no (randomised) algorithm can solve 3-SAT on n variables in time $O(2^{\epsilon n})$; moreover, this holds even when restricted to formulae in which each variable appears in at most three clauses.*

We will also recall a stronger hypothesis called the strong exponential time hypothesis (SETH); see [31, 32], for example.

Hypothesis 2.2 (Strong exponential time hypothesis (SETH))**.** *For every $\varepsilon > 0$, there exists $k = k(\varepsilon) \in \mathbb{N}$ such that no (randomised) algorithm can solve k-SAT on n variables in time $O(2^{(1-\varepsilon)n})$; moreover, this holds even when the number of clauses is at most $c(\varepsilon)n$ where $c(\varepsilon)$ denotes a constant that depends only on ε.*

In this paper, we prove tight running time lower bounds for k-SetIntersection (with definitions to follow) assuming ETH by providing a *fine-grained RUR reduction* from 3-SAT to k-SetIntersection, such that YES instances of 3-SAT map to YES instances of k-SetIntersection with high probability and NO instances of 3-SAT always map to NO instances of k-SetIntersection. Using standard techniques, fine-grained RUR reductions can be used to transform a Monte Carlo one-sided randomised algorithm for k-SetIntersection into a Monte Carlo one-sided randomised algorithm for SAT (for example, see [9]). Our SETH based lower bounds are proved similarly, except now relying on k-SAT instead.

Next, we recall the MaxCover problem introduced in [14] which turned out to be the centrepiece of many results in parameterise inapproximability.

k-MaxCover **problem.** A k-MaxCover instance Γ consists of a bipartite graph $G = (V \sqcup W, E)$ where V is partitioned into $V = V_1 \sqcup \cdots \sqcup V_k$ and W is partitioned into $W = W_1 \sqcup \cdots \sqcup W_\ell$. We refer to the V_i's and W_j's as *left super-nodes* and *right super-nodes* of Γ respectively. A solution to k-MaxCover is called a *labeling*, which is a subset of vertices $v_1 \in V_1, \ldots, v_k \in V_k$. We say that a labeling $v_1, \ldots, v_k$ *covers* a right super-node W_i, if there exists a vertex $w_i \in W_i$ which is a common neighbour of all of $v_1, \ldots v_k$. We denote by $\mathsf{MaxCover}(\Gamma)$ the maximal fraction of right super-nodes that can be simultaneously covered, i.e.,

$$\mathsf{MaxCover}(\Gamma) = \frac{1}{\ell}\left(\max_{\text{labeling } v_1,\ldots,v_k} |\{i \in [\ell] \mid W_i \text{ is covered by } v_1, \ldots v_k\}|\right).$$

Given an instance $\Gamma(G, r, s)$ of the k-MaxCover problem as input, the task of distinguishing between the following two cases.

Completeness: $\mathsf{MaxCover}(\Gamma) \geq r$.
Soundness: $\mathsf{MaxCover}(\Gamma) \leq s$.

We also define the *unique* MaxCover *problem*, which is just the MaxCover problem with the following additional structure: for every labeling $S \subset V$ and any right super-node W_i, there is at most one vertex in W_i which is a common neighbour of all vertices nodes in S.

Next, we define the two central computational problems of interest in this paper, namely, k-SetIntersection and its coloured variant, panchromatic k-SetIntersection.

k-SetIntersection **problem.** A k-SetIntersection instance Γ consists of a collection $\mathcal{C}$ of n subsets of a universe $\mathcal{U}$ (typically $[n]$) and integer parameters $r > s$. Given such an input $\Gamma(\mathcal{C}, r, s)$, the goal is to distinguish between the following two cases.

Completeness: There exist k sets $S_{i_1}, \ldots, S_{i_k}$ in $\mathcal{C}$ such that

$$\left|\bigcap_{j\in[k]} S_{i_j}\right| \geq r.$$

Soundness: For every k-tuple of sets $(S_{i_1}, \ldots, S_{i_k})$ in $\mathcal{C}$, we have

$$\left|\bigcap_{j\in[k]} S_{i_j}\right| \leq s.$$

The panchromatic k-SetIntersection problem is the following coloured version of the k-SetIntersection problem.

Panchromatic k-SetIntersection **problem.** A panchromatic k-SetIntersection instance Γ consists of k collections $\mathcal{C}_1, \ldots, \mathcal{C}_k$ each containing n subsets of a universe $\mathcal{U}$ and integer parameters $r > s$. In the panchromatic k-SetIntersection problem, given input $\Gamma(\mathcal{C}_1, \ldots, \mathcal{C}_k, r, s)$, the goal is to distinguish between the following two cases.

Completeness: There exist k sets $S_{i_1}, \ldots, S_{i_k}$ in $\mathcal{C}_1 \times \cdots \times \mathcal{C}_k$ such that

$$\left| \bigcap_{j \in [k]} S_{i_j} \right| \geq r.$$

Soundness: For every k-tuple of sets $(S_{i_1}, \ldots, S_{i_k})$ in $\mathcal{C}_1 \times \cdots \times \mathcal{C}_k$, we have

$$\left| \bigcap_{r \in [k]} S_{i_r} \right| \leq s.$$

An important quantity associated with an instance Γ of panchromatic k-SetIntersection, the *monochromatic number* $z(\Gamma)$ of Γ, is defined by

$$z(\Gamma) = \max_{\substack{X \subset \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_k \\ |X| = k}} \left| \bigcap_{S \in X} S \right|$$

The following connection between the unique k-MaxCover problem and the panchromatic k-SetIntersection problem will prove useful.

Proposition 2.3. *Every unique* MaxCover *instance*

$$\Gamma(V = V_1 \sqcup \cdots \sqcup V_k, W = W_1 \sqcup \cdots \sqcup W_\ell, E, r, s)$$

is also a panchromatic k-SetIntersection *instance* $\Gamma'(\mathcal{C}_1, \ldots, \mathcal{C}_k, r', s')$ *over universe* $\mathcal{U}$ *with monochromatic number* z *where we have*

(1) $|\mathcal{U}| = |W|$,
(2) $|\mathcal{C}_i| = |V_i|$ *for all* $i \in [k]$,
(3) $r' = r\ell$,
(4) $s' = s\ell$, *and*
(5) $z \leq |W|$.

Proof. For every $w \in W$, we create a universe element $u_w \in \mathcal{U}$. For every $v \in V_i$, we create a set $S_v \in \mathcal{C}_i$, and we include u_w in S_v if there is an edge between w and v in Γ. Note that w is a common neighbour of $(v_1, \ldots, v_k) \in V_1 \times \cdots V_k$ if and only if u_w is in $\cap_{i \in [k]} S_{v_i}$. Furthermore, note that since Γ is an instance of unique k-MaxCover, the quantity $\ell \cdot \mathsf{MaxCover}(\Gamma)$ is simply the maximum number of common neighbours of any k vertices in V subject to picking one vertex from each V_i; the result follows. □

3. Extremal bipartite graphs

Here, we define panchromatic and threshold graphs a little more carefully, and also state precisely what our constructions accomplish.

We start with *panchromatic* graphs.

Definition 3.1 ((n, m, k, t, s, p)-panchromatic graph)**.** *A bipartite graph $G(A, B)$ where A is partitioned into k parts $A_1, \dots, A_k$ with $|A_1| = \cdots = |A_k| = n$ and $|B| \leq m$ satisfying the following pair of conditions.*

Completeness: *For every k-set $X \subset A$ for which $A_i \cap X \neq \emptyset$ for all $i \in [k]$, the number of common neighbours of X in B is at most t, and a p-fraction of such k-sets have exactly t common neighbours in B.*

Soundness: *For every set $X \subset A$ of size k for which $A_i \cap X$ is empty for some $i \in [k]$, the number of common neighbours of X in B is at most s.*

Next, we turn to *threshold* graphs.

Definition 3.2 ((n, m, k, t, s, p)-threshold graph)**.** *A bipartite graph $G(A, B)$ with $|A| = n$ and $|B| \leq m$ satisfying the following pair of conditions.*

Completeness: *For a p-fraction of k-sets of vertices $X \subset A$, the number of common neighbours of X in B is at least t.*

Soundness: *For every $(k+1)$-set of vertices X in A, the number of common neighbours of X in B is at most s.*

We show that both types of graphs may be constructed with reasonable dependencies between the various parameters involved. Both constructions are easy to describe, with the edge sets of the graphs in question coming from the varieties cut out by (carefully chosen) random polynomials; the analysis of these constructions is far from trivial however, and relies on some amount of machinery from algebraic geometry.

For panchromatic graphs, we have the following result which, in particular, ensures that such graphs exist.

Theorem 3.3. *For each $k \in \mathbb{N}$ and any integer $\lambda > 1$, there is a strictly increasing sequence $\{n_i \in \mathbb{N}\}_{i \in \mathbb{N}}$ such that for every $i \in \mathbb{N}$, there exists a distribution $\mathcal{D}_{k,\lambda,n_i}$ over bipartite graphs on $(k+1)n_i$ vertices with the following properties.*

(1) *A graph can be sampled from $\mathcal{D}_{k,\lambda,n_i}$ in $O_k(n_i^2)$ time using $O_k(n_i \log n_i)$ random coins.*

(2) *For $G \sim \mathcal{D}_{k,\lambda,n_i}$, writing $D = \lambda(k^2+2)$, we have*

$$\mathbb{P}\left(G \text{ is a } (n_i, n_i, k, D^k, D^k/\lambda, (4(D^k)!)^{-1})\text{-panchromatic graph}\right) \geq (4(D^k)!)^{-1}.$$

Moreover, for every $n \in \mathbb{N}$, there exists $i \in \mathbb{N}$ such that $n \leq n_i \leq 2^k n$.

For threshold graphs, we have the following analogous result, which again, in particular, ensures that such graphs exist.

Theorem 3.4. *For each $k \in \mathbb{N}$, there is a strictly increasing sequence $\{n_i \in \mathbb{N}\}_{i \in \mathbb{N}}$ such that for every $i \in \mathbb{N}$, there exists a distribution $\mathcal{D}_{k,n_i}$ over bipartite graphs on $2n_i$ vertices with the following properties.*

(1) *A graph can be sampled from* $\mathcal{D}_{k,n_i}$ *in* $O_k(n_i^2)$ *time using* $O_k(n_i \log n_i)$ *random coins.*
(2) *For* $G \sim \mathcal{D}_{k,n_i}$, *writing* $d = (k+1)^2 + 1$, *we have*

$$\mathbb{P}\left(G \text{ is a } (n_i, n_i, k, n_i^{1/(k+1)}/2, d^{k+1}, 1)\text{-threshold graph}\right) \geq 1 - o(1).$$

Moreover, for every $n \in \mathbb{N}$, *there exists* $i \in \mathbb{N}$ *such that* $n \leq n_i \leq 2^k n$.

4. Zero sets of Random Polynomials

The aim of this section is to collect together the requisite tools from algebraic geometry that we require to prove Theorems 3.3 and 3.4. While we have attempted to keep the presentation self-contained for the most part, some of the arguments (unavoidably) assume some familiarity with algebraic geometry; for more background, we refer the reader to [48, 26].

A variety over an algebraically closed field $\overline{\mathbb{F}}$ is a set of the form

$$V = \{x \in \overline{\mathbb{F}}^k : f_1(x) = \cdots = f_t(x) = 0\}$$

for some collection of polynomials $f_1, \ldots, f_t \colon \overline{\mathbb{F}}^k \to \overline{\mathbb{F}}$; when we wish to make these polynomials explicit, we write $V(f_1, \ldots, f_t)$ for V. A variety is said to be *irreducible* if it cannot be written as the union of two proper subvarieties. The *dimension* $\dim V$ of a variety V is then the maximum integer d such that there exists a chain of irreducible subvarieties of V of the form

$$\emptyset \subsetneq V_0 \subsetneq V_1 \subsetneq V_2 \subsetneq \cdots \subsetneq V_d \subset V,$$

where V_0 consists of a single point. The *degree* of an irreducible variety of dimension d is the number of intersection points of the variety with d hyperplanes in general position, and for an arbitrary variety V, we define its degree $\deg V$ to be the sum of the degrees of its irreducible components.

We need Bézout's theorem in the following form; for a proof, see [26], for example.

Lemma 4.1. *For a collection of polynomials* $f_1, \ldots, f_k \colon \overline{\mathbb{F}}^k \to \overline{\mathbb{F}}$, *if the variety*

$$V = \{x \in \overline{\mathbb{F}}^k : f_1(x) = \cdots = f_k(x) = 0\}$$

has $\dim V = 0$, *then*

$$|V| \leq \prod_{i=1}^{k} \deg(f_i).$$

Moreover, for a collection of polynomials $f_1, \ldots, f_t \colon \overline{\mathbb{F}}^k \to \overline{\mathbb{F}}$, *the variety*

$$V = \{x \in \overline{\mathbb{F}}^k : f_1(x) = \cdots = f_t(x) = 0\}$$

has at most $\prod_{i=1}^{t} \deg(f_i)$ *irreducible components.*

In what follows, we let q be a prime power and work with polynomials over $\mathbb{F}_q$, where $\mathbb{F}_q$ is the finite field of order q. All varieties below are over $\mathbb{A}$, where $\mathbb{A} = \overline{\mathbb{F}}_q$ is the algebraic closure of $\mathbb{F}_q$, unless explicitly specified otherwise.

We let $\mathbb{F}_q[X_1, \dots, X_k]_{\leq d}$ be the subset of $\mathbb{F}_q[X_1, \dots, X_k]$ of polynomials in k variables of degree at most d, i.e., the set of linear combinations over $\mathbb{F}_q$ of monomials of the form $X_1^{a_1} \dots X_k^{a_k}$ with $\sum_{i=1}^k a_i \leq d$. Let us note that one may sample a uniformly random element of $\mathbb{F}_q[X_1, \dots, X_k]_{\leq d}$ by taking the coefficients of the monomials above to be independent random elements of $\mathbb{F}_q$.

The first lemma we state estimates the probability of a randomly chosen polynomial passing through each of m distinct points; see [10, 21] for similar statements.

Lemma 4.2. *Suppose that $q > \binom{m}{2}$ and $d \geq m - 1$. Let f be a uniformly random k-variate polynomial chosen from $\mathbb{F}_q[X_1, \dots, X_k]_{\leq d}$.*

(1) *If $x_1, \dots, x_m$ are m distinct points in $\mathbb{F}_q^k$, then*

$$\mathbb{P}\left(f(x_i) = 0 \text{ for all } i = 1, \dots, m\right) = q^{-m}.$$

(2) *If $x_1, \dots, x_m$ are m distinct points in $\mathbb{A}^k$, then*

$$\mathbb{P}\left(f(x_i) = 0 \text{ for all } i = 1, \dots, m\right) \leq q^{-m}.$$

Proof. We prove the first statement below, and later outline the proof of the second statement.

Let $x_i = (x_{i,1}, \dots, x_{i,k})$ for each $i = 1, \dots, m$. We choose elements $a_2, \dots, a_k \in \mathbb{F}_q$ such that $x_{i,1} + \sum_{j=2}^k a_j x_{i,j}$ is distinct for all $i = 1, \dots, m$. To see that this is possible, note that there are exactly $\binom{m}{2}$ equations

$$x_{i,1} + \sum_{j=2}^{k} a_j x_{i,j} = x_{i',1} + \sum_{j=2}^{k} a_j x_{i',j},$$

each with at most q^{k-2} solutions $(a_2, \dots, a_k)$. Therefore, since the total number of choices for $(a_2, \dots, a_k)$ is q^{k-1} and $q^{k-1} > q^{k-2}\binom{m}{2}$, we can make an appropriate choice.

We now consider $\mathbb{F}_q[Z_1, \dots, Z_k]_{\leq d}$, the set of polynomials of degree at most d in the variables $Z_1, \dots, Z_k$, where $Z_1 = X_1 + \sum_{j=2}^k a_j X_j$ and $Z_j = X_j$ for all $2 \leq j \leq k$. Since this change of variables is an invertible linear map, $\mathbb{F}_q[Z_1, \dots, Z_k]_{\leq d}$ is identical to $\mathbb{F}_q[X_1, \dots, X_k]_{\leq d}$. It will therefore suffice to show that a randomly chosen polynomial from $\mathbb{F}_q[Z_1, \dots, Z_k]_{\leq d}$ passes through all of the points $z_1, \dots, z_m$ corresponding to $x_1, \dots, x_m$ with probability exactly q^{-m}. For this, we will use the fact that, by our choice above, $z_{i,1} \neq z_{i',1}$ for any $1 \leq i < i' \leq m$.

For any f in $\mathbb{F}_q[Z_1, \dots, Z_k]_{\leq d}$, we may write $f = g + h$, where h contains all monomials of the form Z_1^j for $j = 0, 1, \dots, m-1$ and g contains all other monomials. For any fixed choice of g, there is, by Lagrange interpolation, exactly one choice of h with

coefficients in $\mathbb{F}_q$ such that $f(z_i) = 0$ for all $i = 1, \ldots, m$, namely, the unique polynomial of degree at most $m-1$ which takes the value $-g(z_i)$ at $z_{i,1}$ for all $i = 1, 2, \ldots, m$, where uniqueness follows from the fact that the $z_{i,1}$ are distinct. Since this is out of a total of q^m possibilities, we see that the probability of f passing through all of the z_i is exactly q^{-m}, as required.

For the second statement, we may argue identically, now working over $\mathbb{A}$ and noting that the unique polynomial of degree at most $m-1$ which takes the value $-g(z_i)$ at $z_{i,1}$ for all $i = 1, 2, \ldots, m$ may now have coefficients in $\mathbb{A}$ as opposed to $\mathbb{F}_q$, whence we get an inequality as opposed to the equality in the first statement. □

The next result we prove allows us to upper bound the size of the $\mathbb{F}_q$-variety cut out by multiple random polynomials.

Theorem 4.3. *Fix $t, k \in \mathbb{N}$ with $t \leq k$, and fix positive integers $d_1, \ldots, d_t \in \mathbb{N}$. Independently for each $i \in [t]$, sample f_i from $\mathbb{F}_q[X_1, \ldots, X_k]_{\leq d_i}$ uniformly at random. Then*

$$\mathbb{P}\left(\dim V(f_1, \ldots, f_t) > k - t\right) \leq C_t q^{-\min(d_1, \ldots, d_t)} \tag{1}$$

for some constant $C_t = C_t(d_1, \ldots, d_k) > 0$. In particular, if $t = k$, then

$$\mathbb{P}\left(\left|V(f_1, \ldots, f_k) \cap \mathbb{F}_q^k\right| > \prod_{i=1}^{k} d_i\right) \leq C q^{-\min(d_1, \ldots, d_k)}$$

for some constant $C = C(d_1, \ldots, d_k) > 0$.

Proof. For terminology not defined here, and standard facts about dimension that we call upon without proof, see the first and sixth chapter of [48].

To establish (1) it suffices show that

$$\mathbb{P}\left(\dim V(f_1, \ldots, f_{t-1}, f_t) > k - t \mid \dim V(f_1, \ldots, f_{t-1}) = k - t + 1\right) \leq q^{-d_t} \prod_{i=1}^{t-1} d_i \tag{2}$$

since (1) follows from (2) by induction on t.

Now, sample polynomials $f_1, \ldots, f_{t-1}$, and assume that the variety $U = V(f_1, \ldots, f_{t-1})$ is of dimension $d - t + 1$. By Lemma 4.1, U has at most $d_1 \cdots d_{t-1}$ components, which we name $U_1, \ldots, U_m$. Note that since $\dim U_i \leq \dim U = d - t + 1$, and U_i is intersection of $t-1$ hypersurfaces, each U_i is of dimension exactly $d - t + 1$. For each U_i, pick d_t distinct points $x_{i,1}, \ldots, x_{i,d_t}$ on U_i.

Since f_t is a random polynomial of degree d_t, from Lemma 4.2 we infer that

$$\mathbb{P}\left(U_i \subset V(f_t)\right) \leq \mathbb{P}\left(f_t(x_i, j) = 0 \text{ for all } j = 1, \ldots, d_t\right) \leq q^{-d_t}$$

for each $1 \le i \le m$. Hence, by the union bound

$$\mathbb{P}\left(\dim V(f_1,\dots,f_{t-1},f_t) > k-t\right) \le \sum_{i=1}^{m} \mathbb{P}\left(U_i \subset V(f_t)\right) \le q^{-d_t} \prod_{i=1}^{t-1} d_i.$$

proving (2), and hence (1).

If $t = k$, then

$$\begin{aligned}\mathbb{P}\left(\left|V(f_1,\dots,f_k)\cap \mathbb{F}_q^k\right| > \prod_{i=1}^{k} d_i\right) &\le \mathbb{P}\left(\left|V(f_1,\dots,f_k)\right| > \prod_{i=1}^{k} d_i\right)\\ &\le \mathbb{P}(\dim V(f_1,\dots,f_k) > 0)\\ &\le C_k q^{-\min(d_1,\dots,d_k)},\end{aligned}$$

where the first inequality is trivial, the second is a consequence of Lemma 4.1, i.e., Bézout's theorem, and the third is just (1) for $t = k$. □

Finally, we need a way to lower bound the size of the $\mathbb{F}_q$-variety cut out by multiple random polynomials, and the following result gives us what we need. While the arguments thus far have been mostly elementary, this result relies on some nontrivial rigidity properties of varieties over finite fields.

Theorem 4.4. *Fix positive integers $k, d_1, \dots, d_k \in \mathbb{N}$. Independently for each $i \in [k]$, sample f_i from $\mathbb{F}_q[X_1,\dots,X_k]_{\le d_i}$ uniformly at random. Then*

$$\mathbb{P}\left(\left|V(f_1,\dots,f_k)\cap \mathbb{F}_q^k\right| = \prod_{i=1}^{k} d_i\right) \ge \frac{1 - cq^{-1/2}}{\left(\prod_{i=1}^{k} d_i\right)!}$$

for some constant $c = c(d_1,\dots,d_k) > 0$.

Proof. For terminology not defined here, and standard results that we quote without proof, see the first three chapters of [48].

We set $r_i = \binom{k+d_i}{k}$ for $1 \le i \le k$, write $\vec{r} = (r_1,\dots,r_k)$ and $|\vec{r}|$ for $r_1 + \cdots + r_k$. For $1 \le i \le k$, we identify $\mathbb{A}^{r_i}$ with $\mathbb{A}[X]_{\le d_i}$, i.e., the space of polynomials in k variables of degree at most d_i with coefficients in $\mathbb{A}$. For brevity, we write $\mathbb{A}^{\vec{r}}$ in place of $\mathbb{A}^{r_1} \times \cdots \times \mathbb{A}^{r_k}$ (and $\mathbb{F}_q^{\vec{r}}$ in place of $\mathbb{F}_q^{r_1} \times \cdots \times \mathbb{F}_q^{r_k}$), and to distinguish the space where we evaluate our polynomials from these spaces of polynomials themselves, we set $Y = \mathbb{A}^k$.

Also, for $\mathbf{f} = (f_1,\dots,f_k) \in \mathbb{A}^{\vec{r}}$, we abbreviate the variety $V(f_1,\dots,f_k) \subset Y$ by $V(\mathbf{f})$. Now, set $t = d_1 \cdots d_k$ and call $\mathbf{f} \in \mathbb{F}_q^{\vec{r}}$ *good* if the variety $V(\mathbf{f})$ is zero-dimensional and has t distinct points that are defined over $\mathbb{F}_q$. In this language, note that we are trying to show, for large q, that roughly $1/t!$ of all the points in $\mathbb{F}_q^{\vec{r}}$ are good. To this end, we set

$$W = \{(\mathbf{f}, y_1,\dots,y_t) \in \mathbb{A}^{\vec{r}} \times Y^t : y_j \in V(\mathbf{f}) \text{ for all } j = 1,\dots,t\},$$

and deduce the result from the following claim.

Claim 4.5. *Suppose that* $(\mathbf{f}^*, \mathbf{y}^*)$ *is a simple point of* W *such that* $\mathbf{f}^*$ *is good and the coordinates of* $\mathbf{y}^* = (y_1^*, \dots, y_t^*)$ *are all distinct, and that for generic* $\mathbf{f}$*, the variety* $V(\mathbf{f})$ *is zero-dimensional of degree* t*. Then there are at least*

$$\frac{1 - cq^{-1/2}}{t!} q^{|\vec{r}|}$$

good points in $\mathbb{F}_q^{\vec{r}}$*, for some constant* $c = c(d_1, \dots, , d_k) > 0$*.*

Proof. Since $(\mathbf{f}^*, \mathbf{y}^*)$ is simple, the irreducible component of W containing it is unique. Let W_1 be the irreducible component of W containing $(\mathbf{f}^*, \mathbf{y}^*)$ and note that $\dim W_1 = \dim W$. Since the variety $V(\mathbf{f})$ is generically zero-dimensional of degree t, the fibres $W_{\mathbf{f}} = \{\mathbf{y} \in Y^t : (\mathbf{f}, \mathbf{y}) \in W\}$ of W are generically finite, whence we get $\dim W_1 = \dim W = |\vec{r}|$.

Let $\{W_1, \dots, W_m\}$ be the orbit of W_1 under the action of the Frobenius endomorphism. Since W is defined over $\mathbb{F}_q$, and hence invariant under this action, each such W_i is an irreducible component of W. Note that $(\mathbf{f}^*, \mathbf{y}^*) \in W_i$ for each $i \in [m]$, so if $m > 1$, this contradicts the uniqueness of the component containing $(\mathbf{f}^*, \mathbf{y}^*)$. Thus, $m = 1$, i.e., W_1 is defined over $\mathbb{F}_q$.

Since $(\mathbf{f}^*, \mathbf{y}^*) \in W_1$, the variety W_1 is not contained in

$$U = \bigcup_{i \neq j} \{(\mathbf{f}, \mathbf{y}) : y_i = y_j\}.$$

Hence, $W_1 \cap H$ is a proper subvariety of W_1, and therefore contains $O_{\deg W_1}(q^{|\vec{r}|-1})$ points by the Schwartz–Zippel lemma for varieties [12, Lemma 14]. Since W_1 is defined over $\mathbb{F}_q$ and is irreducible over $\mathbb{A}$, the Lang–Weil estimate [41] implies that W_1 contains at least

$$q^{\dim W_1} \left(1 - O_{\deg W_1}(q^{-1/2})\right)$$

points defined over $\mathbb{F}_q$. Hence, $W_1 \setminus H$ contains at least

$$q^{|\vec{r}|} \left(1 - O_{\deg W_1}(q^{-1/2}) - O_{\deg W_1}(q^{-1})\right) = q^{|\vec{r}|} \left(1 - O_{\deg W_1}(q^{-1/2})\right)$$

points defined over $\mathbb{F}_q$ as well. Since each good point $\mathbf{f}$ corresponds to exactly $t!$ points of $W_1 \setminus H$ defined over $\mathbb{F}_q$, the result follows. □

To finish, it remains to show that the simplicity and genericity hypotheses in Claim 4.5 are satisfied.

For $1 \leq i \leq k$, pick an arbitrary set $A_i \subset \mathbb{F}_q$ of size d_i. Define $\mathbf{f}^* = (f_1^*, \dots, f_k^*)$ by setting $f_i^* = \prod_{a \in A_i} (X_i - a)$ for $1 \leq i \leq k$ and let $\mathbf{y}^*$ be the vector of length $d_1 \cdots d_k$ whose coordinates are all the elements of $A_1 \times \cdots \times A_k$.

To prove that $(\mathbf{f}^*, \mathbf{y}^*)$ is simple, consider the tangent space of W at $(\mathbf{f}^*, \mathbf{y}^*)$, which we denote T_*W. An element $(\delta\mathbf{f}, \delta\mathbf{y}) \in \mathbb{A}^{\vec{r}} \times Y^t$ is in T_*W if it is a solution to the system of equations

$$\delta f_i(y_j^*) + \frac{\partial f_i}{\partial x_i}(y_j^*)(\delta y_j)_i = 0$$

for all $i \in [k]$ and $j \in [t]$. From these equations, it is clear that for every $\delta\mathbf{f} \in \mathbb{A}^{\vec{r}}$ there is a unique $\delta\mathbf{y}$ such that $(\delta\mathbf{f}, \delta\mathbf{y})$ is in the tangent space. Hence $\dim T_*W = \dim \mathbb{A}^{\vec{r}} = \dim W$, so it follows that $(\mathbf{f}^*, \mathbf{y}^*)$ is simple.

Next, the statement that for generic $\mathbf{f}$, the variety $V(\mathbf{f})$ (is zero-dimensional and) has at most $t = d_1 \cdots d_k$ points is the generalise Bézout's theorem. The construction of $(\mathbf{f}^*, \mathbf{y}^*)$ above shows that $V(\mathbf{f})$ generically has at least t points as well.

We have established the hypotheses under which Claim 4.5 applies; the result follows. □

5. Constructions of panchromatic graphs and threshold graphs

Here, we shall use the algebraic machinery we have developed to give the proofs of Theorems 3.3 and 3.4

5.1. **Panchromatic graphs.** First, we give the construction of panchromatic graphs using random polynomials.

Proof of Theorem 3.3. Let q be a prime power, and let $\mathbb{F}_q$ be the finite field of order q. We shall assume that $k \in \mathbb{N}$ and $\lambda > 1$ are fixed, and that q is sufficiently large as a function of k. Finally, let us fix $d = k^2 + 2$, $D = \lambda d$ and $n = q^k$. In the rest of the proof, all asymptotic notation will be in the limit of $q \to \infty$.

We shall construct a panchromatic graph between two sets A and B as follows. First, choose polynomials $w_1, \ldots, w_k \in \mathbb{F}_q[X_1, \ldots, X_k]_{\leq D}$ independently and uniformly at random. Next, for $i \in [k]$, let A_i be a set of n vertices each associated with a polynomial $w_i + p$, where $p \in \mathbb{F}_q[X_1, \ldots, X_k]_{\leq d}$ is chosen uniformly at random and independently for each vertex; note here that the distribution of the resulting polynomial $w_i + p$ is also uniform on $\mathbb{F}_q[X_1, \ldots, X_k]_{\leq D}$. Let A be the disjoint union $\sqcup_{i=1}^k A_i$, and set $B = \mathbb{F}_q^k$, so that $|A| = kq^k$ and $|B| = q^k$. Finally, let G be the (random) graph between A and B where a polynomial $f \in A$ is joined to a point $x \in B$ if $f(x) = 0$. We shall show that G has the requisite properties with probability at least $(4(D^k)!)^{-1}$.

First, we count the number of k-sets $U = \{f_1, f_2, \ldots, f_k\}$ with $f_i \in A_i$ for which the size of the common neighbourhood $N(U)$ in G exceeds D^k. For such a set U, observe that $N(U)$ is the set of $\mathbb{F}_q$-solutions of k polynomials from $\mathbb{F}_q[X_1, \ldots, X_k]_{\leq D}$ chosen independently and uniformly at random, so by Theorem 4.3, we have

$$\mathbb{P}(|N(U)| > D^k) = O(q^{-D}).$$

Writing Z_1 for the number of such k-sets, we get

$$\mathbb{E}[Z_1] = O\left(n^k q^{-D}\right) = O\left(q^{k^2} q^{-\lambda(k^2+2)}\right) = O(q^{-2}) \leq 1/q. \tag{3}$$

Next, we count the number of k-sets $U = \{f_1, f_2, ..., f_k\}$ with $f_i \in A_i$ for $i \in [k]$ for which size of the common neighbourhood $N(U)$ in G is exactly D^k. As above, for such a set U, observe that $|N(U)|$ is distributed as the number of $\mathbb{F}_q$-solutions of k polynomials from $\mathbb{F}_q[X_1, \dots X_k]_{\leq D}$ chosen independently and uniformly at random, so by Theorem 4.4, we have

$$\mathbb{P}(|N(U)| = D^k) \geq (2(D^k)!)^{-1}.$$

Writing Z_2 for the number of such k-sets, we get

$$\mathbb{E}[Z_2] \geq n^k (2(D^k)!)^{-1}. \tag{4}$$

Finally, we count the number of k-sets $U \subset A$ with $A_i \cap U$ being empty for some $i \in [k]$ for which the size of the common neighbourhood $N(U)$ in G exceeds $dD^{k-1} = D^k/\lambda$. For such a set U, observe that $|N(U)|$ is distributed as the number of $\mathbb{F}_q$-solutions of a collection of k random polynomials. To understand the distribution of this random collection of polynomials, for each $i \in [k]$ for which $U \cap A_i \neq \emptyset$, we pick one element $U \cap A_i$ and subtract that from every other element of $U \cap A_i$; observe that by doing so, we get a set $\{g_1, \dots, g_k\}$ of independent random polynomials, each uniform over either $\mathbb{F}_q[X_1, \dots X_k]_{\leq d}$ or $\mathbb{F}_q[X_1, \dots, X_k]_{\leq D}$, and at least one of which is uniform over $\mathbb{F}_q[X_1, \dots X_k]_{\leq d}$. Since $|N(U)|$ is then number of $\mathbb{F}_q$-solutions of $\{g_1, \dots, g_k\}$, we deduce from Theorem 4.3 that

$$\mathbb{P}(|N(U)| > dD^{k-1}) = O(q^{-d}).$$

Writing Z_3 for the number of such k-sets, we get

$$\mathbb{E}[Z_3] = O\left((kn)^k q^{-d}\right) = O\left(q^{k^2} q^{-k^2-2}\right) = O(q^{-2}) \leq 1/q. \tag{5}$$

We combine (3), (4) and (5) as follows. Clearly, $\mathbb{E}[Z_1 + Z_3] = o(1)$, so by Markov's inequality, both Z_1 and Z_2 are zero with probability $1 - o(1)$. Finally, since Z_2 is trivially at most n^k and $\mathbb{E}[Z_2] \geq n^k (2(D^k)!)^{-1}$, it is easily checked that

$$\mathbb{P}\left(Z_2 \geq n^k (4(D^k)!)^{-1}\right) \geq (2(D^k)!)^{-1}.$$

By the union bound, we see that G is a $(n, n, k, D^k, D^k/\lambda, (4(D^k)!)^{-1})$-panchromatic graph with probability at least $(4(D^k)!)^{-1}$, completing the proof. □

5.2. **Threshold graphs.** Next, we give the construction of threshold graphs, once again using random polynomials.

Proof of Theorem 3.4. As before, let q be a prime power, and let $\mathbb{F}_q$ be the finite field of order q. We shall assume that $k \in \mathbb{N}$ is fixed, and that q is sufficiently large as a function of k. Let $d = (k+1)^2 + 1$ and $n = q^{k+1}$. We shall construct a threshold graph between two sets A and B both of size q^{k+1}. In the rest of the proof, all asymptotic notation will be in the limit of $q \to \infty$.

We construct A by sampling q^{k+1} random polynomials from $\mathbb{F}_q[X_1, \ldots, X_{k+1}]_{\leq d}$ uniformly and independently, set $B = \mathbb{F}_q^{k+1}$, and define a (random) bipartite graph G between A and B by joining $f \in A$ to $x \in B$ if $f(x) = 0$. We shall show that G has the requisite properties with probability $1 - o(1)$.

First, we consider the soundness properties of G. Fix a set $U \subset A$ of size $k+1$. The size of its common neighbourhood $N(U)$ in G is distributed as the number of $\mathbb{F}_q$-solutions of $k+1$ polynomials from $\mathbb{F}_q[X_1, \ldots X_{k+1}]_{\leq d}$ chosen independently and uniformly at random, so by Theorem 4.3, we have

$$\mathbb{P}(|N(U)| > d^{k+1}) = O(q^{-d}).$$

Call a set of $k+1$ vertices of G *bad* if their common neighbourhood has more than d^{k+1} vertices. The number Z_1 of bad $(k+1)$-sets then satisfies

$$\mathbb{E}[Z_1] = O\left(\binom{n}{k+1} q^{-d}\right) = O\left(\binom{q^{k+1}}{k+1} q^{-(k+1)^2-1}\right) = O(q^{-1}) = o(1). \tag{6}$$

Next, we turn to the completeness properties of G. Fix a set $U \subset A$ of size k. For $v \in B$, put $I(v) = 1$ if $f(v) = 0$ for all $f \in U$, and $I(v) = 0$ if $f(v) \neq 0$ for some $f \in U$. For $1 \leq m \leq d$ and distinct $v_1, \ldots, v_m \in B$, we have

$$\mathbb{P}\left(I(v_1) \cdots I(v_m) = 1\right) = \prod_{f \in U} \mathbb{P}\left(f(v_j) = 0 \text{ for all } j = 1, \ldots, m\right) = q^{-mk},$$

where the first equality is by independence, and the second is by Lemma 4.2. Small moments of the random variable $Z = |N(U)|$ are now easily computed: for $1 \leq m \leq d$, we have

$$\begin{aligned}
\mathbb{E}\left[Z^m\right] &= \mathbb{E}\left[\left(\sum_{v \in B} I(v)\right)^m\right] \\
&= \mathbb{E}\left[\sum_{v_1, \ldots, v_m \in B} I(v_1) \cdots I(v_m)\right] \\
&= \sum_{v_1, \ldots, v_m \in B} \mathbb{E}[I(v_1) \cdots I(v_m)] \\
&= \sum_{r=1}^{m} \binom{q^{k+1}}{r} M_{r,m} q^{-rk},
\end{aligned} \tag{7}$$

where $M_{r,m}$ is the number of surjective functions from an m-element set onto an r-element set. Combining (7) with standard identities for the Stirling numbers of the second kind, we get that

$$\mathbb{E}\left[(Z-\mathbb{E}[Z])^d\right] = O(q) \text{ and } \mathbb{E}[Z] = q,$$

whence it follows that

$$\mathbb{P}(Z < q/2) \leq \mathbb{P}(|Z-\mathbb{E}[Z]| < q/2) \leq \frac{\mathbb{E}\left[(Z-\mathbb{E}[Z])^d\right]}{(q/2)^d} = O\left(q^{1-d}\right).$$

Call a set of k vertices of G *bad* if their common neighbourhood has fewer than $q/2$ vertices. The number Z_2 of bad k-sets then satisfies

$$\mathbb{E}[Z_2] = O\left(\binom{n}{k} q^{1-d}\right) = O\left(\binom{q^{k+1}}{k} q^{-(k+1)^2}\right) = O(q^{-1-k}) = o(1). \tag{8}$$

Combining (6) and (8), we see that

$$\mathbb{E}[Z_1 + Z_2] = o(1);$$

it follows from Markov's inequality that $Z_1 + Z_2 = 0$ (and hence $Z_1 = Z_2 = 0$) with probability $1-o(1)$, so G is a $(q^{k+1}, q^{k+1}, k, q/2, d^{k+1}, 1)$-threshold graph with probability $1-o(1)$, completing the proof. □

We remark in passing that a quantitatively weaker version of Theorem 3.4 that utilises less randomness can alternately be proved by building a bipartite graph between two copies of $\mathbb{F}_q^{k+1}$ by choosing a single random polynomial f in $2k+2$ variables of degree $2k^2$ and joining pairs of points $x, y \in \mathbb{F}_q^{k+1}$ for which $f(x,y) = 0$; however, the analysis of this construction relies on more machinery (and in particular, on the Lang–Weil estimate), and furthermore, yields ineffective parameter dependencies.

6. Time lower bounds for k-SetIntersection

Armed with our graph constructions, we are now ready to prove the formal versions of Theorems 1.5 and 1.6. However, before we jump into these proofs, we need to lay out the PGC framework.

6.1. Panchromatic graph composition.

Given a panchromatic problem and a panchromatic graph, we would like to compose them in some way such that we obtain a *monochromatic* version of the panchromatic problem having the property that every optimal solution of the monochromatic version can be traced back to an optimal solution of the panchromatic version. When we talk about the PGC technique, we use it as an umbrella term for this composition operation; typically, the composition is a product operation, as below for the k-SetIntersection problem.

Theorem 6.1 (Panchromatic graph composition)**.** *There is an algorithm that given as input*

(1) *an instance* $\Gamma(\mathcal{C}_1, \ldots, \mathcal{C}_k, r, s)$ *of panchromatic* k-SetIntersection *over universe* $\mathcal{U}$ *with monochromatic number* $z = z(\Gamma)$*, and*
(2) *an* (n,m,k,t,w,p)*-panchromatic graph* $H(A = A_1 \sqcup \cdots \sqcup A_k, B)$ *with* $|A_j| = |\mathcal{C}_j|$ *for all* $j \in [k]$*,*

then outputs an instance $\Gamma'(\mathcal{C}', rt, \max(st, zw))$ *of* k-SetIntersection *over universe* $\mathcal{U}'$ *such that the following hold.*

Size: $|\mathcal{C}'| = |\mathcal{C}_1| + \cdots + |\mathcal{C}_k|$ *and* $|\mathcal{U}'| = |\mathcal{U}||B|$.
Completeness: *If there exists a* k*-tuple of sets* $(S_{i_1}, \ldots, S_{i_k})$ *in* $\mathcal{C}_1 \times \cdots \times \mathcal{C}_k$ *such that*

$$\left| \bigcap_{j \in [k]} S_{i_j} \right| \geq r,$$

then with probability p*, there exist* k *sets* $S'_{i_1}, \ldots, S'_{i_k}$ *in* $\mathcal{C}'$ *such that*

$$\left| \bigcap_{j \in [k]} S'_{i_j} \right| \geq rt.$$

Soundness: *If for every* k*-tuple of sets* $(S_{i_1}, \ldots, S_{i_k})$ *in* $\mathcal{C}_1 \times \cdots \times \mathcal{C}_k$ *we have*

$$\left| \bigcap_{j \in [k]} S_{i_j} \right| \leq s,$$

then for every k *sets* $S'_{i_1}, \ldots, S'_{i_k}$ *in* $\mathcal{C}'$ *we have*

$$\left| \bigcap_{j \in [k]} S'_{i_j} \right| \leq \max(st, zw).$$

Running Time: *The reduction runs in* $\tilde{O}(|\mathcal{C}'||\mathcal{U}'|)$ *time.*

Proof. We set $\mathcal{U}' = \mathcal{U} \times B$, and for every $j \in [k]$, take $\pi_j \colon \mathcal{C}_j \to A_j$ to be a uniformly random one-to-one mapping. Additionally, for every $j \in [k]$, let $\zeta_j : \mathcal{C}_j \to 2^{\mathcal{U}'}$ be a function which maps a set in $\mathcal{C}_j$ to a subset of $\mathcal{U}'$ in $\mathcal{C}'$ in the following way: for every $S \in \mathcal{C}_j$, we include $(u, b) \in \mathcal{U} \times B$ in $\zeta_j(S)$ if and only if $u \in S$ and $(\pi_j(S), b) \in E(H)$.

First, let us suppose that there exists a k-tuple of sets $(S_{i_1}, \ldots, S_{i_k})$ in $\mathcal{C}_1 \times \cdots \times \mathcal{C}_k$ such that

$$\left| \bigcap_{j \in [k]} S_{i_j} \right| \geq r.$$

Consider the k-tuple of vertices $(\pi_1(S_{i_1}), \ldots, \pi_k(S_{i_k}))$ in $A_1 \times \cdots \times A_k$. Since $\pi_1, \ldots, \pi_k$ were picked independently and uniformly at random, the aforementioned k-tuple of vertices in A is uniformly random, and thus from the completeness of the panchromatic graph, there exists, with probability p, a set of t vertices in B, denoted by B', which are all common neighbours of $(\pi_1(S_{i_1}), \ldots, \pi_k(S_{i_k}))$. For $u \in \cap_{j\in[k]} S_{i_j}$ and $b \in B'$, we clearly have $(u, b) \in \zeta_j(S_{i_j})$, or in other words,

$$\left|\bigcap_{j\in[k]} \zeta_j(S_{i_j})\right| \geq r|B'| \geq rt.$$

Next, let us suppose that for every k-tuple of sets $(S_{i_1}, \ldots, S_{i_k})$ in $\mathcal{C}_1 \times \cdots \times \mathcal{C}_k$, we have

$$\left|\bigcap_{j\in[k]} S_{i_j}\right| \leq s.$$

For the sake of contradiction, let there be k sets $S'_{i_1}, \ldots, S'_{i_k}$ in $\mathcal{C}'$ such that

$$\left|\bigcap_{j\in[k]} S'_{i_j}\right| > \max(st, zw).$$

By our construction of $\mathcal{C}'$, there exists $\ell_j \in [k]$ and $S_{i_j} \in \mathcal{C}_{\ell_j}$ such that $\zeta_{\ell_j}(S_{i_j}) = S'_{i_j}$ for every $j \in [k]$. Let $D = \{\ell_j \mid j \in [k]\}$.

First, suppose that $|D| = k$. From the completeness of the panchromatic graph, we see that the set of common neighbours of $(\pi_{\ell_1}(S_{i_1}), \ldots, \pi_{\ell_k}(S_{i_k}))$ in B, denoted by B', is of size at most t. It follows that

$$\left|\bigcap_{j\in[k]} S'_{i_j}\right| \leq \left|\bigcap_{j\in[k]} S_{i_j}\right| \cdot |B'| \leq st,$$

which is a contradiction. Next, suppose that $|D| < k$. By the soundness of the panchromatic graph, the set of common neighbours of $(\pi_{\ell_1}(S_{i_1}), \ldots, \pi_{\ell_k}(S_{i_k}))$ in B, denoted by B', is of size at most w. It follows that

$$\left|\bigcap_{j\in[k]} S'_{i_j}\right| \leq \left|\bigcap_{j\in[k]} S_{i_j}\right| \cdot |B'| \leq zw,$$

where z is the monochromatic number of Γ, again leading to a contradiction.

Finally, from the construction of Γ', the claim on the runtime follows immediately; this completes the proof. □

6.2. **SETH-based Time Lower Bound.** In this subsection, we prove the following result.

Theorem 6.2. *Let $F : \mathbb{N} \to \mathbb{N}$ be any computable increasing function. Assuming* SETH, *for every $\varepsilon > 0$ and integer $k > 1$, no randomised $O(n^{k(1-\varepsilon)})$-time algorithm can decide every instance $\Gamma(\mathcal{C}, r, r/F(k))$ of k-*SetIntersection *over universe $[n^{1+o(1)}]$ with $|\mathcal{C}| = n$.*

Our proof builds on the following SETH based lower bound for gap k-MaxCover proved in [36].

Theorem 6.3. *Let $F : \mathbb{N} \to \mathbb{N}$ be any computable increasing function. Assuming* SETH, *for every $\varepsilon > 0$ and integer $k > 1$, no randomised $O(n^{k(1-\varepsilon)})$-time algorithm can decide every instance $\Gamma(G = (V \sqcup W, E), 1, 1/F(k))$ of* Unique k-MaxCover, *even when*

(1) *$V = V_1 \sqcup \cdots \sqcup V_k$ with $|V_j| = n$ for all $j \in [k]$, and*
(2) *$W = W_1 \sqcup \cdots \sqcup W_\ell$ with $|W_i| = O_{k,\varepsilon}(1)$ for all $j \in [\ell]$, where $\ell = (\log n)^{O_k(1)}$.* □

We are now ready for the proof of Theorem 6.2.

Proof of Theorem 6.2. The case $k = 2$ was already proved in [37], so we assume that $k > 2$. Fix $F : \mathbb{N} \to \mathbb{N}$ as in the theorem, and suppose, for some fixed $\varepsilon > 0$, that there is a randomised $O(n^{k(1-\varepsilon)})$-time algorithm that can decide every instance $\Gamma(\mathcal{C}, r, r/F(k))$ of k-SetIntersection over universe $[n^{1+o(1)}]$ with $|\mathcal{C}| = n$. We claim that this algorithm can be used to solve every instance $\Gamma'(G = (V \sqcup W, E), 1, 1/F(k))$ of k-MaxCover in time $O(n^{k(1-\varepsilon/2)})$ where

(1) $V = V_1 \sqcup \cdots \sqcup V_k$ with $|V_j| = n$ for all $j \in [k]$, and
(2) $W = W_1 \sqcup \cdots \sqcup W_\ell$ with $|W_i| = O_{k,\varepsilon}(1)$ for all $j \in [\ell]$, where $\ell = (\log n)^{O_k(1)}$,

contradicting Theorem 6.3.

Fix an instance $\Gamma'(G = (V \sqcup W, E), 1, 1/F(k))$ of k-MaxCover as above. By applying Proposition 2.3 to Γ', we obtain an instance

$$\Gamma''(\mathcal{C}_1, \ldots, \mathcal{C}_k, \ell, \ell/F(k))$$

of panchromatic k-SetIntersection over a universe of size $O_\varepsilon((\log n)^{O_k(1)})$ with monochromatic number bounded above by $c_{k,\varepsilon}\ell$ for some $c_{k,\varepsilon}$ depending only on k and ε.

Put $m = \sqrt{n}$, $\lambda = c_{k,\varepsilon}F(k)$, $D = \lambda(k^2 + 2)$, and choose $i^* \in \mathbb{N}$ in Theorem 3.3 such that $m \leq n_{i^*} \leq 2^k m$.

We sample $w = \widetilde{\Omega}(4(D^k)!)$ many graphs $G_1, \ldots, G_w$ from $\mathcal{D}_{k,\lambda,n_{i^*}}$ in time $O_k(n)$. By Theorem 3.3, we know that with high probability one of these graphs, say G^*, is a panchromatic graph with parametrisation

$$(n_{i^*}, n_{i^*}, k, D^k, D^k/\lambda, (4(D^k)!)^{-1}),$$

and we can find it in time $wn_{i^*}^{k+1} = O_k(n^{k/2+1})$. We arbitrarily delete $n_{i^*} - m$ vertices in each colour class of G^* to make it a panchromatic graph with parametrisation

$$(m, n_{i^*}, k, D^k, D^k/\lambda, (4(D^k)!)^{-1}).$$

For every $i \in [k]$, arbitrarily equipartition $\mathcal{C}_i$ into $\mathcal{C}_i^1, \ldots, \mathcal{C}_i^m$. For every $(t_1, \ldots, t_k) \in [m]^k$, we consider the instance

$$\Gamma''_{(t_1,\ldots,t_k)}(\mathcal{C}_1^{t_1}, \ldots, \mathcal{C}_k^{t_k}, \ell, \ell/F(k))$$

of panchromatic k-SetIntersection obtained from Γ'' above; each such instance is also clearly defined over a universe of size $O_\varepsilon((\log n)^{O_k(1)})$ and has monochromatic number bounded above by $c_{k,\varepsilon}\ell$. For each such $(t_1, \ldots, t_k) \in [m]^k$, we apply Theorem 6.1 to $\Gamma''_{(t_1,\ldots,t_k)}$ using G^* to thus obtain an instance

$$\Gamma_{(t_1,\ldots,t_k)}(\mathcal{C}, \ell D^k, (\ell D^k)/F(k))$$

of k-SetIntersection over a universe $\mathcal{U}$, with $|\mathcal{U}| = m(\log n)^{O_k(1)}$ and $|\mathcal{C}| = mk$, in time $\tilde{O}(n^{1+o(1)})$.

In summary, if Γ' is complete, then there exists $(t_1, \ldots, t_k) \in [m]^k$ such that $\Gamma''_{(t_1,\ldots,t_k)}$ is also complete, and consequently, as is $\Gamma_{(t_1,\ldots,t_k)}$. On the other hand, if Γ' is sound, then for every $(t_1, \ldots, t_k) \in [m]^k$, $\Gamma''_{(t_1,\ldots,t_k)}$ is also sound, and consequently, as is $\Gamma_{(t_1,\ldots,t_k)}$. Thus, this reduction shows that we can decide Γ' by testing all the instances $\Gamma_{(t_1,\ldots,t_k)}$. The total runtime of the resulting algorithm is

$$O\left(n^{k/2} \cdot \left(n^{k(1-\varepsilon)/2} + n^{1+o(1)}\right) + n^{k/2+1}\right) = O\left(n^{k(1-\frac{\varepsilon}{2})}\right),$$

yielding the desired contradiction. □

6.3. **ETH-based Time Lower Bound.** In this subsection, we prove the following result.

Theorem 6.4. *Let $F : \mathbb{N} \to \mathbb{N}$ be any computable increasing function. Assuming* ETH, *for sufficiently large $k \in \mathbb{N}$, no randomised $n^{o(k)}$-time algorithm can decide an instance $\Gamma(\mathcal{C}, r, r/F(k))$ of k-*SetIntersection *over universe $[n^{1+o(1)}]$ with $|\mathcal{C}| = n$.*

Our proof builds on the following ETH based lower bound for gap k-MaxCover proved in [36].

Theorem 6.5. *Let $F : \mathbb{N} \to \mathbb{N}$ be any computable increasing function. Assuming* ETH, *for sufficiently large $k \in \mathbb{N}$, no randomised $n^{o(k)}$-time algorithm can decide an instance $\Gamma(G = (V \sqcup W, E), 1, 1/F(k))$ of* Unique k-MaxCover, *even when*

(1) *$V = V_1 \sqcup \cdots \sqcup V_k$ with $|V_j| = n$ for all $j \in [k]$, and*
(2) *$W = W_1 \sqcup \cdots \sqcup W_\ell$ with $|W_i| = O_k(1)$ for all $j \in [\ell]$, where $\ell = (\log n)^{O_k(1)}$.* □

We are now ready for the proof of Theorem 6.4.

Proof of Theorem 6.4. Fix $F : \mathbb{N} \to \mathbb{N}$ as in the theorem, and suppose that there is a randomised $n^{o(k)}$-time algorithm that can decide every instance $\Gamma(\mathcal{C}, r, r/F(k))$ of k-SetIntersection over universe $[n^{1+o(1)}]$ with $|\mathcal{C}| = n$. Notice that such an algorithm can immediately yields a search algorithm that finds a witness in the YES case by making nk calls to the decision algorithm. We claim that this search algorithm can be used to solve every instance $\Gamma'(G = (V \sqcup W, E), 1, 1/F(k))$ of k-MaxCover in time $O(n^{o(k)})$ where

(1) $V = V_1 \sqcup \cdots \sqcup V_k$ with $|V_j| = n$ for all $j \in [k]$, and
(2) $W = W_1 \sqcup \cdots \sqcup W_\ell$ with $|W_i| = O_{k,\varepsilon}(1)$ for all $i \in [\ell]$, where $\ell = (\log n)^{O_k(1)}$,

contradicting Theorem 6.5. The proof mirrors that of Theorem 6.2, so stop with an outline.

Fix an instance $\Gamma'(G = (V \sqcup W, E), 1, 1/F(k))$ of k-MaxCover as above. By applying Proposition 2.3 to Γ', we obtain an instance

$$\Gamma''(\mathcal{C}_1, \ldots, \mathcal{C}_k, \ell, \ell/F(k))$$

of panchromatic k-SetIntersection over a universe of size $(\log n)^{O_k(1)}$ with monochromatic number bounded above by $c_k\ell$ for some constant c_k depending only on k.

Put $\lambda = c_k F(k)$, $D = \lambda(k^2 + 2)$, and choose $i^* \in \mathbb{N}$ in Theorem 3.3 such that $n \leq n_{i^*} \leq 2^k n$. We sample $w = \widetilde{\Omega}(4(D^k)!)$ many graphs $G_1, \ldots, G_w$ from $\mathcal{D}_{k,\lambda,n_{i^*}}$ in time $O_k(n^2)$. By Theorem 3.3, we know that with high probability, one of these graphs is a panchromatic graph with parametrisation

$$(n_{i^*}, n_{i^*}, k, D^k, D^k/\lambda, (4(D^k)!)^{-1});$$

In each of these graphs, we arbitrarily delete $n_{i^*} - n$ vertices in each colour class so that one of these graphs is now guaranteed to be panchromatic (with high probability) with parametrisation

$$(n, n_{i^*}, k, D^k, D^k/\lambda, (4(D^k)!)^{-1}).$$

For each G_i, we apply Theorem 6.1 to Γ'' using G_i. If G_i is a panchromatic graph with the right parametrisation, then we obtain an instance $\Gamma(\mathcal{C}, \ell D^k, (\ell D^k)/F(k))$ of k-SetIntersection over a universe $\mathcal{U}$, with $|\mathcal{U}| = n(\log n)^{O_k(1)}$ and $|\mathcal{C}| = nk$, in time $O(n^{2+o(1)})$, and solving this instance Γ allows us to solve Γ'. On the other hand, if G_i is not a suitable panchromatic graph, then we still obtain some instance Γ of k-SetIntersection, and while our hypothetical search algorithm would output a witness if we are in the YES case of Γ, this would not yield any meaningful solution to Γ', and consequently, we can discard such instances Γ. □

7. Conclusion

We finish with a discussion of some open problems that we think are particularly deserving of attention.

Closest pair. In [37], the authors constructed panchromatic graphs (for $k = 2$) with various parameters, and used these to prove conditional hardness of approximation results for the closest pair problem, where we are given a set of n points in $\mathbb{R}^d$ and would like to find the closest pair of points in the ℓ_p-metric. Using these, the authors showed that assuming SETH, no algorithm running in $n^{1.5-\delta}$ time can approximate the closest pair problem to factor $(1+\varepsilon)$, where $\delta = \delta(\varepsilon) > 0$. If there exist $(n, m, 2, t, t(1-\varepsilon), 1/n^{o(1)})$-panchromatic graphs with $m = n^{o(1)}$ and $t = \Omega(m)$, then this would prove subquadratic time inapproximability for the closest pair problem; do such panchromatic graphs exist?

Biclique. Using a more intricate composition technique and weaker objects than our threshold graphs, Lin [42] showed that the k-Biclique problem is W[1]-hard; in the k-Biclique problem, we are given a balanced bipartite graph on n vertices, and our goal is to determine if it contains a $K_{k,k}$. Lin further showed that under ETH, no $n^{o(\sqrt{k})}$ time algorithm can decide k-Biclique. If there exist $(n, n, k, t, t-1, 1/n)$-threshold graphs with $t = O(k)$, then we would obtain a tight time lower bound for k-Biclique under ETH; do such threshold graphs exist?

Derandomisation. In this paper, we provide distributions from which we can efficiently sample panchromatic and threshold graphs. Since these graphs seem to be widely useful building blocks, a natural derandomisation question presents itself: can we find explicit constructions of panchromatic and threshold graphs?

Applications of threshold graphs. Norm-graphs have various applications in theoretical computer science such as proving lower bounds for span-programs [8, 27], rectifier networks [34] and circuit lower bounds [35]. In each of these cases, our threshold graph constructions match the lower bounds obtained by using norm-graphs. Are there applications where the stronger completeness property of our threshold graph constructions can prove useful? Additionally, can our constructions yield (semi-explicit) rigid matrices? If yes, this would serve as a very interesting follow-up to [29].

Acknowledgements

The first author was supported in part by NSF CAREER grant DMS-1555149, the second author was financially supported by Subhash Khot's Simons Investigator Award and by Grant 825876 from the Simons Foundation, and the third author was supported by NSF grants CCF-1814409 and DMS-1800521.

References

1. *Handbook of theoretical computer science. Vol. A*, Elsevier Science Publishers, B.V., Amsterdam; MIT Press, Cambridge, MA, 1990, Algorithms and complexity. 11

2. N. Alon, *Problems and results in extremal combinatorics–I*, Discret. Math. **273** (2003), 31–53. 1
3. _____, *Problems and results in extremal combinatorics – II*, Discret. Math. **308** (2008), 4460–4472. 1
4. _____, *Problems and results in extremal combinatorics—III*, J. Comb. **7** (2016), 233–256. 1
5. N. Alon, O. Goldreich, J. Håstad, and R. Peralta, *Simple constructions of almost k-wise independent random variables*, Random Structures Algorithms **3** (1992), 289–304. 2
6. N. Alon, R. Yuster, and U. Zwick, *Color-coding*, J. Assoc. Comput. Mach. **42** (1995), 844–856. 4, 8
7. A. E. Ashikhmin, A. Barg, and S. G. Vladut, *Linear codes with exponentially many light vectors*, J. Comb. Theory, Ser. A **96** (2001), 396–399. 2
8. L. Babai, A. Gál, J. Kollár, L. Rónyai, T. Szabó, and A. Wigderson, *Extremal bipartite graphs and superpolynomial lower bounds for monotone span programs*, Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing, ACM, New York, 1996, pp. 603–611. 2, 6, 29
9. A. Bhattacharyya, É. Bonnet, L. Egri, S. Ghoshal, Karthik C. S., B. Lin, P. Manurangsi, and D. Marx, *Parameterized intractability of even set and shortest vector problem*, **68** (2021), A16, 40pp. 3, 11
10. B. Bukh, *Random algebraic construction of extremal graphs*, Bull. Lond. Math. Soc. **47** (2015), 939–945. 8, 9, 16
11. B. Bukh and D. Conlon, *Rational exponents in extremal graph theory*, J. Eur. Math. Soc. (JEMS) **20** (2018), 1747–1757. 9
12. B. Bukh and J. Tsimerman, *Sum-product estimates for rational functions*, Proc. Lond. Math. Soc. **104** (2012), 1–26. 19
13. M. L. Carmosino, J. Gao, R. Impagliazzo, I. Mihajlin, R. Paturi, and S. Schneider, *Nondeterministic extensions of the strong exponential time hypothesis and consequences for non-reducibility*, ITCS'16—Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science, ACM, New York, 2016, pp. 261–270. 11
14. P. Chalermsook, M. Cygan, G. Kortsarz, B. Laekhanukit, P. Manurangsi, D. Nanongkai, and L. Trevisan, *From gap-exponential time hypothesis to fixed parameter tractable inapproximability: clique, dominating set, and more*, SIAM J. Comput. **49** (2020), 772–810. 11
15. A. K. Chandra, M. L. Furst, and R. J. Lipton, *Multi-party protocols*, Proceedings of the 15th Annual ACM Symposium on Theory of Computing, ACM, New York, 1983, pp. 94–99. 2
16. E. Chattopadhyay and D. Zuckerman, *Explicit two-source extractors and resilient functions*, Ann. of Math. **189** (2019), 653–705. 2

17. Y. Chen and B. Lin, *The constant inapproximability of the parameterized dominating set problem*, SIAM J. Comput. **48** (2019), 513–533. 3
18. R. Clifford and A. Popa, *Maximum subset intersection*, Inf. Process. Lett. **111** (2011), 323–325. 7
19. A. Cobham, *The intrinsic computational difficulty of functions*, Logic, Methodology and Philos. Sci., North-Holland, Amsterdam, 1965, pp. 24–30. 2
20. G. Cohen, *Two-source dispersers for polylogarithmic entropy and improved Ramsey graphs*, SIAM J. Comput. **50** (2021), 30–67. 2
21. D. Conlon, *Some remarks on the Zarankiewicz problem*, Preprint, arXiv:2007.12816. 9, 16
22. M. Cygan, F. V. Fomin, Ł. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, and S. Saurabh, *Parameterized algorithms*, Springer, Cham, 2015. 2
23. R. David, C. S. Karthik, and B. Laekhanukit, *On the complexity of closest pair via polar-pair of point-sets*, SIAM J. Discrete Math. **33** (2019), 509–527. 4, 10
24. R. G. Downey and M. R. Fellows, *Fundamentals of parameterized complexity*, Texts in Computer Science, Springer, London, 2013. 2
25. J. Edmonds, *Paths, trees, and flowers*, Canadian J. Math. **17** (1965), 449–467. 2
26. W. Fulton, *Introduction to intersection theory in algebraic geometry*, CBMS Regional Conference Series in Mathematics, vol. 54, American Mathematical Society, Providence, RI, 1984. 15
27. A. Gál, *A characterization of span program size and improved lower bounds for monotone span programs*, Comput. Complex. **10** (2001), 277–296. 29
28. A. Ganor, G. Kol, and R. Raz, *Exponential separation of information and communication for boolean functions*, J. ACM **63** (2016), 1–31. 2
29. O. Goldreich and A. Tal, *Matrix rigidity of random Toeplitz matrices*, Comput. Complex. **27** (2018), 305–350. 29
30. V. Guruswami, C. Umans, and S. Vadhan, *Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes*, J. ACM **56** (2009), A20, 34. 2
31. R. Impagliazzo and R. Paturi, *On the complexity of k-SAT*, J. Comput. Syst. Sci. **62** (2001), 367–375. 11
32. R. Impagliazzo, R. Paturi, and F. Zane, *Which problems have strongly exponential complexity*?, J. Comput. Syst. Sci. **63** (2001), 512–530. 11
33. S. Jukna, *Extremal combinatorics*, second ed., Texts in Theoretical Computer Science, Springer, Heidelberg, 2011. 1
34. ———, *Computational complexity of graphs*, Advances in network complexity, Quant. Netw. Biol., vol. 4, Wiley-Blackwell, Weinheim, 2013, pp. 99–153. 29
35. S. Jukna and I. Sergeev, *Complexity of linear Boolean operators*, Found. Trends Theor. Comput. Sci. **9** (2013), 1–123. 2, 29
36. Karthik C. S., B. Laekhanukit, and P. Manurangsi, *On the parameterized complexity of approximating dominating set*, J. ACM **66** (2019), 1–38. 7, 9, 10, 26, 27

37. Karthik C. S. and P. Manurangsi, *On closest pair in Euclidean metric: monochromatic is as hard as bichromatic*, Combinatorica **40** (2020), 539–573. 4, 5, 10, 26, 29
38. S. Khot, *Ruling out PTAS for graph min-bisection, dense k-subgraph, and bipartite clique*, SIAM J. Comput. **36** (2006), 1025–1071. 6
39. J. Kollár, L. Rónyai, and T. Szabó, *Norm-graphs and bipartite Turán numbers*, Combinatorica **16** (1996), 399–406. 6
40. E. Kushilevitz and N. Nisan, *Communication complexity*, Cambridge University Press, New York, 1997. 2
41. S. Lang and A. Weil, *Number of points of varieties in finite fields*, Amer. J. Math. **76** (1954), 819–827. 9, 19
42. B. Lin, *The parameterized complexity of the k-biclique problem*, J. ACM **65** (2018), 1–23. 3, 6, 7, 29
43. ———, *A simple gap-producing reduction for the parameterized set cover problem*, 46th International Colloquium on Automata, Languages, and Programming, vol. 132, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019, pp. A81, 15. 3
44. J. Matoušek, *On discrepancy bounds via dual shatter function*, Mathematika **44** (1997), 42–49. 9
45. D. Micciancio and S. Goldwasser, *Complexity of lattice problems*, The Kluwer International Series in Engineering and Computer Science, vol. 671, Kluwer Academic Publishers, Boston, MA, 2002. 11
46. M. Naor, L. J. Schulman, and A. Srinivasan, *Splitters and near-optimal derandomization*, 36th Annual Symposium on Foundations of Computer Scienc, IEEE Computer Society, 1995, pp. 182–191. 2
47. W. M. Schmidt, *Equations over finite fields*, Lecture Notes in Mathematics, Vol. 536, Springer-Verlag, Berlin-New York, 1976. 9
48. I. R. Shafarevich, *Basic algebraic geometry*, Springer-Verlag, Berlin-New York, 1977. 15, 17, 18
49. D. A. Spielman, *Linear-time encodable and decodable error-correcting codes*, IEEE Trans. Inf. Theory **42** (1996), 1723–1731. 2
50. C. A. Tovey, *A simplified NP-complete satisfiability problem*, Discrete Appl. Math. **8** (1984), 85–89. 11
51. E. C. Xavier, *A note on a maximum k-subset intersection problem*, Inf. Process. Lett. **112** (2012), 471–472. 6

Department of Mathematics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Email address: bbukh@math.cmu.edu

Department of Computer Science, Rutgers University, Piscataway, NJ 08854, USA

Email address: karthik.cs@rutgers.edu

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA

Email address: narayanan@math.rutgers.edu