Author:
King, Dylan A

Title:
Hypergraph Turán Problems for Daisies

Hypergraph Turán Problems for Daisies

Dylan King

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Master's of Science by Research in the Faculty of Science

School of Mathematics
June 2022

Abstract

Starting with Willem Mantel in 1907 and continuing with the work of Pál Turán and Paul Erdős in the mid-twentieth century, the extremal properties of graphs (or hypergraphs) avoiding a fixed subgraph (or subhypergraph) have been extensively studied, up to the present day. This is the field of so-called 'Turán problems'; it has had much impact on other areas of Mathematics, such as number theory and geometry, as well as on combinatorics and theoretical computer science.

The classical 'Turán problem' for a fixed r-uniform hypergraph F is the following: for each positive integer n, what is the maximum number $ex(n, F)$ of edges we may take in a r-uniform hypergraph H on n vertices that contains no copy of F? The limit of $ex(n, F)/\binom{n}{r}$, as n tends to infinity, is called the *Turán density* of F, and is usually denoted by $\pi(F)$. In the graph case ($r = 2$), all Turán densities are known (by the Erdős-Stone theorem), but for hypergraphs (when $r \geq 3$) even some of the most basic questions remain open.

In this thesis we study a natural and important class of Turán problems for hypergraphs, posed by Bollobás, Leader and Malvenuto, and independently by Johnson and Talbot [11] and (again independently) by Bukh. For integers $r \geq 3$ and $t \geq 2$, an r-uniform *t-daisy* $\mathcal{D}_r^t$ is a family of $\binom{2t}{t}$ r-element sets of the form

$$\{S \cup T \ : T \subset U,\ |T| = t\}$$

for some sets S, U with $|S| = r - t$, $|U| = 2t$ and $S \cap U = \emptyset$. In this thesis we consider the Turán problem for $F = \mathcal{D}_r^t$.

The exact value of $\pi(\mathcal{D}_r^t)$ is not known for any $t \geq 2, r \geq 3$, but we are actually more interested in the behavior of the Turán density as $r \to \infty$. It was conjectured by Bollobás, Leader and Malvenuto in [3] (and independently by Bukh; an equivalent conjecture was made independently by Johnson and Talbot) that the Turán densities of t-daisies satisfy $\lim\limits_{r\to\infty} \pi(\mathcal{D}_r^t) = 0$ for all $t \geq 2$; this has become a well-known problem, and it is still open for all values of t.

In this thesis, we give lower bounds for the Turán densities of r-uniform t-daisies, improving the best known lower bound from exponential to polynomial in r. To do so, we introduce (and make some progress on) the following natural problem in additive combinatorics: for integers $m \geq 2t \geq 4$, what is the maximum cardinality $g(m, t)$ of a subset R of $\mathbb{Z}/m\mathbb{Z}$ such that for any $x \in \mathbb{Z}/m\mathbb{Z}$ and any $2t$-element subset X of $\mathbb{Z}/m\mathbb{Z}$, there are t distinct elements of X whose sum is not in the translate $x + R$? This is a slice-analogue of the extremal Hilbert cube problem considered in [4] and [10]. Finally we conclude the thesis by connecting $\pi(\mathcal{D}_r^t)$ to a problem on the Boolean cube.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... DATE: ....

Contents

1. Introduction

In this thesis, all graphs are *simple* (meaning, without multiple edges or loops), and finite (meaning, their vertex-set is finite). A graph is therefore an ordered pair (V, E) where V is a finite set and E is a set of unordered pairs of elements of V. Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are said to be *isomorphic* if there exists a bijection $f : V_1 \to V_2$ such that $f(\{u, v\}) \in E_2$ if and only if $\{u, v\} \in E_1$.

Given a graph F, we say that a graph G is F*-free* if it does not contain an isomorphic copy of F as a subgraph. The nth *Turán number* of F is defined to be

$$ex(n, F) := \max\{|E| : G = (V, E) \text{ is a } F\text{-free graph with } |V| = n\}.$$

Since the number of graphs on n vertices is finite, $ex(n, F)$ exists, and is an integer between zero and $\binom{n}{2}$. In the case of the triangle K_3, Mantel showed in [19] that $ex(n, K_3) = \lfloor n^2/4 \rfloor$ and that the only graphs attaining equality are the bipartite graphs with balanced vertex-classes (here, *balanced* means the sizes of the vertex-classes can differ by at most one). In 1941, Pál Turán [18] generalized this result to any complete graph K_t, showing that $ex(n, K_t) = (1 - \frac{1}{t-1} + o(1))\binom{n}{2}$ (in fact, determining $ex(n, K_t)$ exactly for all n and t), and showing that the maximum is attained only by a complete $(t-1)$-partite graph with classes of sizes as balanced as possible. As already suggested by our use of $o(1)$ notation, oftentimes we are interested only in the behavior of $ex(n, F)$ as $n \to \infty$, and this is encapsulated in the following Lemma and the definition found within.

Lemma 1.1. *For any graph F, we define the* Turán density *of F by*

$$\pi(F) := \lim_{n\to\infty} \frac{ex(n, F)}{\binom{n}{2}};$$

this limit exists and is contained in $[0, 1]$.

Proof. To prove this lemma, it suffices to show that the sequence $x_n := \frac{ex(n,F)}{\binom{n}{2}}$ is non-increasing; since $x_n \in [0, 1]$ for all n, we will then be done by monotone convergence. Indeed, consider a graph G on n vertices which is F-free and contains $ex(n, F)$ edges. Then any induced subgraph $G' \subset G$ on $n - 1$ vertices is also F-free, and therefore contains at most $ex(n - 1, F)$ edges. There are n choices for G', and summing over all such G' counts each edge of G exactly $n - 2$ times, so

$$(n - 2)ex(n, F) \le nex(n - 1, F) \implies x_n \le x_{n-1}$$

as needed. □

In the language of Turán densities, Turán's result implies that $\pi(K_t) = 1 - \frac{1}{t-1}$, agreeing with intuition that as $t \to \infty$ this density tends to 1. This problem is monotone in F, since for $F \subset F'$ we have $ex(n, F) \le ex(n, F')$, and so far we have considered the increasing sequence of complete graphs K_t. A foundational result comes from Erdős and Stone in [8], who proved that if F has chromatic number $\chi(F) = r \ge 3$, then $\pi(F) = \frac{r-2}{r-1}$. While this asymptotic result cannot tell us $ex(n, F)$ exactly, as Turán's Theorem does for K_t, it is applicable to a much greater number of graphs (since every F has a well-defined chromatic number).

From this starting point, one can branch out into several related lines of questioning. One problem is to consider $ex(n, F)$ for bipartite F, in which case the Erdős-Stone theorem simply says that $ex(n, F) = o(n^2)$, and therefore does not tell us the order of magnitude of $ex(n, F)$. It was proved in [14] that the 4-cycle has $ex(n, C_4) \leq \frac{1}{2}n^{3/2} + \frac{1}{2}n$; a construction of Erdős-Rényi-Sós [7] and (independently) Brown [2] shows that this is asymptotically sharp for infinitely many values of n. The order of magnitude of $ex(n, F)$ is known for $F = K_{3,3}$, but is unknown for $F = K_{t,t}$ for all $t > 3$. One can also consider what happens to G when a class of multiple graphs $\mathcal{F} = \{F_1, F_2, \dots\}$ are forbidden. Other researchers have studied *stability*, showing that F-free graphs with a nearly-maximal number of edges are (quantitatively) close to the special class of extremal graphs [6],[16]. In this thesis we are interested in Turán densities for 'higher-dimensional' versions of graphs.

Hypergraphs are a natural generalization of the graphs we have considered thus far. We will consider finite and simple r-uniform hypergraphs $H = (V, E)$ on the vertex-set $V = [n] = \{1, \dots, n\}$, where now each 'edge' is a subset of $[n]$ of size r so that $E \subset \binom{n}{r}$. For a fixed r-uniform F we can extend the notation from $r = 2$ and again write $ex(n, F)$ for the largest number of edges found in any r-uniform H on n vertices which contains no isomorphic copy of F. The Turán density is defined now as

$$\pi(F) := \lim_{n \to \infty} \frac{\mathrm{ex}(n, F)}{\binom{n}{r}},$$

a limit which exists by essentially the same averaging argument used above in the case $r = 2$.

We have much less understanding of $\pi(F)$ for r-uniform F with $r \geq 3$, compared to in the graph case. For example, let $\mathcal{K}_\ell^{(r)}$ denote the complete r-graph on ℓ vertices, so that Turán's Theorem gives $\pi(\mathcal{K}_\ell^{(2)}) = 1 - \frac{1}{\ell - 1}$, since $\mathcal{K}_\ell^{(2)} = K_\ell$. In contrast, the value $\pi(\mathcal{K}_\ell^{(r)})$ is unknown for any $\ell > r \geq 3$ (when $\ell = r$, $\mathcal{K}_r^{(r)}$ is a single edge, so trivially $\pi(\mathcal{K}_r^{(r)}) = 0$). Erdős offered a \$1000 reward for showing that $\pi(\mathcal{K}_4^{(3)}) = 5/9$ [9], but this remains one of the most notorious open problems in extremal combinatorics.

In lieu of the broadly applicable theorems we have concerning Turán densities of graphs, some progress has been made by considering very special families of hypergraphs F. One example is the $(2r)$-graph known as the expanded triangle, $C_3^{(2r)}$. An *expanded ((2r)-uniform) triangle* consists of 3 edges, $\{S_1 \cup S_2, S_1 \cup S_3, S_2 \cup S_3\}$, for S_1, S_2, S_3 disjoint sets of size r. In general one can 'expand' any graph into a $2r$-graph (without altering the number of edges) by replacing each graph vertex with a set of r vertices. Independently of r, it turns out that $\pi(C_3^{(2r)}) = \frac{1}{2}$. For a discussion of this problem, its generalization to expansions of other graphs, and other Turán problems for hypergraphs, the reader is referred for example to the survey of Keevash [13].

In this thesis we will be interested in what happens when we forbid a different kind of hypergraph, which we introduce now. For integers $r \geq 2$ and $t \geq 2$, an r-uniform *t-daisy*

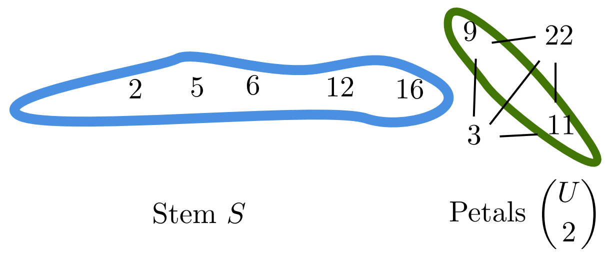

FIGURE 1. A schematic of a $\mathcal{D}_7^2$ with $S = \{2, 5, 6, 12, 16\}$ and $U = \{3, 9, 11, 22\}$. Each of the $\binom{4}{2} = 6$ elements of this $\mathcal{D}_7^2$ are formed by combining the stem S with a petal (a member of $\binom{U}{2}$). Here we have highlighted in blue the stem and in green the petal $\{9, 11\}$, leaving the other petals as edges between elements of U.

$\mathcal{D}_r^t$ is a collection of $\binom{2t}{t}$ r-element sets of the form

$$\{S \cup T \ : T \subset U, \ |T| = t\}$$

for some sets S, U with $|S| = r - t$, $|U| = 2t$ and $S \cap U = \emptyset$. This collection can be seen visually in Figure 1.

We will typically consider this definition for fixed values of t, allowing r to grow. In the case $t = r = 2$, we have $\pi(\mathcal{D}_2^2) = 2/3$ by Turán's theorem for K_4s. The first unknown case occurs when $r = 3$; in this case, Bollobás, Leader and Malvenuto showed in [3] that $\pi(\mathcal{D}_3^2) \geq \frac{1}{2}$ (by taking the complement of the Fano plane, blowing up and iterating) and conjectured that in fact equality holds; this is still open.

For larger $t \geq 2$ and $r \geq 3$, even less is known concerning $\pi(\mathcal{D}_r^t)$. We observe that the value of $\pi(\mathcal{D}_r^t)$ is clearly nondecreasing in t, since an r-uniform family that is free of t-daisies is also free of t'-daisies for all $t' > t$. Note also that if H is a $\mathcal{D}_r^t$-free r-graph and $x \in V$ is an element of the ground set, then the $(r-1)$-graph $H' = (V', E')$ with vertex-set $V' = V \setminus \{x\}$ and edge-set

$$E' = \{e \setminus x \ : \ e \in E \ \text{ with } x \in e\},$$

(also known as the *link of H at x*), is $\mathcal{D}_{r-1}^t$-free. It follows by averaging over all such links (similarly to the reasoning used in Lemma 1.1) that $\pi(\mathcal{D}_r^t)$ is nonincreasing in r.

The following conjecture was made by Bollobás, Leader and Malvenuto in [3] (and independently by Bukh, see [3]).

Conjecture 1.2. *For all $t \geq 2$, $\lim_{r\to\infty} \pi(\mathcal{D}_r^t) = 0$.*

This is still open, even for $t = 2$. (It is not even known whether the sequence $(\pi(\mathcal{D}_r^t))_r$ is strictly decreasing, for any value of $t \geq 2$.) Johnson and Talbot independently made an equivalent conjecture in [11], which we now describe. We will need the standard definition of the Boolean cube, and a subcube thereof.

Definition 1.3. *For $n \in \mathbb{N}$ the* n-dimensional Boolean cube *is* $\{0,1\}^n$*. For* $1 \le d \le n$*, a* d-dimensional subcube *of* $\{0,1\}^n$ *is a subset of* $\{0,1\}^n$ *of the form*

$$\{x \in \{0,1\}^n \ : \ x_i = a_i \ \forall \ i \in I\}$$

for some set $I \in \binom{[n]}{n-d}$ *(called the set of* fixed coordinates*) and values* $a_i \in \{0,1\} \ \forall \ i \in I$*.*

Conjecture 1.4. *Let* $d \ge 2$ *and* $\delta \in (0,1]$*. Then for* n *sufficiently large depending on* d *and* δ*, and any set* $A \subset \{0,1\}^n$ *with* $|A| \ge \delta 2^n$*, there exists a* d*-dimensional subcube* $\mathcal{C}$ *with* $|A \cap \mathcal{C}| \ge \binom{d}{\lfloor d/2 \rfloor}$*.*

It is easy to verify Conjecture 1.4 for $d = 2$ and $d = 3$, but it remains open for all $d \ge 4$. By setting

$$A = \{x \in \{0,1\}^n \ : \ \sum_{i=1}^{n} x_i \equiv 0 \ (\text{mod. } d+1)\},$$

i.e. every $(d+1)$th slice, we have $\lim_{n\to\infty} |A|/2^n = 1/(d+1)$, but any d-dimensional subcube meets exactly one of these slices, and therefore we cannot ask for more than $\binom{d}{\lfloor d/2 \rfloor}$ in Conjecture 1.4. Before showing that Conjectures 1.2 and 1.4 are equivalent we recall the classic Chernoff bound which shows that sums of independent random variables fall within a constant fraction of their expectation with probability exponentially close to 1. We do not state the strongest known form of this statement, but instead exactly the weaker form we require.

Lemma 1.5 (Chernoff Bound)**.** *Suppose that* $X_1, \dots, X_n$ *are independent random variables with* $X := \sum_{i=1}^{n} X_i$ *and* $\mu := \mathbb{E}[X]$*. Then for any* $0 \le \delta \le 1$*,* $\mathbb{P}(|X - \mu| \ge \delta\mu) \le 2e^{-\delta^2 \mu/3}$*.*

We will use this bound in the following Lemma, a detailed proof of an argument outlined in [3].

Lemma 1.6. *Conjectures 1.2 and 1.4 are equivalent for* $d = 2t$*.*

Proof. First we show that Conjecture 1.2 implies Conjecture 1.4. Assume the validity of Conjecture 1.2. Let X_i take values 0 and 1 each with probability 1/2 independently at random, for $1 \le i \le n$. By allowing these X_i to specify an element of $\{0,1\}^n$ and applying Lemma 1.5 with $\mu = n/2$ and $\delta = 1/2$ we have

$$|\{x \in \{0,1\}^n \ : \sum_{i=1}^{n} x_i \notin [n/4, 3n/4]\}| \le 2^{n+1} \cdot e^{-n/24}.$$

Let $A \subset \{0,1\}^n$ with $|A| \ge \delta 2^n$; then

$$|\{x \in A \ : \ n/4 < \sum_{i=1}^{n} x_i < 3n/4\}| \ge 2^n(\delta - 2e^{-n/24}) \ge (\delta/2)2^n$$

provided n is sufficiently large depending on δ. Hence, by averaging, there exists some r with $n/4 < r < 3n/4$ such that

$$|\{x \in A \ : \ \sum_{i=1}^{n} x_i = r\}| \ge \frac{\delta}{2}\binom{n}{r}.$$

Let $\mathcal{F}$ denote the corresponding family of r-element subsets of $[n]$; this has density at least $\delta/2$. Since we are assuming that $\pi(\mathcal{D}_r^t) \to 0$ as $r \to \infty$, there exists r_0 such that $\pi(\mathcal{D}_r^t) < \delta/2$ for all $r \geq r_0$. We would like to conclude immediately from this that $\mathcal{F}$ contains a copy of $\mathcal{D}_r^t$ for $r \geq r_0$, but there is the small problem that n may be too small for us to immediately conclude this. To get around this problem, we average over appropriate links. Observe that, by averaging, for any $u < r$ there exists a u-element subset U such that the link of $\mathcal{F}$ at U, i.e.

$$\{S \setminus U : \ S \in \mathcal{F}, \ U \subset S\} \subset \binom{[n] \setminus U}{r-u}$$

has density at least $\delta/2$. Choose r_0 such that $\pi(\mathcal{D}_r^t) < \delta/2$ for all $r \geq r_0$; then by definition, for all $r \geq r_0$ there exists $n_0(r)$ such that $ex(n, \mathcal{D}_r^t) < (\delta/2)\binom{n}{r}$ for all $n \geq n_0(r)$. Now set $u_0 = r - r_0$; the difference $n - r > n/4$ is unchanged when taking the link at U, so if $n/4 \geq n_0(r_0) - r_0$ we have $n - u_0 \geq n_0(r_0)$. Since there exists a u_0-element subset U such that the link of $\mathcal{F}$ at U has density at least $\delta/2$, it follows that this link contains a copy of $\mathcal{D}_{r_0}^t$, hence so does $\mathcal{F}$. Hence, A contains at least $\binom{2t}{t}$ points of a $(2t)$-dimensional subcube, as required. Next we show that Conjecture 1.4 implies Conjecture 1.2. Suppose that 1.2 is false. Then the monotonicity in r (and in n) of $ex(n, F)$ discussed above implies that there exists $\delta > 0$ such that for any $n \in \mathbb{N}$ we may find, for each $0 \leq r \leq n$, a $\mathcal{D}_r^t$-free r-graph H_r with $V(H_r) = [n]$ and $|E(H_r)| \geq \delta\binom{n}{r}$ (note that any r-graph on $[n]$ is trivially $\mathcal{D}_r^t$-free if $r < t$ or $r > n - t$). Now for each $0 \leq j \leq 2t$ consider the set

$$A_j = \{x \in \{0,1\}^n \ : \ \sum_{i=1}^{n} x_i \equiv j \ (\text{mod. } 2t+1) \text{ and } x \in E(H_r) \ \text{ for } r = \sum_{i=1}^{n} x_i\}.$$

That is, we have taken a dense $\mathcal{D}_r^t$-free graph on every $(2t+1)$th layer. (By averaging over all j there exists j_0 such that A_{j_0} has density at least $\delta/(2t+1)$.) Any d-dimensional subcube $\mathcal{C}$ intersects exactly one of these layers, and $|A \cap \mathcal{C}| < \binom{2t}{t}$ since otherwise H_r would contain a copy of $\mathcal{D}_r^t$. This implies that Conjecture 1.4 is false. □

As a brief reminder, upper bounds on $\pi(\mathcal{D}_r^t)$ show that an r-uniform hypergraph with many edges must contain a daisy. Conjecture 1.2 suggests that as $r \to \infty$ the threshold for "many" tends toward density 0. On the other hand, lower bounds on $\pi(\mathcal{D}_r^t)$ must come from constructions of large daisy-free hypergraphs.

In [3] a lower bound of $\pi(\mathcal{D}_r^2) \geq r!/r^r$ is observed, which comes from considering the r-partite r-uniform hypergraph on $[n]$ with parts of sizes as equal as possible. That is, partition $[n]$ into disjoint sets $V_1, \ldots, V_r$ each of size approximately n/r. Then take only edges which include exactly one element from each V. This gives an edge density of $r!/r^r$ and cannot contain any daisies. To see the edge density consider that to build an edge one element at a time, when adding the ith element we have forbidden $(i-1)n/r$ choices by the elements we have already added. To see that this construction cannot contain any daisies, let $U = \{i, j, k, l\}$ be the petal elements with stem S, for $|S| = r - 2$ and $S \cap U = \emptyset$. Containing the daisy would require including each of $S \cup \{i, j\}$, $S \cup \{i, k\}$, and $S \cup \{j, k\}$.

The first two sets imply that j and k must belong to the same partite class, say W, in which case the third set cannot be taken.

Unfortunately, the lower bound obtained by this short argument is exponentially small in r. In this thesis, we obtain (by discussion of the results shown in [5]) the following improved lower bound (which is polynomial in r), using an additive-combinatorial construction. We also raise a question in additive combinatorics which may be of interest in its own right.

Theorem 1.7. *There exists an absolute constant $c > 0$ such that $\pi(D_r^2) \geq c/r$ for all $r \geq 3$. Furthermore, for each $t \geq 3$, we have*

$$\pi(D_r^t) \geq r^{-\frac{4t-2}{\binom{2t}{t}-1}-O(1/\sqrt{\log r})}.$$

Our proof of Theorem 1.7 relies upon the existence of a subset of $\mathbb{Z}_m := \mathbb{Z}/m\mathbb{Z}$ that avoids a certain additive structure, which we define now.

Definition 1.8. *For positive integers $m, t \geq 2$ with $m \geq 2t$, let $g(m,t)$ denote the maximum possible size of a subset $R \subset \mathbb{Z}_m$ such that for any $x_0 \in \mathbb{Z}_m$ and any $(2t)$-element subset X of $\mathbb{Z}_m$, there are t distinct elements of X whose sum is not contained in $R - x_0$, i.e.*

$$\left\{ x_0 + \sum_{x \in T} x \ : \ T \subset X, \ |T| = t \right\} \not\subset R.$$

For brevity, given a set $X \in \binom{\mathbb{Z}_m}{2t}$ we write

$$C(X) := \left\{ \sum_{x \in T} x \ : \ T \subset X, \ |T| = t \right\}$$

for the set of sums of t distinct elements of X; $g(m,t)$ is the maximum size of a subset of $\mathbb{Z}_m$ containing no translate of $C(X)$ for any $|X| = 2t$.

The function $g(m,t)$ is related to a question raised by Gunderson and Rödl in [10], concerning Hilbert cubes.

Definition 1.9. *If R is a ring, the d-*dimensional Hilbert cube *generated by $x_1, \ldots, x_d \in R$ is the set*

$$\left\{ \sum_{i \in I} x_i \ : \ I \subset \{1, 2, \ldots, d\} \right\} \subset R.$$

Gunderson and Rödl considered large sets of integers which do not contain any translate of a Hilbert cube (working over the integers, i.e., the $R = \mathbb{Z}$ case of Definition 1.9). In particular, they prove the following (Theorem 2.3 and Theorem 2.5 of [10]).

Theorem 1.10. *For each integer $d \geq 3$, there exists $c_d > 0$ such that any set of integers $A \subset [m]$ with $|A| \geq c_d(\sqrt{m}+1)^{2-\frac{1}{2^{d-2}}}$ contains a translate of a d-dimensional Hilbert cube. Furthermore, for all m there exists a set of integers $A \subset [m]$ with $|A| \geq m^{1-\frac{d}{2^d-1}-O(1/\sqrt{\log m})}$ that does not contain any translate of a Hilbert cube.*

(Here, and henceforth, we write $[m] := \{1, 2, \ldots, m\}$ for the standard m-element set.)

We consider the case $R = \mathbb{Z}_m$, but for our purposes, the structural differences between $[m]$ and $\mathbb{Z}_m$ will not be particularly important. Estimating $g(m, t)$ is a natural variant of the Gunderson-Rödl problem, where we avoid only the middle slice of a Hilbert cube of dimension $2t$. We make a small but important change in that, in Definition 1.8, we require that X be composed of $2t$ distinct elements, while a Hilbert cube may even have $x_1 = x_2 \cdots = x_d$. (In this case, the Hilbert cube is a d-element arithmetic progression, but $C(X)$ is a singleton.)

We obtain the following bounds on $g(m, t)$.

Theorem 1.11. *For all $t \geq 3$ and $m \geq 4$ we have*

$$g(m, t) \geq m^{1 - \frac{2t}{\binom{2t}{t} - 1} - O(1/\sqrt{\log m})},$$

and if furthermore m is prime, then

$$g(m, t) \geq m^{1 - \frac{2t-1}{\binom{2t}{t} - 1} - O(1/\sqrt{\log m})}.$$

For $t = 2$ and $m \geq 64$ we have $g(m, 2) \geq \sqrt{m}/8$.

Theorem 1.12. *For each $t \geq 2$ and all m sufficiently large depending on t, we have*

$$g(m, t) \leq 4^{1-1/2^{2t}} (\sqrt{m} + \sqrt{2t})^{2-1/2^{2t-1}}.$$

(Here, we use the standard asymptotic notation: if X is a set and $f, h : X \to \mathbb{R}^+$, we write $f = O(h)$ if there exists an absolute constant $C > 0$ such that $f(x) \leq Ch(x)$ for all $x \in X$.)

It would be of interest to narrow the gap between our upper and lower bounds on $g(m, t)$.

The proofs of Theorem 1.12 and of the first part of Theorem 1.11 (i.e. the $t > 2$ case) are very similar to those used in [10] to prove Theorem 1.10 (above). Our lower bound for $t > 2$ consists of a probabilistic construction very similar to that of Gunderson and Rödl in [10]; we have to choose a random set of slightly lower density as we must avoid the middle slice of a Hilbert cube, as opposed to an entire Hilbert cube. At first sight, it might seem that an upper bound on $g(m, t)$ follows from the upper bound given in Theorem 1.10, since a set which contains a Hilbert cube contains its middle layer, but one must make small changes to the proof in [10] so as to ensure that the generators of the Hilbert cube we find, the x_i, are distinct. We give an explicit construction for the $t = 2$ case in Theorem 1.11: this construction outperforms the probabilistic one in the case $t = 2$.

The remainder of this thesis is structured as follows. In Section 2 we prove Theorem 1.7 using Theorem 1.11, and in Section 3 we prove Theorems 1.11 and 1.12 (note that we do not require the latter in our study of the Turán density of daisies, but it may be of independent interest). Finally in Section 4 we describe the relationship between $\pi(\mathcal{D}_r^t)$ and a Ramsey-theoretic variant.

2. The Proof of Theorem 1.7

Proof. In proving Theorem 1.7, by an appropriate choice of c (and of the absolute constant implicit in the big-Oh notation), we may clearly assume that $r \geq 8$. For $r, t \in \mathbb{N}$ with $r \geq 8$ and n sufficiently large depending on r and t, we proceed to construct a $\mathcal{D}_r^t$-free family of r-element subsets of $[n]$. We use a 'partite' construction, partitioning $[n]$ into L blocks, and then taking only r-sets containing at most one element from each block. We may assume that $n \geq 2r^2$. Let L be a prime number such that $r^2 \leq L \leq 2r^2$ (such exists, by Bertrand's postulate), and for each $i \in [n]$, set $x_i = \left\lfloor L\frac{i-1}{n} \right\rfloor$; note that $0 \leq x_i < L$ for each $i \in [n]$. Thus in our 'partite' construction the kth block are those $i \in [n]$ with $x_i = k$. By Theorem 1.11, there exists a set $R \subset \mathbb{Z}_L$ of size $|R| = g(L,t)$ with the property that for any $X \subset \mathbb{Z}_L$ with $|X| = 2t$ and for any $x_0 \in \mathbb{Z}_L$, we have $x_0 + C(X) \not\subset R$. Define a family $\mathcal{F}_R \subset \binom{[n]}{r}$ by

$$\mathcal{F}_R = \left\{ S \in \binom{[n]}{r} \ : \ \sum_{i \in S} x_i \in R, \quad (\forall i, j \in S)(x_i = x_j \Rightarrow i = j) \right\}.$$

First we check that $\mathcal{F}_R$ is $\mathcal{D}_r^t$-free. Indeed, consider the daisy $\mathcal{D} = \{S_0 \cup T \ : \ T \subset U, \ |T| = t\}$, where $S_0, \ U \subset [n]$ with $|S_0| = r - t$, $|U| = 2t$ and $S_0 \cap U = \emptyset$. Let $x_0 := \sum_{i \in S_0} x_i$. Then the $(2t)$-element set

$$X = \{x_i \ : \ i \in U\} \subset \mathbb{Z}_L$$

must satisfy $x_0 + C(X) \not\subset R$, and therefore there is a t-sum (indexed by $T = \{i_1, i_2, \ldots, i_t\} \subset U$, say) such that

$$x_0 + x_{i_1} + x_{i_2} + \cdots + x_{i_t} = x_0 + \sum_{i \in T} x_i \notin R.$$

It follows that $S := S_0 \cup T \notin \mathcal{F}_R$ and therefore $\mathcal{D} \not\subset \mathcal{F}_R$, as required.

Now to finish the proof of Theorem 1.7 we bound $|\mathcal{F}_R|$ from below. First note that there are at least $\frac{1}{2}\binom{n}{r}$ sets $S \in \binom{[n]}{r}$ with $x_i \neq x_j$ for all $i \neq j, \ i, j \in S$. Indeed, choose a set S uniformly at random from $\binom{[n]}{r}$. Since the probability that a uniformly random two-element subset $\{i, j\}$ of $[n]$ has $x_i = x_j$ is at most $1/L$, we have

$$\begin{aligned} \mathbb{P}(x_i = x_j \text{ for some } i \neq j, \ i, j \in S) &\leq (1/L)\binom{n}{2}\binom{n-2}{r-2} / \binom{n}{r} \\ &= r(r-1)/(2L) \\ &\leq 1/2. \end{aligned}$$

The family $\mathcal{F} = \mathcal{F}_R$, defined above, is $\mathcal{D}_r^t$-free even if the set R is replaced by a translate $R_a := R + a$ for some $a \in \mathbb{Z}_L$. We may sum over all such a to find

$$\sum_{a \in \mathbb{Z}_L} |\mathcal{F}_{R_a}| \geq \frac{1}{2}\binom{n}{r}|R|$$

since any $S \in \binom{[n]}{r}$ has $\sum_{i\in S} x_i \in R_a$ for exactly $|R|$ values of a. Therefore by averaging over all such translates there must be some translate R_a of R such that $|\mathcal{F}_{R_a}| \geq \frac{1}{2}\binom{n}{r}\frac{|R|}{L}$, and therefore

$$\pi(D_r^t) \geq \frac{g(L,t)}{2L}.$$

Now we may apply Theorem 1.11 (recalling that $r^2 \leq L \leq 2r^2$ is prime): when $t = 2$ and $L = m \geq 64$ (which follows from $r \geq 8$), we have

$$\pi(D_r^2) \geq \frac{g(L,2)}{2L} \geq \frac{\sqrt{L}}{16L} \geq \frac{\sqrt{r^2}}{32r^2} = \frac{1}{32r}$$

and when $t > 2$ we have

$$\pi(D_r^t) \geq \frac{L^{1-\frac{2t-1}{\binom{2t}{t}-1}-O(1/\sqrt{\log L})}}{L} \geq \frac{(r^2)^{1-\frac{2t-1}{\binom{2t}{t}-1}-O(1/\sqrt{\log r})}}{2r^2} = r^{-\frac{4t-2}{\binom{2t}{t}-1}-O(1/\sqrt{\log r})},$$

as required. □

3. Bounds on $g(m,t)$

The focus of this section is the analysis of $g(m,t)$.

3.1. **Proof of Theorem 1.11.** Before we begin a formal proof we briefly and informally describe some potential methods for finding a large subset of $\mathbb{Z}_m$ which contains no translates of any $C(X)$. Suppose that we include every element of $\mathbb{Z}_m$ with probability p; then we will typically obtain a set R of size pm, and we will look to choose p in a way so that the number of translates of $C(X)$ present in R is at most, say, $pm/10$ (it would also be possible to ask for 0 translations to be present but allowing $pm/10$ yields slightly better bounds). Then we could delete from R one element of each such $C(X)$, which gives the property we need without reducing by too much the size of R.

Then our primary consideration is the number of translates of $C(X)$ present in our randomly selected R. Given a specific translate $x_0 + C(X)$, it is present with probability $p^{|C(X)|}$. For a rough intuition consider two cases:

(1) If X is an arithmetic progression then $|C(X)| = t^2+1$, and the number of arithmetic progressions in $\mathbb{Z}_m$ is roughly m^2. Then we would require $m^2p^{t^2+1} \leq pm/10$.
(2) If $|C(X)| = \binom{2t}{t}$ then we would have to consider all m^{2t+1} choices for x_0 and X. Then we would require $m^{2t+1}p^{\binom{2t}{t}} \leq pm/10$.

Of course our aim is to choose p as large as possible, and (especially as t grows), the first of these two scenarios is more limiting on p. Furthermore, this analysis does not cover any of the intermediary cases, where $|C(X)|$ is strictly smaller than $\binom{2t}{t}$ but X is not an arithmetic progression.

Here we use the idea that Gunderson and Rödl employed in [10]. Instead of including elements of $\mathbb{Z}_m$ uniformly at random, we first pass to a fairly dense subset of $\mathbb{Z}_m$ which is free of 3-term arithmetic progressions; this 'destroys' a lot of the additive structure we want to avoid. In fact there is so little arithmetic structure that we are always in the ideal second case above, where $|C(X)| = \binom{2t}{t}$.

Proof. First assume $t \geq 3$. By the well-known construction[1] of Behrend in [1], there exists a set $R_0 \subset [\lfloor m/5 \rfloor]$ with $|R_0| = m^{1-\gamma(m)}$ (where $\gamma(m) := \frac{4}{\sqrt{\log{(m/5)}}}$, log denoting the natural logarithm) that contains no 3-term arithmetic progression. Let $R_1 \subset \mathbb{Z}_m$ be the natural embedding of R_0 into $\mathbb{Z}_m$ (here the factor $\frac{1}{5}$ ensures that no 3-term arithmetic progressions appear in $\mathbb{Z}_m$ which were not present in $\mathbb{Z}$). Then R_1 also contains no 3-term arithmetic progression. Set

$$p = \begin{cases} \frac{1}{8} m^{-\frac{2t-1+\gamma(m)}{\binom{2t}{t}-1}} & \text{if } m \text{ is prime,} \\ \frac{1}{8} m^{-\frac{2t+\gamma(m)}{\binom{2t}{t}-1}} & \text{otherwise,} \end{cases}$$

and choose a set $R_2 \subset R_1$ by including each element of R_1 independently at random with probability p. A standard Chernoff bound (for example Theorem 3.5 in [15]) yields

$$\mathbb{P}(|R_2| \leq |R_1|p/2) \leq e^{-|R_1|p/8}. \tag{1}$$

Define the random variable

$$Y = |\{x_0 + C(X) : \ x_0 \in \mathbb{Z}_m, \ X \subset \mathbb{Z}_m, \ |X| = 2t, \ x_0 + C(X) \subset R_2\}|.$$

Since R_2 does not contain any 3-term arithmetic progressions, for any set of the form $x_0 + C(X)$ lying within R_2, we must have $|x_0 + C(X)| = |C(X)| = \binom{2t}{t}$. Indeed, suppose for a contradiction that $X = \{x_1, \ldots, x_{2t}\}$ is a $(2t)$-element subset of $\mathbb{Z}_m$ with $x_0 + C(X) \subset R_2$, where $x_0 \in \mathbb{Z}_m$ and $|C(X)| < \binom{2t}{t}$. Then there exist two distinct t-element subsets of X, $\{x_{i_1}, \ldots, x_{i_t}\} = S_1$ and $\{x_{i'_1}, \ldots, x_{i'_t}\} = S_2$ say, such that

$$x_{i_1} + \cdots + x_{i_t} = x_{i'_1} + \cdots + x_{i'_t};$$

we may assume without loss of generality that $x_{i_1} \in S_1 \setminus S_2$ and $x_{i'_1} \in S_2 \setminus S_1$, so that $x_{i_1} \neq x_{i'_j}$ for all j and $x_{i'_1} \neq x_{i_j}$ for all j. Then

$$\{x_{i'_1} + x_{i_2} + \cdots + x_{i_t}, x_{i_1} + x_{i_2} + \cdots + x_{i_t}, x_{i_1} + x_{i'_2} + \cdots + x_{i'_t}\} \subset R_2$$

is a (nontrivial) 3-term arithmetic progression (with common difference $x_{i_1} - x_{i'_1}$) in R_2, a contradiction.

We now proceed to bound $\mathbb{E}Y$ from above. In the case that m is not prime we may crudely bound the number of possible sets of the form $x_0 + C(X)$ from above by m^{2t+1} (which is the number of choices for $x_0, x_1, \ldots, x_{2t} \in \mathbb{Z}_m$). If m is prime then we may assume each such set has $x_0 = 0$, by translating each of $x_1, \ldots, x_{2t}$ by $t^{-1}x_0$, leaving only

[1] This short and elegant construction considers numbers whose digits, written in an appropriately chosen base, have a fixed Euclidean norm when considered as a vector. Norm inequalities prevent additive collisions and elementary estimates show the set obtained is quite large.

m^{2t} choices. The probability that each fixed set of the form $x_0 + C(X)$ lies in R_2 is of course $p^{\binom{2t}{t}}$. It follows that

$$\mathbb{E}Y \leq \begin{Bmatrix} m^{2t} p^{\binom{2t}{t}} & \text{if } m \text{ is prime} \\ m^{2t+1} p^{\binom{2t}{t}} & \text{otherwise} \end{Bmatrix} \leq m^{1-\gamma(m)} \frac{p}{8}.$$

It follows from Markov's inequality that

$$\mathbb{P}(Y \geq m^{1-\gamma(m)} p/4) \leq 1/2. \tag{2}$$

Combining (1) and (2), we obtain

$$\mathbb{P}\left(|R_2| > m^{1-\gamma(m)} p/2 \text{ and } Y < m^{1-\gamma(m)} p/4 \right) \geq 1 - e^{-m^{1-\gamma(m)} p/8} - \tfrac{1}{2}. \tag{3}$$

Clearly, for any $t \geq 2$ and m sufficiently large depending on t, we have $1 - \gamma(m) - \frac{\gamma(m)+2t}{\binom{2t}{t}-1} > 0$, so for large enough m, the probability in (3) is positive, and therefore there exists a set $R_2 \subset \mathbb{Z}_m$ with $|R_2| > m^{1-\gamma(m)} p/2$ and $Y < m^{1-\gamma(m)} p/4$. Now for each set of the form $x_0 + C(X) \subset R_2$ for $(x_0, X) = (x_0, \{x_1, \ldots, x_{2t}\})$ we remove a single element from R_2, chosen arbitrarily from $x_0 + C(X)$. The total number of elements deleted from R_2 is at most $Y < m^{1-\gamma(m)} p/4$ and we are still left with

$$|R_2| - Y \geq m^{1-\gamma(m)} \frac{p}{4} = \begin{cases} \frac{1}{32} m^{1-\gamma(m) - \frac{2t-1+\gamma(m)}{\binom{2t}{t}-1}} & \text{if } m \text{ is prime} \\ \frac{1}{32} m^{1-\gamma(m) - \frac{2t+\gamma(m)}{\binom{2t}{t}-1}} & \text{otherwise} \end{cases}$$

elements, finishing the proof of the first statement of Theorem 1.11.

Finally, in the case $t = 2$, we give an algebraic construction that improves upon the random one. First we recall the definition of a Sidon set.

Definition 3.1. *A* Sidon set *in a group G is a subset $S \subset G$ such that the only solutions to the equation $a + b = c + d$ with $a, b, c, d \in S$, are the trivial ones (meaning, those with $\{a, b\} = \{c, d\}$).*

It follows from the classical construction of Singer [17] that for any prime p there is a Sidon set of size $p + 1$ inside $\mathbb{Z}_{p^2+p+1}$. Assume that $m \geq 64$ and let p be a prime with $\sqrt{m}/8 \leq p \leq \sqrt{m}/4$ (such exists, by Bertrand's postulate). Let R_0 be a Sidon set of size at least $\sqrt{m}/8$ inside $\mathbb{Z}_{p^2+p+1}$. The image R of R_0 under the natural inclusion map from $\mathbb{Z}_{p^2+p+1}$ to $\mathbb{Z}_m$ is a Sidon set in $\mathbb{Z}_m$ (here we use $p^2+p+1 \leq m/16+\sqrt{m}/4+1 < m/2$). Now we will show that for any x_0 and $X = \{x_1, x_2, x_3, x_4\} \in \binom{\mathbb{Z}_m}{4}$ we have $x_0 + \{x_1 + x_2, x_1 + x_3, x_1 + x_4, x_2 + x_3, x_2 + x_4, x_3 + x_4\} \not\subset R$. Suppose for a contradiction that $x_0 + C(X) \subset R$; then

$$(x_0 + x_1 + x_2) + (x_0 + x_3 + x_4) = (x_0 + x_1 + x_3) + (x_0 + x_2 + x_4) \tag{4}$$

and each term in brackets is an element of R. Since R is a Sidon set, this implies $x_2 = x_3$ or $x_1 = x_4$, contradicting the fact that the x_i are distinct. We have $|R| \geq \sqrt{m}/8$ and therefore we are done in the case $t = 2$. $\square$

3.2. **Proof of Theorem 1.12.**

Proof. We begin with a quick calculation.

Lemma 3.2. *If* $m \geq d+1$ *and* $b \geq \max\{\frac{\sqrt{m}+\sqrt{d}}{2\sqrt{d}}, 4d+1\}$, *then* $\frac{\binom{b}{2}-db}{m-d} \geq \frac{b^2}{4(\sqrt{m}+\sqrt{d})^2}$.

Proof. Since
$$\frac{1}{b} \leq \frac{2\sqrt{d}}{\sqrt{m}+\sqrt{d}},$$
we have
$$\frac{b-1}{b} \geq \frac{\sqrt{m}-\sqrt{d}}{\sqrt{m}+\sqrt{d}}.$$
Since $b \geq 4d+1$ we have $\binom{b}{2} - db \geq \frac{b(b-1)}{4}$, and therefore
$$\binom{b}{2} - db \geq \frac{b^2}{4}\frac{\sqrt{m}-\sqrt{d}}{\sqrt{m}+\sqrt{d}}.$$
Dividing by $m-d$ yields the result. □

We may now obtain our upper bound on $g(m,t)$. Let $A \subset \mathbb{Z}_m$ such that

(5) $$|A| \geq 4^{1-1/2^{2t}}(\sqrt{m}+\sqrt{2t})^{2-1/2^{2t-1}}.$$

We will show that $x_0 + C(X) \subset A$ for some $x_0 \in \mathbb{Z}_m$ and $X \in \binom{\mathbb{Z}_m}{t}$. For $x_1, \ldots, x_d \in \mathbb{Z}_m$ and $A \subset \mathbb{Z}_m$ we define $A_{x_1} := A \cap (A - x_1)$, $A_{x_1,x_2} := A_{x_1} \cap (A_{x_1} - x_2)$, and more generally,
$$A_{x_1,\ldots,x_{d-1},x_d} := A_{x_1,\ldots,x_{d-1}} \cap (A_{x_1,\ldots,x_{d-1}} - x_d).$$
Then $A_{x_1,\ldots,x_d} = \{x \in \mathbb{Z}_m : \ x + \sum_{i\in I} x_i \in A \ \forall\ I \subset [d]\}$ and so A will contain a translate of $C(X)$ if we can find $2t$ distinct elements $x_1, \ldots, x_{2t} \in \mathbb{Z}_m$ with $|A_{x_1,\ldots,x_{2t}}| \geq 1$. We will find such elements inductively, using the following claim.

Claim 3.3. *Provided* m *is sufficiently large depending on* d, *for each* $0 \leq d \leq 2t$ *there exist* d *distinct elements* $x_1, \ldots, x_d \in \mathbb{Z}_m$ *such that*
$$|A_{x_1,\ldots,x_d}| \geq \frac{|A|^{2^d}}{4^{2^d-1}(\sqrt{m}+\sqrt{2t})^{2^{d+1}-2}}$$

(Here, when $d = 0$ the left-hand side is $|A|$.)

Proof of Claim. The proof is by induction on d (with base case $d = 0$, for which the claim holds trivially). Suppose the claim holds for d for elements $x_1, \ldots, x_d$. Every pair of elements of $A_{x_1,\ldots,x_d}$ has a unique difference y, so
$$\sum_{y\in\mathbb{Z}_m} |A_{x_1,\ldots,x_d,y}| = \binom{|A_{x_1,\ldots,x_d}|}{2}.$$

However, we will need to exclude y from those x_i we have already chosen, and therefore crudely we have

$$\sum_{y\in\mathbb{Z}_m\setminus\{x_1,\ldots,x_d\}} |A_{x_1,\ldots,x_d,y}| \geq \binom{|A_{x_1,\ldots,x_d}|}{2} - d|A_{x_1,\ldots,x_d}|.$$

By averaging over $y \in \mathbb{Z}_m \setminus \{x_1,\ldots,x_d\}$, there exists $y' \in \mathbb{Z}_m \setminus \{x_1,\ldots,x_d\}$ such that

$$|A_{x_1,\ldots,x_d,y'}| \geq \frac{\binom{|A_{x_1,\ldots,x_d}|}{2} - d|A_{x_1,\ldots,x_d}|}{m-d}.$$

We now wish to apply Lemma 3.2 with $b = |A_{x_1,\ldots,x_d}|$. The hypotheses that $m \geq d+1$ and $b \geq 4d+1$ are satisfied for m large enough (depending on t), so to apply Lemma 3.2 it remains only to check that $b \geq \frac{\sqrt{m}+\sqrt{d}}{2\sqrt{d}}$, which follows from our inductive hypothesis and our lower bound (5) on $|A|$:

$$\begin{aligned} b &= |A_{x_1,\ldots,x_d}| \\ &\geq \frac{|A|^{2^d}}{4^{2^d-1}(\sqrt{m}+\sqrt{2t})^{2^{d+1}-2}} \\ &\geq \frac{4^{2^d-2^{d-2t}}(\sqrt{m}+\sqrt{2t})^{2^{d+1}-2^{d-2t+1}}}{4^{2^d-1}(\sqrt{m}+\sqrt{2t})^{2^{d+1}-2}} \\ &\geq 4^{1-2^{d-2t}}(\sqrt{m}+\sqrt{2t})^{2-2^{d-2t+1}} \\ &\geq \frac{\sqrt{m}+\sqrt{d}}{2\sqrt{d}}. \end{aligned}$$

Hence, applying Lemma 3.2 we have

$$\begin{aligned} |A_{x_1,\ldots,x_d,y'}| &\geq \frac{\binom{|A_{x_1,\ldots,x_d}|}{2} - d|A_{x_1,\ldots,x_d}|}{m-d} \\ &\geq \frac{|A_{x_1,\ldots,x_d}|^2}{4(\sqrt{m}+\sqrt{d})^2} \\ &\geq \frac{|A|^{2^{d+1}}}{4^{2^{d+1}-2+1}(\sqrt{m}+\sqrt{d})^{2^{d+2}-4+2}}, \end{aligned}$$

as required, so we set $x_{d+1} = y'$. □

Applying Claim 3.3 with $d = 2t$, and using our lower bound (5) on $|A|$, we obtain distinct $x_1,\ldots,x_{2t} \in \mathbb{Z}_m$ such that $|A_{x_1,\ldots,x_{2t}}| \geq 1$, completing the proof of Theorem 1.12. □

4. Connections to Other Problems

To conclude this thesis we place our study of $\pi(\mathcal{D}_r^t)$ in the context of two closely related problems. Results in extremal combinatorics that guarantee the existence of 'nice' substructures (under some condition), come in two varieties, a 'Ramsey' (or 'coloring') version or a (stronger) 'density' version. To see an example (and non-example) in action we state and discuss four classical results.

Theorem 4.1 (Van der Waerden's Theorem)**.** *For all $r, k \in \mathbb{N}$ there exists $N = W(r, k) \in \mathbb{N}$ such that for any k-coloring $c : [N] \to [k]$, there is a nontrivial monochromatic arithmetic progression in $[N]$ of length r.*

Theorem 4.2 (Szemerédi's Theorem)**.** *For all $r \in \mathbb{N}$ and $\delta \in (0, 1]$ there exists $N = S(r, \delta) \in \mathbb{N}$ such that every subset $A \subset [N]$ with $|A| \geq \delta N$ contains a nontrivial arithmetic progression of length r.*

Theorem 4.3 (Finitary Ramsey's Theorem for Hypergraphs)**.** *For all $t, k, r \in \mathbb{N}$ with $t \geq r$, there exists an $N = R(t, k, r) \in \mathbb{N}$ such that for any k-coloring $c : \binom{[N]}{r} \to [k]$ there exists $A \subset [N]$ with $|A| = t$ such that c is constant on $\binom{A}{r}$.*

Example 4.4. *Let $t = 3$, $r = 2$, $\delta < 1/2$, N even, and $G = (V, E)$ be the complete bipartite graph between two vertex sets of size $N/2$. Then $|E| > \frac{1}{2}\binom{N}{2} > \delta\binom{N}{2}$, but G is triangle free, i.e. for any $T \subset [N]$ with $|T| = 3$, we have $\binom{T}{2} \not\subset E$.*

Theorem 4.2 implies Theorem 4.1, if we do not worry about optimizing $W(r, k)$ (although doing so is an interesting pursuit in itself with a rich mathematical theory). Any k-coloring of $[N]$ must include a monochromatic subset of size at least N/k, and we can deduce Theorem 4.1 by applying Theorem 4.2 with $\delta = 1/k$. While Van der Waerden showed that some color class will contain arithmetic progressions, Szemerédi showed (using much more powerful techniques, introducing the celebrated and widely useful Regularity Lemma for the purpose) that in fact those arithmetic progressions can be found in any color class of positive density. One might hope for the same density strengthening of Theorem 4.3, but Example 4.4 shows that we cannot do so (even in the special case $r = 2$).

Returning to the daisy problem, we see that Conjecture 1.2 could be equivalently written as follows.

Conjecture 4.5. *For all $t \geq 2$ and $\delta \in (0, 1]$, provided r is sufficiently large depending on t and δ and N is sufficiently large depending on r, t and δ, any subset $A \subset \binom{[N]}{r}$ with $|A| \geq \delta\binom{N}{r}$ contains an isomorphic copy of $\mathcal{D}_r^t$.*

In the same manner discussed above, this conjecture would immediately imply a 'coloring' version, since a k-coloring of $\binom{[n]}{r}$ will contain at least one color class with density at least $1/k$. In fact, we could go further and find many daisies in this color class by removing daisies as we find them (we remove a daisy simply by recolouring one of the daisy's r-sets). For r large enough so that $1/k - \pi(\mathcal{D}_r^t) > 1/2k$ (guaranteed by Conjecture 4.5), find a $\mathcal{D}_r^t$ and remove one edge from it, until the number of edges remaining in the large color class drops below $\pi(\mathcal{D}_r^t)\binom{n}{r}$. The number of $\mathcal{D}_r^t$ this algorithm discovers is at least $\frac{1}{2k}\binom{n}{r}$.

This argument relies critically on Conjecture 4.5, but in fact a coloring version can be immediately deduced from Ramsey's Theorem (Theorem 4.3).

Theorem 4.6 (Coloring version of Conjecture 4.5)**.** *Let $t \geq 2$, $r \geq t$ and $k \geq 2$. Provided N is sufficiently large depending on k, t and r, for any k-coloring $c : \binom{[N]}{r} \to [k]$, there exists a monochromatic copy of $\mathcal{D}_r^t$.*

Proof. By Theorem 4.3 there exists an integer $m = R(2t, k, t)$ such that every k-coloring of the complete t-graph on m vertices induces a monochromatic $\mathcal{K}_{2t}^{(t)}$ (a complete t-graph on $2t$ vertices). Provided $N \geq r - t + m$ we may choose disjoint sets $S, S' \subset [N]$ with $|S| = r - t$ and $|S'| = m$. We k-color the t-element subsets of S' by the colouring c' defined by $c'(A) = c(S \cup A)$, for $A \in \binom{S'}{t}$. Since $|S'| = m$ there exists $U \subset S'$ with $|U| = 2t$ such that c is constant on $\{S \cup A \ : A \in \binom{U}{t}\}$, and this hypergraph is exactly a t-daisy. □

We remark that an easy variant of the above argument shows that there are many monochromatic copies of $\mathcal{D}_r^t$.

Acknowledgement. This thesis was supervised by David Ellis at the University of Bristol, and the main results can be found in a preprint (coauthored with him) available at https://arxiv.org/abs/2204.08930. Many thanks also go to the University of Bristol (in particular the Mathematics Department) for hosting this research, and the Marshall Scholarship which provided funding. We are also very grateful to Victor Souza for bringing the reference [10] to our attention.

References

[1] F.A. Behrend, On sets of integers which contain no three term arithmetical progression. *Proc. Natl. Acad. Sci. USA* 32 (1946), 331-332.
[2] W. G. Brown, On graphs that do not contain a Thomsen graph, *Canad. Math. Bull.* 9 (1966), 281-285.
[3] B. Bollobás, I. Leader and C. Malvenuto, Daisies and other Turán problems. *Combin. Probab. Comput.* 20 (2011), 743-747.
[4] J. Cilleruelo and R. Tesoro, On sets free of sumsets with summands of prescribed size. *Combinatorica* 38 (2017), 511-546.
[5] D. Ellis and D. King, Lower bounds for the Turán densities of daisies, https://arxiv.org/abs/2204.08930.
[6] P. Erdős, Some recent results on extremal problems in graph theory (Results), *Theory of Graphs (Internl. Symp. Rome)* (1966), 118-123.
[7] P. Erdős, A. Rényi and V. Sós, On a problem of graph theory, *Studia Sci. Math. Hungar.* 1 (1966), 215-235.
[8] P. Erdős and A.H. Stone, On the structure of linear graphs, *Bull. Amer. Math. Soc.* **52** (1946), 1087-1091.
[9] Z. Fu uredi, Turán type problems, *Surveys in combinatorics*, London Math. Soc. Lecture Note Ser. 166, Cambridge Univ. Press, Cambridge, 1991, 253-300.
[10] D.S Gunderson and V. Rödl, Extremal problems for afine cubes of integers. *Combin. Probab. Comput.* 7 (1998), 65-79.
[11] J. R. Johnson and J. Talbot, Vertex Turán problems in the hypercube, *J. Combinatorial Theory (A)* 117 (2010), 454-465.
[12] G. Katona, T. Nemetz, and M. Simonovits, On a problem of Turán in the theory of graphs, *Mat. Lapok* **15** (1964), 228-238.
[13] P. Keevash. Hypergraph Turán problems, *Surveys in combinatorics*, London Math. Soc. Lecture Note Ser. 392, Cambridge Univ. Press, Cambridge, 2011, 83-139.

[14] Kóvari, T., Sós, V., and Turán, P.. On a problem of K. Zarankiewicz, *Colloq. Math. 3.1* (1954), 50-57.
[15] M. Mitzenmacher, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.
[16] M. Simonovits, A method for solving extremal problems in graph theory, stability problems, *Theory of Graphs (Proc. Colloq. Tihany, 1966)*, Academic Press, New York (1968), 279-319.
[17] J. Singer, A theorem in finite projective geometry and some applications to number theory. *Trans. Amer. Math. Soc.* 43 (1938), 377-385.
[18] P. Turán, On an extremal problem in graph theory (in Hungarian), *Mt. Fiz. Lapok* **48** (1941), 436-452.
[19] W. Mantel. Problem 28. *Wiskundige Opgaven*, 10:60-61, 1907.