ABUNDANCE: ASYMMETRIC GRAPH REMOVAL LEMMAS AND INTEGER SOLUTIONS TO LINEAR EQUATIONS

ANTÓNIO GIRÃO, EOIN HURLEY, FREDDIE ILLINGWORTH, AND LUKAS MICHEL

ABSTRACT. We prove that a large family of pairs of graphs satisfy a polynomial dependence in asymmetric graph removal lemmas. In particular, we give an unexpected answer to a question of Gishboliner, Shapira, and Wigderson [GSW23] by showing that for every $t \geqslant 4$, there are K_t-abundant graphs of chromatic number t. Using similar methods, we also extend work of Ruzsa [Ruz93] by proving that a set $\mathcal{A} \subset \{1, \ldots, N\}$ which avoids solutions with distinct integers to an equation of genus at least two has size $\mathcal{O}(\sqrt{N})$. The best previous bound was $N^{1-o(1)}$ and the exponent of $1/2$ is best possible in such a result. Finally, we investigate the relationship between polynomial dependencies in asymmetric removal lemmas and the problem of avoiding integer solutions to equations. The results suggest a potentially deep correspondence. Many open questions remain.

1. INTRODUCTION

Graph theory and additive combinatorics are distinct areas of discrete mathematics with a deep and interesting interplay. One particularly elegant example is the interplay between the *triangle removal lemma* and sets of integers avoiding 3-term arithmetic progressions. The former is an influential result in extremal graph theory proved in the seminal paper of Ruzsa and Szemerédi [RS78]. It states:

> If an n-vertex graph has $o(n^3)$ triangles, then all triangles can be deleted by removing $o(n^2)$ edges.

Contrapositively, if an n-vertex graph requires the deletion of at least εn^2 edges to be made triangle-free, then it contains at least δn^3 triangles where $\delta = \delta(\varepsilon) > 0$ depends only on ε. In many applications it would be useful if the function $\delta(\varepsilon)$ is not too small and, in particular, if it is at least some polynomial. Unfortunately, this is not the case. Ruzsa and Szemerédi made brilliant use of Behrend's construction [Beh46] of large sets $\mathcal{A} \subset [N] := \{1, \ldots, N\}$ with no non-trivial 3-term arithmetic progressions to prove a sub-polynomial upper bound on $\delta(\varepsilon)$. If one is interested in removal lemmas for graphs other than triangles, then one must consider

large sets of integers avoiding additive structures other than 3-term arithmetic progressions. This provides a link between polynomial dependence in removal lemmas and the question:

> How large can a set of integers $\mathcal{A} \subset [N]$ be if it has no non-trivial solutions to some linear diophantine equation?

Such problems are central to additive combinatorics and were systematically studied by Ruzsa [Ruz93] who investigated the properties of equations E that determine the size of the largest subset of $[N]$ containing no non-trivial solutions to E.

In this paper we make progress in three directions:

- We develop a number of tools to construct graphs with polynomial dependencies of δ on ε in *asymmetric* graph removal lemmas. This yields a surprising answer to a question posed by Gishboliner, Shapira and Wigderson [GSW23], uncovering richer behaviour than anticipated.
- We extend the work of Ruzsa [Ruz93] by proving that a set $\mathcal{A} \subset [N]$ which avoids solutions with distinct integers to an equation of genus at least two has size $\mathcal{O}(\sqrt{N})$.
- We probe the relationship between these two problems, proving both positive and negative results.

We now provide a miniature introduction for each direction, before proving the relevant results for the three bullet points in Sections 2, 3 and 4 respectively.

Notation. All logarithms are natural unless explicitly stated otherwise. We always consider ε to be a positive real number much smaller than 1. We use the standard $o(\cdot), \mathcal{O}(\cdot)$ and $\Omega(\cdot)$ notation, and use $x(y) = \text{poly}(y)$ as shorthand for *there exists polynomials P and Q such that $P(y) < x(y) < Q(y)$ for all relevant y*. We denote the number of edges and vertices of a graph G by $e(G)$ and $|G|$, respectively. Moreover, we use P_k to denote the path on k vertices.

1.1. **Asymmetric graph removal lemma.** Generalising the triangle removal lemma, Erdős, Frankl, and Rödl [EFR86] proved the *graph removal lemma*: if an n-vertex graph cannot be made F-free by deleting εn^2 edges, then it contains at least $\delta n^{|F|}$ copies of F where $\delta > 0$ depends only on ε and F. The aforementioned bound of Ruzsa and Szemerédi shows that, for triangles and many other graphs, δ must be sub-polynomial in ε (in fact, at most $\varepsilon^{\Omega(\log(\varepsilon^{-1}))}$). There is particular interest in determining for which F we have $\delta = \text{poly}(\varepsilon)$. If this is the case, then the removal lemma is said to be *efficient* as it leads to algorithms for efficient property testing. Polynomial dependence in removal lemmas has been studied not only for subgraphs but also in the settings of arithmetic cycles [FL17, FLS18], induced subgraphs [AF15, AS06], digraphs [AS04], ordered graphs [GT22], matrices [ABE20, AFN07], posets [KF23], and graph properties [GS21].

Alon [Alo02] proved that the only graphs for which $\delta = \text{poly}(\varepsilon)$ are bipartite graphs. Gishboliner, Shapira, and Wigderson [GSW23], building on work of Csaba [Csa21], observed that the situation is more complicated for *asymmetric* removal lemmas, where the graph that cannot be

eliminated by deleting εn^2 edges and the graph of which there are many copies are different. They showed that if an n-vertex graph cannot be made C_{2k-1}-free by deleting εn^2 edges, then it contains at least $\mathrm{poly}(\varepsilon) \cdot n^{2k+1}$ copies of C_{2k+1}. To capture such pairs of graphs with a polynomial dependency, Gishboliner, Shapira, and Wigderson defined a graph H to be *F-abundant* if any n-vertex graph that cannot be made F-free by deleting εn^2 edges must contain at least $\mathrm{poly}(\varepsilon) \cdot n^{|H|}$ copies of H. Being F-abundant is a natural property that is preserved under taking subgraphs and blow-ups (see Lemma 2.3). The obvious question of course is which graphs are F-abundant?

Certainly we require that H is homomorphic F, written $H \to F$ (see Lemma 2.3). Since F-abundance is preserved under taking blow-ups and subgraphs, if G is F-abundant and H is homomorphic to G, then H is F-abundant.

Therefore, we can summarise the positive results of [GSW23] as follows: for all $k \geqslant 2$, if H is homomorphic to C_{2k+1}, then H is C_{2k-1}-abundant. While this is a surprising result, odd cycles are a very particular family of graphs with many unusual properties. Thus it is natural to wonder, given the results of [GSW23], if abundance is just another quirk of odd cycles. Indeed, Gishboliner, Shapira, and Wigderson asked [GSW23, Problem 1.9] whether there are any K_t-abundant graphs of chromatic number t for $t \geqslant 4$ and mentioned that they were inclined to believe that the answer was no. We show that the answer to the question is *yes* in a strong sense: F-abundant graphs have no restriction whatsoever on their homomorphic images beyond the trivial requirement that they are homomorphic to F.

Theorem 1.1. *Let F and G be graphs with $F \nrightarrow G$. There is an F-abundant graph H with $H \nrightarrow G$.*

In particular, taking G to be the complete graph on $\chi(F) - 1$ vertices shows that for any graph F there is an F-abundant graph with the same chromatic number as F. By standard arguments one can also take H to have arbitrarily high girth (see Corollary 2.17). Our method of proof for Theorem 1.1 employs a number of glueing type operations (see Theorems 2.7 and 2.9) on graphs that we show maintain F-abundance. This gives lots of flexibility when constructing F-abundant graphs, and shows that the family of pairs of graphs that satisfy an abundance relation is very rich indeed (see Theorem 2.14).

For the simplest abundant pair, which is C_5 and K_3, we also demonstrate how to prove effective bounds on the degree of the polynomial dependency. This analysis also extends to the sparse case, where ε may depend on n, see Section 2.4.

1.2. **Avoiding solutions to equations in the integers.** Consider the equations

$$x + y = 2z \quad \text{and} \quad x + y = z + w.$$

An integer solution to the first is a 3-term arithmetic progression, while the second is the famed Sidon equation [Sid32]. It is known that any set $\mathcal{A} \subset [N]$ that contains no non-trivial solution

to the Sidon equation has size at most $\mathcal{O}(\sqrt{N})$ while there exist sets with no 3-term arithmetic progressions of size $N^{1-o(1)}$.

In order to understand the source of this wildly differing behaviour, Ruzsa [Ruz93] initiated the systematic study of *translation-invariant equations*, namely equations

$$a_1 x_1 + a_2 x_2 + \cdots + a_k x_k = 0 \tag{1}$$

with non-zero integer coefficients satisfying $a_1 + \cdots + a_k = 0$. Such equations always have trivial solutions where all variables are equal. For a translation-invariant equation E, Ruzsa defined the two quantities (see Section 3 for the definitions of non-trivial, genus, and symmetric in what follows)

$$r_E(N) = \max\{|\mathcal{A}| : \mathcal{A} \subset [N], \text{there is no non-trivial solution to } E \text{ with the } x_i \text{ all in } \mathcal{A}\},$$

$$R_E(N) = \max\{|\mathcal{A}| : \mathcal{A} \subset [N], \text{there is no solution to } E \text{ with the } x_i \text{ all distinct and in } \mathcal{A}\}.$$

For any equation E, $r_E(N) \leqslant R_E(N) \leqslant N(\log N)^{-\alpha_E}$ for some positive constant α_E, but this upper bound may be far from tight as in the case of the Sidon equation. In fact, Ruzsa showed that if E has genus m, then $r_E(N) = \mathcal{O}(N^{1/m})$. In the special case that E is a symmetric equation in at least 4 variables (and so has genus at least two), Ruzsa showed that $R_E(N) = \mathcal{O}(\sqrt{N})$. However, for a general equation E of genus at least two it was not known whether there is any such improvement for the upper bound on R_E. We show that there is.

Theorem 1.2. *Let E be an equation of genus at least 2. Then $R_E(N) = \mathcal{O}(\sqrt{N})$.*

In fact we also show that there are many such solutions, see Corollary 3.5. We note that for any integer $\ell \geqslant 2$ and any $\varepsilon > 0$, Ruzsa showed that there is an equation E of genus ℓ for which $R_E(N) = \Omega(N^{1/2-\varepsilon})$. Thus the exponent in Theorem 1.2 is the best one can hope for.

1.3. **From graphs to equations and back.** An F-coloured graph (G, σ) is a graph G along with a homomorphism $\sigma \colon G \to F$, and one can analogously define F-abundance for F-coloured graphs (see Section 2.1). Ruzsa and Szemerédi [RS78] provided a way of associating a collection of systems of equations $\mathcal{E}(G, \sigma)$ to an F-coloured graph, with each system $S \in \mathcal{E}(G, \sigma)$ containing one equation for each cycle of the graph. This translated the problem of proving non-F-abundance of an F-coloured-graph to the problem of proving that there exists a set $\mathcal{A} \subset [N]$ of size $N^{1-o(1)}$ containing few solutions (with distinct integers) to any of the systems of equations. We thus extend the definition of $R_E(N)$ to systems of equations in the obvious way, where a solution to a system is simultaneously a solution to all of the equations. Further, we say a system of equations S is *avoidable* if $R_S(N) = N^{1-o(1)}$. In this language, the Ruzsa-Szemerédi construction shows:

> *If a coloured graph (G, σ) is F-abundant, then every system $S \in \mathcal{E}(G, \sigma)$ is unavoidable.*

The only method we have for proving non-abundance is this construction of Ruzsa and Szemerédi, and the only equations that are known to be avoidable are *convex* equations [Ruz93, Thm. 2.3], in which all but one of the coefficients in (1) have the same sign, such as $x + y = 2z$. Theorem 1.2 says that equations of genus at least 2 are unavoidable. This leaves open the case of equations of genus 1 that are not convex. Indeed, Ruzsa remarked that the following[1] remained a possibility, although there was "too little evidence to call it a conjecture".

Question 1.3 ([Ruz93, (9.1)]). *Are all genus 1 equations avoidable?*

The answer to Ruzsa's question has immediate implications for graph removal lemmas. For example, as far as we know, it may be that all tripartite triangle-free graphs are K_3-abundant. However, Gishboliner, Shapira, and Wigderson [GSW23] showed that *if the answer to Ruzsa's question is yes*, then non-abundant tripartite triangle-free graphs exist. Indeed, they showed that an appropriately chosen random graph would bear witness. Nevertheless, they also asked whether one could *unconditionally* show that a non-abundant tripartite triangle-free graph exists. They suggested a way to do this via convex equations. We make two contributions here. We show that the proposed method to use convex equations cannot work (see Lemma 4.1) and we provide a small explicit example of a tripartite triangle-free graph that is non-abundant *if* Ruzsa's question has a positive answer.

Both within this paper and outside of it, the connections between equations and graphs have thus far been one way, using avoidable equations to prove non-abundance of graphs. It therefore remained possible that the additive angle is simply a convenient way of proving non-abundance and nothing more. We show that the connection between the two questions is deeper than that. As some setup is required we only informally state our two results here (see Section 4 for details). Firstly, we show that in the case of coloured cycles, a family of graphs containing both abundant and non-abundant graphs, we have a complete picture (see Theorem 4.3) which can be paraphrased as:

> For an F-coloured cycle (C, σ), the avoidability of $\mathcal{E}(C, \sigma)$ determines the F-abundance of (C, σ).

Of course, cycles are a particularly simple family, but we also prove a result for general K_3-coloured graphs (see Theorem 4.4) which can be paraphrased as:

> For a K_3-coloured graph (G, σ), if $\mathcal{E}(G, \sigma)$ is symmetric (in Ruzsa's sense), then G is triangle-abundant.

In fact we show a stronger if and only if statement, but that requires some setup.

2. F-ABUNDANT GRAPHS

We start by formally introducing abundance, giving some of its basic properties, and by developing useful machinery. With this in hand we prove that certain glueing type operations

---

[1] In fact he asked a stronger question but [Buk08] proved this to be false.

(see Theorems 2.7 and 2.9) maintain abundance which allows us to construct a large family of F-abundant graphs (see Theorem 2.14). We use this to resolve some open cases such as showing that the Petersen graph is triangle-abundant and prove the main result of this section, Theorem 1.1.

We start with some useful definitions. An n-vertex graph is *ε-far from F-free* if it cannot be made F-free by deleting fewer than εn^2 edges. F-abundance can now be phrased as follows.

Definition 2.1 (F-abundant). *Let H and F be graphs. H is F-abundant if for all $\varepsilon > 0$ and for all sufficiently large[2] n, every n-vertex graph that is ε-far from F-free contains $\mathrm{poly}(\varepsilon) \cdot n^{|H|}$ copies of H. The polynomial may only depend on H and F.*

There is a simple condition that H must satisfy to have any chance of being F-abundant. A *blow-up* of a graph F is obtained by replacing each vertex i of F by a non-empty independent set V_i and each edge ij by a complete bipartite graph between parts V_i and V_j. Blow-ups of F where each V_i has the same size are ε-far from F-free and so, for H to have any chance of being F-abundant, H must be a subgraph of a blow-up of F. This last condition is very natural in the language of graph homomorphisms.

Definition 2.2 (Homomorphisms). *A graph G is F-partite if there is a partition $(V_i : i \in V(F))$ of $V(G)$ into independent sets such that if $uv \in E(G)$ for $u \in V_i$ and $v \in V_j$, then $ij \in E(F)$. The partition $(V_i : i \in V(F))$ is an F-partition of G.*

The map $\sigma\colon G \to F$ defined so that $v \in V_{\sigma(v)}$ for every vertex of G is called an F-colouring of G or a homomorphism from G to F. If G has an F-colouring, then we say that G is homomorphic to F and denote this by $G \to F$.

We now collect together some basic facts about F-abundance before developing some machinery in the next section.

Lemma 2.3. *Let H and F be graphs.*

1. *If H is F-abundant, then $H \to F$.*
2. *F is F-abundant if and only if F is bipartite.*
3. *F-abundance is preserved under taking subgraphs and blow-ups. In particular, if H is F-abundant and $H' \to H$, then H' is F-abundant.*
4. *If $e(F) > 0$, then every bipartite graph is F-abundant.*

Proof. Note that H is F-partite if and only if H is a subgraph of some blow-up of F. This together with the discussion preceding Definition 2.2 proves part 1. Part 2 is just a restatement of Alon's result [Alo02] that there is polynomial dependence in the F-removal lemma if and only if F is bipartite.

Suppose an n-vertex graph G contains $\varepsilon n^{|H|}$ copies of H. Let H' be a subgraph of H. Each copy of H' in G is contained in at most $n^{|H|-|H'|}$ copies of H and so G contains at least $\varepsilon n^{|H'|}$

---

[2]The 'sufficiently large' condition was not stated explicitly in [GSW23] but was implicitly used there.

copies of H'. Now let H'' be a blow-up of H. A standard supersaturation argument (such as [Kee11, Lem. 2.1]) together with Erdős's result on the extremal function for complete ℓ-uniform ℓ-partite hypergraphs [Erd64] shows that G contains $\text{poly}(\varepsilon) \cdot n^{|H''|}$ copies of H''. Thus F-abundance is preserved under taking subgraphs and blow-ups which proves part 3.

Finally, if $e(F) > 0$, then a single edge is plainly F-abundant. Every bipartite graph is homomorphic to an edge and so is F-abundant by part 3. $\qquad\square$

2.1. Canonical structure of graphs that are ε-far from F-free.

To understand F-abundance it will first be useful to understand the structure of graphs G that are ε-far from F-free. We will show that we may pass to an F-partite subgraph of G that is $\Omega(\varepsilon)$-far from F-free and has some additional structure that will be helpful for proving F-abundance.

We say that a graph G is *uniformly ε-far from F-free* if it has an F-partition $(V_i : i \in V(F))$ and a collection $\mathcal{C}$ of edge-disjoint copies of F such that

- for every copy of F in $\mathcal{C}$ and each $i \in V(F)$, the vertex corresponding to i is in V_i, and
- every vertex of G is in at least $\varepsilon \cdot |G|$ copies of F from $\mathcal{C}$.

We first note a few simple properties of graphs that are uniformly ε-far from F-free.

Lemma 2.4. *Let G be uniformly ε-far from F-free and let $(V_i : i \in V(F))$ be the corresponding F-partition of G. Then,*

- *G is $\varepsilon/|F|$-far from F-free,*
- *for all $i \in V(F)$, every vertex of V_i has at least $\varepsilon \cdot |G|$ neighbours in V_j for every neighbour j of i in F, and*
- *$|V_i| \geqslant \varepsilon \cdot |G|$ for every non-isolated vertex i of F.*

Proof. Let V_i be the largest part in the F-partition, so $|V_i| \geqslant |G|/|F|$. Each vertex of V_i is in at least $\varepsilon \cdot |G|$ copies of F from $\mathcal{C}$ which implies that

$$|\mathcal{C}| \geqslant |V_i| \cdot \varepsilon \cdot |G| \geqslant \varepsilon/|F| \cdot |G|^2.$$

To make G F-free requires deleting an edge from each copy of F in $\mathcal{C}$, and these copies are edge-disjoint. Hence, at least $|\mathcal{C}|$ edges need deleting and so G is $\varepsilon/|F|$-far from F-free.

Fix two adjacent vertices i and j of F and let $v \in V_i$. There are at least $\varepsilon \cdot |G|$ copies of F in $\mathcal{C}$ that contain v and every such copy also contains a vertex u in V_j that is adjacent to v. As the copies of F in $\mathcal{C}$ are edge-disjoint, the vertices u in V_j are all distinct and so v has at least $\varepsilon \cdot |G|$ neighbours in V_j.

Finally, let i be a non-isolated vertex of F. Let j be a neighbour of i in F and fix some $v \in V_j$. Vertex v has at least $\varepsilon \cdot |G|$ neighbours in V_i and so $|V_i| \geqslant \varepsilon \cdot |G|$. $\qquad\square$

Lemma 2.4 shows that graphs which are uniformly far from F-free are far from F-free (up to a multiplicative loss of $|F|$). Next we prove a converse: every graph that is far from F-free contains a large subgraph that is uniformly far from F-free. In particular, graphs which are

7

uniformly ε-far from F-free are canonical for graphs that are ε-far from F-free with only a constant multiplicative loss in the parameter ε, and so we can use such graphs for showing F-abundance.

Lemma 2.5. *Let G be an n-vertex graph that is ε-far from F-free. Then G has a subgraph G' of order at least $(\varepsilon/|F|^{|F|})^{1/2} \cdot n$ that is uniformly $\varepsilon/(2 \cdot |F|^{|F|})$-far from F-free.*

Proof. Since G is ε-far from F-free, it contains a collection $\mathcal{C}$ of εn^2 edge-disjoint copies of F. First take a random F-partition of G and say a $C \in \mathcal{C}$ *survives* if the vertices of C are in the corresponding parts of the F-partition. Each $C \in \mathcal{C}$ survives with probability $1/|F|^{|F|}$ so, for some F-partition, at least $1/|F|^{|F|}$ of copies of F in $\mathcal{C}$ survive. Let G_1 be the corresponding F-partite subgraph and $\mathcal{C}_1$ be the set of surviving copies of F. Note that $|\mathcal{C}_1| \geqslant |\mathcal{C}|/|F|^{|F|} \geqslant \varepsilon/|F|^{|F|} \cdot n^2$.

Next we sparsify. If $v \in G_1$ is in fewer than $\varepsilon/(2 \cdot |F|^{|F|}) \cdot n$ copies of F from $\mathcal{C}_1$, then delete the edges of all these copies of F. Do this until every vertex is either in no copy of F from $\mathcal{C}_1$ or in at least $\varepsilon/(2 \cdot |F|^{|F|}) \cdot n$ of them. The number of deletions is less than $\varepsilon/(2 \cdot |F|^{|F|}) \cdot n^2$ and so at least $\varepsilon/(2 \cdot |F|^{|F|}) \cdot n^2$ copies of F from $\mathcal{C}_1$ remain. Call the set of remaining copies of F $\mathcal{C}'$. Let G' be the graph with vertex-set $\bigcup_{C \in \mathcal{C}'} V(C)$ and edge-set $\bigcup_{C \in \mathcal{C}'} E(C)$.

Note that G' is an F-partite subgraph of G, $\mathcal{C}' \subset \mathcal{C}$ is a collection of edge-disjoint copies of F in G', and every vertex of G' is in at least $\varepsilon/(2 \cdot |F|^{|F|}) \cdot n$ copies of F from $\mathcal{C}'$. Hence, G' is uniformly $\varepsilon/(2 \cdot |F|^{|F|})$-far from F-free. Finally, since $|\mathcal{C}'| \geqslant \varepsilon/(2 \cdot |F|^{|F|}) \cdot n^2$, G' has at least that many edges and so has order at least $(\varepsilon/|F|^{|F|})^{1/2} \cdot n$. $\qquad\square$

Lemmas 2.4 and 2.5 give the canonical F-partite structure of graphs that are ε-far from F-free. With this in mind, it is natural to talk about *F-coloured graphs* being F-abundant. An F-coloured graph is a pair (H, σ) where H is a graph and σ is an F-colouring of H.

Definition 2.6. *Let (H, σ) be an F-coloured graph. (H, σ) is F-abundant if for all $\varepsilon > 0$ and for all sufficiently large n, every n-vertex graph that is uniformly ε-far from F-free graph with corresponding F-partition $(V_i : i \in V(F))$ contains $\mathrm{poly}(\varepsilon) \cdot n^{|H|}$ copies of (H, σ) (that is, each $v \in H$ is embedded in part $V_{\sigma(v)}$). The polynomial may only depend on F, H, and σ.*

To show that a graph H is F-abundant, it suffices to show that there is some F-colouring σ of H such that (H, σ) is F-abundant. In fact, all known proofs of abundance implicitly do this. Whether or not this is necessary is an important question, related to the compactness of abundance, see Section 5.

2.2. **Building up abundance.** Now we develop some operations that allow us to build F-abundant F-coloured graphs from smaller ones. The first lets us add vertices with monochromatic neighbourhoods.

Theorem 2.7. *Let (H, σ) be an F-coloured graph that is F-abundant. Let $v_1, \ldots, v_s \in V(H)$ be vertices satisfying $\sigma(v_1) = \sigma(v_2) = \cdots = \sigma(v_s)$. Let (H', σ') be an F-coloured graph obtained*

from (H, σ) by joining a new vertex u to $v_1, \ldots, v_s$ and taking $\sigma'(u)$ to be any neighbour of $\sigma(v_1)$ in F. Then (H', σ') is F-abundant.

Proof. Without loss of generality, we may assume that $V(F) = \{1, 2, \ldots, |F|\}$, $\{1, 2\}$ is an edge of F, $\sigma(v_1) = \sigma(v_2) = \cdots = \sigma(v_s) = 1$, and $\sigma'(u) = 2$. Let G be an n-vertex graph that is uniformly ε-far from F-free. Let $(V_i : i \in V(F))$ be the corresponding F-partition of G, and let $\mathcal{C}$ be the corresponding collection of edge-disjoint copies of F.

Fix any vertex $x \in V_2$ (this will play the role of u) and let $U_1 = N(x) \cap V_1$ (we will find the vertices corresponding to $v_1, \ldots, v_s$ in U_1). Note that, by Lemma 2.4, $|U_1| \geqslant \varepsilon n$. Consider $G_1 = G[U_1, V_2, V_3, \ldots, V_{|F|}] - x$ and let $\mathcal{C}_1 = \{C \in \mathcal{C} : V(C) \cap U_1 \neq \varnothing \text{ and } x \notin V(C)\}$. Since every vertex of U_1 is in at least εn copies of F from $\mathcal{C}$ and x is in at most n copies of F from $\mathcal{C}$ (the copies of F from $\mathcal{C}$ are edge-disjoint), $\mathcal{C}_1$ is a collection of at least $\Omega(\varepsilon^2) \cdot n^2$ edge-disjoint copies of F in G_1.

By Lemma 2.5, there is an uniformly $\Omega(\varepsilon^2)$-far from F-free subgraph G' of G_1 of order $\Omega(\varepsilon) \cdot |G_1| = \Omega(\varepsilon^2) \cdot n$. Let $A_1 \subset U_1$, $A_2 \subset V_2$, $\ldots$, $A_{|F|} \subset V_{|F|}$ be the corresponding F-partition of G'. Since (H, σ) is F-abundant, G' contains $\mathrm{poly}_1(\varepsilon) \cdot n^{|H|}$ copies of (H, σ). In particular, each v_i is embedded into $A_1 \subset U_1 \subset N(x)$. Thus, adding x to any such copy of (H, σ) gives a copy of (H', σ'). Since there were $|V_2| \geqslant \varepsilon n$ choices for x in V_2, there are $\mathrm{poly}_2(\varepsilon) \cdot n^{|H'|}$ copies of (H', σ') in G. $\qquad\square$

Example 2.8. *One can see that C_5 is K_3-abundant by noting that the 3-colouring of C_5 in the middle of Figure 1 has a vertex with a monochromatic neighbourhood (in fact, any 3-colouring has this property). Once this vertex has been deleted the resulting graph (depicted on the left of the figure) is a path, which is abundant (ignore the right hand picture for now).*
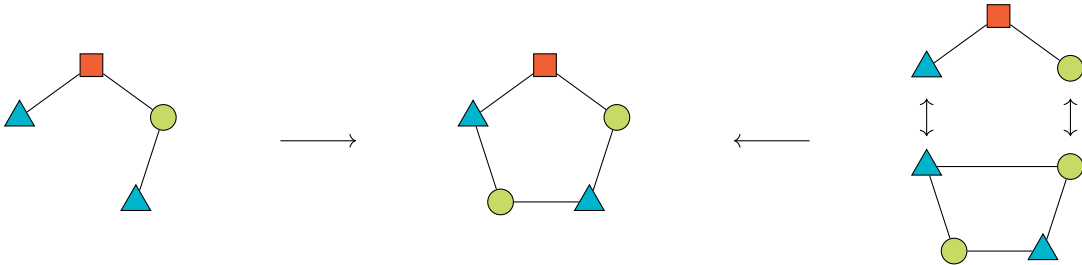

FIGURE 1. Two ways to show the K_3-abundance of C_5.

The second operation allows us to glue along an edge.

Theorem 2.9. *Let (H, σ) and (H', σ') be F-coloured graphs that are F-abundant. Let $uv \in E(H)$, $U', V' \subset V(H')$ satisfy $\sigma'(U') = \sigma(u)$, and $\sigma'(V') = \sigma(v)$. Let the F-coloured graph (H'', σ'') be obtained by*

- *deleting the edge uv from H,*
- *blowing up u by $|U'|$ and v by $|V'|$,*

- *identifying each vertex of U' with a unique copy of u and each vertex of V' with a unique copy of v.*

Then (H'', σ'') is F-abundant.

The following example is particularly simple and worth bearing in mind for the proof. Let H be the C_4 and H' be the path.

Example 2.10. *The K_3-coloured C_4 on the right of Figure 1 is K_3-abundant (for example, by Theorem 2.7). Moreover, the path depicted on the right is also K_3-abundant. Thus, by applying Theorem 2.9 we see that the K_3-coloured C_5 in the middle of Figure 1 is K_3-abundant.*

Proof of Theorem 2.9. Without loss of generality, we may assume that $V(F) = \{1, 2, \ldots, |F|\}$, $\{1, 2\}$ is an edge of F, $\sigma(u) = \sigma'(U') = 1$, and $\sigma(v) = \sigma(V') = 2$. Let G be an n-vertex graph that is uniformly ε-far from F-free. Let $(V_i : i \in V(F))$ be the corresponding F-partition of G, and let $\mathcal{C}$ be the corresponding collection of edge-disjoint copies of F. By deleting edges, we may assume that every edge of G is in a copy of F from $\mathcal{C}$.

Since (H, σ) is F-abundant, there are $\mathrm{poly}_1(\varepsilon) \cdot n^{|H|}$ copies of (H, σ) in G. Let $H_1 := H - u - v$. The *weight* of a copy of (H_1, σ) in G is the number of (H, σ) that contain it. This is just the number of $(x, y) \in V_1 \times V_2$ which when added to the copy of (H_1, σ) give a copy of (H, σ). The sum of the weights of the copies of (H_1, σ) is at least $\mathrm{poly}_1(\varepsilon) \cdot n^{|H|}$, the number of copies of (H_1, σ) is at most $n^{|H_1|} = n^{|H|-2}$, and the maximum weight of any single copy of (H_1, σ) is at most n^2. Hence, there are at least $\mathrm{poly}_2(\varepsilon) \cdot n^{|H|-2}$ copies of (H_1, σ) that have weight at least $\mathrm{poly}_3(\varepsilon) \cdot n^2$.

Fix a copy (H_1, σ) of weight at least $\mathrm{poly}_3(\varepsilon) \cdot n^2$ and call it $\mathcal{H}$. Consider the pairs $(x, y) \in V_1 \times V_2$ that extend $\mathcal{H}$ to a copy of (H, σ). Since $uv \in E(H)$, each such pair is an edge in G. So if $X \subset V_1$ and $Y \subset V_2$ are the end-vertices of these edges, there are $\mathrm{poly}_3(\varepsilon) \cdot n^2$ edges from X to Y. Crucially, every pair (x, y) with $x \in X$ and $y \in Y$ extends $\mathcal{H}$ to a copy of $(H - uv, \sigma)$.

Let $\mathcal{C}_1 = \{C \in \mathcal{C} : C \cap X \neq \varnothing \text{ and } C \cap Y \neq \varnothing\}$. Since there are $\mathrm{poly}_3(\varepsilon) \cdot n^2$ edges from X to Y and every edge of G is in a copy of F from $\mathcal{C}$, there are $\mathrm{poly}_3(\varepsilon) \cdot n^2$ edge-disjoint copies of F in $\mathcal{C}_1$. Let G_1 be the subgraph of G induced by $\mathcal{C}_1$. By Lemma 2.5, G_1 has a subgraph G' of order at least $\mathrm{poly}_4(\varepsilon) \cdot n$ that is uniformly $\mathrm{poly}_5(\varepsilon)$-far from F-free. Furthermore, the corresponding F-partition $(V_i' : i \in V(F))$ of G' satisfies $V_1' \subset X$, $V_2' \subset Y$, $V_3' \subset V_3$, $\ldots$, $V_{|F|}' \subset V_{|F|}$. Since (H', σ') is F-abundant, G' contains $\mathrm{poly}_6(\varepsilon) \cdot n^{|H'|}$ copies of (H', σ') and at most $|H| \cdot n^{|H'|-1}$ of these intersect $\mathcal{H}$. Each of the remaining copies extend $\mathcal{H}$ to a copy of (H'', σ''), since U' embeds into $V_1' \subset X$ and V' embeds into $V_2' \subset Y$.

Thus $\mathcal{H}$ extends to at least $\mathrm{poly}_7(\varepsilon) \cdot n^{|H'|}$ copies of (H'', σ''). There were $\mathrm{poly}_2(\varepsilon) \cdot n^{|H|-2}$ choices for $\mathcal{H}$ and so G contains at least $\mathrm{poly}_8(\varepsilon) \cdot n^{|H'|+|H|-2} = \mathrm{poly}_8(\varepsilon) \cdot n^{|H''|}$ copies of (H'', σ''). Thus (H'', σ'') is F-abundant. $\qquad\square$

Example 2.11. *The K_3-coloured Petersen graph depicted on the right in Figure 2 can be seen to be K_3-abundant by a combination of Theorem 2.7 and Theorem 2.9 as indicated in the figure.*

The left graph (without the square vertices) is 2-coloured and therefore K_3-abundant. Then, we add the square vertices to that graph according to Theorem 2.9 by replacing one edge each. Thus, the middle graph (without the circular top vertex) is abundant. Finally, we add the circular top vertex to that graph. The resulting graph, which is a coloured Petersen graph, is then abundant by Theorem 2.7. Notably, neither Theorem 2.7 nor Theorem 2.9 suffice alone.
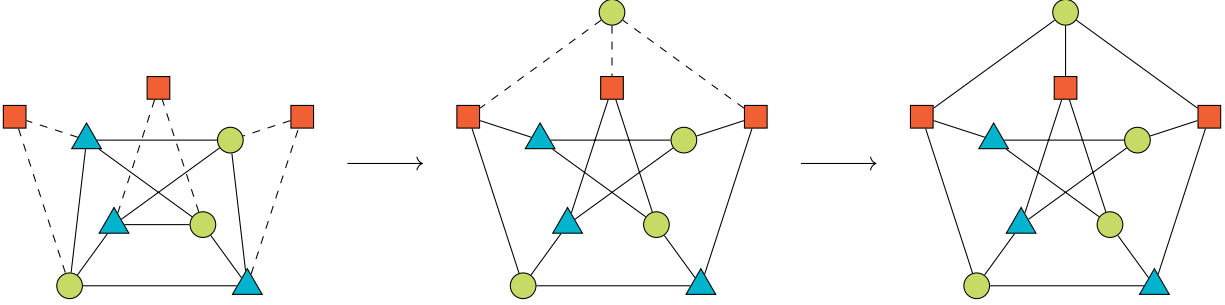

FIGURE 2. K_3-abundance of the Petersen graph.

Our third operation is a particularly useful combination of Theorem 2.7 and Theorem 2.9.

Theorem 2.12. *Let (H, σ) and (H', σ') be F-coloured graphs that are F-abundant. Let a, b, a' be vertices of F with $ab, aa' \in E(F)$. The coloured graph obtained from the disjoint union of (H, σ) and (H', σ') by adding all edges between vertices of colour a in H and colour a' in H' and all edges between vertices of colour b in H and colour a in H' is F-abundant.*

Proof. Let (H_1, σ_1) be the coloured graph obtained from (H, σ) by adding two vertices u and v where:

- u is joined to v,
- u is joined to all vertices of colour a in H,
- v is joined to all vertices of colour b in H,
- $\sigma(u) = a'$, and
- $\sigma(v) = a$.

Applying Theorem 2.7 to (H, σ) and adding the vertices v and u in that order shows that (H_1, σ_1) is F-abundant. Let $U' = \{x \in H' : \sigma'(x) = a'\}$ and $V' = \{x \in H' : \sigma'(x) = a\}$. Applying Theorem 2.9 to (H_1, σ_1) and (H', σ') gives the claimed result. $\square$

The case $a' = b$ in Theorem 2.12 is particularly clean, and yields the following family.

Definition 2.13. *A graph H is F-splittable if there is an F-colouring σ of H with the following property. Every subgraph H' of H has an edge cut-set whose edges span only two colours.*

From the previous theorem it follows that all splittable graphs are abundant.

Theorem 2.14. *Every F-splittable graph is F-abundant.*

Proof. Let H be F-splittable and σ the F-colouring that witnesses this. We prove that $(H', \sigma|_{V(H')})$ is F-abundant for every subgraph H' of H. We do this by induction on the number of edges. The result holds for all subgraphs with no edges.

Let H' be a subgraph of H with at least one edge. As H is F-splittable with witness σ, there are two colours a and b such that the set E_{ab} of edges in H' spanning these two colours is an edge cut-set of H'. Let H_1 and H_2 be the two sides of the cut. Since H' has at least one edge we can assume that it has an edge whose endpoints have colours a and b. In particular, H_1 and H_2 have fewer edges than H' and so, by induction, $(H_1, \sigma|_{V(H_1)})$ and $(H_2, \sigma|_{V(H_2)})$ are F-abundant. Applying Theorem 2.12 with $a' = b$ shows that $(H', \sigma|_{V(H')})$ is F-abundant. $\qquad\square$

2.3. F-splittable graphs with chromatic number $\chi(F)$.

In this subsection we prove Theorem 1.1. We first define an operation that converts an F-abundant graph into a more complicated one. Let σ be a surjective F-colouring of a graph H. For $e = ab \in E(F)$ we define $(H, \sigma)_e$ as the disjoint union of two copies of (H, σ), say (H_1, σ_1) and (H_2, σ_2), where all vertices in $\sigma_1^{-1}(a)$ are joined to all vertices in $\sigma_2^{-1}(b)$ and all vertices in $\sigma_1^{-1}(b)$ are joined to all vertices in $\sigma_2^{-1}(a)$. If (H, σ) is F-splittable, then so is $(H, \sigma)_e$.

Using this operation, we now construct a sequence of increasingly complicated F-splittable graphs. Let $e_1, \ldots e_{e(F)}$ be an enumeration of the edges of F. We start with an F-splittable F-coloured graph (H, σ) where σ is surjective (for example, one could take H to be $|F|$ isolated vertices and σ a bijection from $V(H)$ to $V(F)$). We define a sequence of graphs $(H^1, \sigma_1) = (H, \sigma)$, (H^2, σ_2), $\ldots$, where $(H^{j+1}, \sigma_{j+1}) := (H^j, \sigma_j)_{e_{j \pmod{e(F)}}}$ for each j. By Theorem 2.14, the graph H^m is F-abundant for every positive integer m.

Lemma 2.15. *Let G be a graph with $F \nrightarrow G$. Then, for all sufficiently large m, $H^m \nrightarrow G$.*

Proof. Take m sufficiently large in terms of F and G. Suppose for a contradiction that there is a homomorphism φ from H^m to G. We view φ as a G-colouring of H^m. For all $v \in V(F)$, let $S^m(v)$ be the set of colours appearing on the vertices corresponding to v in H^m. Note that each colour is a vertex of G. We denote the two canonical copies of H^i within H^{i+1} by H_1^i and H_2^i.

Now look inside $H^m = H_1^{m-1} \cup H_2^{m-1}$. For $v \in V(F)$ and $i \in \{1, 2\}$, let $S_i^{m-1}(v)$ be the set of colours appearing on vertices corresponding to v in H_i^{m-1}. Note that $S^m(v) = S_1^{m-1}(v) \cup S_2^{m-1}(v)$. Compare $\sum_{v \in V(F)}|S_1^{m-1}(v)|$ and $\sum_{v \in V(F)}|S_2^{m-1}(v)|$ and let $i \in \{1, 2\}$ be such that $\sum_{v \in V(F)}|S_i^{m-1}(v)|$ is the smaller (or equal) of the two sums. Define $H^{m-1} := H_i^{m-1}$ and $S^{m-1}(v) := S_i^{m-1}(v)$ for each $v \in V(F)$. Now look inside $H^{m-1} = H_1^{m-2} \cup H_2^{m-2}$ and iterate.

Doing this produces a sequence of graphs $H^m \supset H^{m-1} \supset \cdots \supset H_1$ and, for each $v \in V(F)$, a sequence of sets of colours $S^m(v) \supset S^{m-1}(v) \supset \cdots \supset S^1(v)$ where $\sum_{v \in V(F)}|S^j(v)|$ is the minimum of $\sum_{v \in V(F)}|S_1^j(v)|$ and $\sum_{v \in V(F)}|S_2^j(v)|$.

We apply the pigeonhole principle to sequence of k-tuples $(S^j(v) : v \in V(F))$ where j ranges from 1 to m. Provided m is sufficiently large, there will be more than $e(F)$ indices j and some tuple $(S(v) : v \in V(F))$ such that $S_j(v) = S(v)$ for all $v \in V(F)$ and for all such j. Note that,

12

crucially, m just depends on F and G. Since for each v the sequence $S^j(v)$ is nested, we in fact will have an interval $I \subset \{1, 2, \ldots, m\}$ of length greater than $e(F)$ such that $S_j(v) = S(v)$ for all $v \in V(F)$ and $j \in I$.

Now, F is not homomorphic to G and each $S(v)$ is a non-empty set of colours (non-empty set of vertices of G). So, there must be adjacent vertices $x, y \in V(F)$ and non-adjacent vertices $c, d \in V(G)$ such that $c \in S(x)$ and $d \in S(y)$. Since I is an interval of length greater than $e(F)$, there is a j such that j and $j+1$ are both in I and $e_{j \pmod{e(F)}} = xy$. We assume without loss of generality that $H^j = H_1^j$, so $S_1^j(v) = S^j(v) = S(v)$ for all $v \in V(F)$, and

$$\sum_{v \in V(F)} |S_1^j(v)| \leqslant \sum_{v \in V(F)} |S_2^j(v)|. \tag{2}$$

Now, for all $v \in V(F)$, $S_2^j(v) \subset S^{j+1}(v) = S(v) = S_1^j(v)$. Thus, by (2), we must have $S_2^j(v) = S_1^j(v) = S(v)$ for all $v \in V(F)$. Since $c \in S(x) = S_1^j(x)$, there is a vertex $u_x \in V(H_1^j)$ corresponding to x with $\varphi(u_x) = c$. Similarly, since $d \in S(y) = S_2^j(y)$, there is a vertex $u_y \in V(H_2^j)$ corresponding to y with $\varphi(u_y) = d$.

Since $e_{j \pmod{e(F)}} = xy$, every vertex in H_1^j corresponding to x is adjacent to every vertex in H_2^j corresponding to y and so u_x is adjacent to u_y. Since φ is a homomorphism to G, $\varphi(u_x) = c$ must be adjacent to $\varphi(u_y) = d$. However, c and d were chosen to not be adjacent, which is the required contradiction. $\qquad\square$

Since the graph H^m is F-splittable by construction, this proves Theorem 1.1 and shows that F-abundance cannot be characterised by looking at right homomorphisms. We combine this with the following result of Nešetřil and Zhu.

Theorem 2.16 ([NZ04, Thm. 1.1]). *For every graph H' and for all positive integers k and ℓ there exists a graph H with the following properties:*

- *H has girth greater than ℓ, and*
- *for every graph G on at most k vertices, there is a homomorphism $H \to G$ if and only if there is a homomorphism $H' \to G$.*

Corollary 2.17. *Let ℓ be a positive integer and F and G be graphs with $F \nrightarrow G$. There is an F-abundant graph H with girth greater than ℓ and $H \nrightarrow G$.*

Proof. By Theorem 1.1, there is an F-abundant graph H' with $H' \nrightarrow G$. Applying Theorem 2.16 with $k = \max\{|H'|, |G|\}$ gives a graph H with girth greater than ℓ, $H \to H'$, and $H \nrightarrow G$. Since H' is F-abundant and $H \to H'$, the graph H is also F-abundant. $\qquad\square$

2.4. **Effective Abundance.** Note that in the above arguments we were interested in showing abundance without paying too much attention to the degree of the polynomial in ε. In some cases, we can actually show much better bounds on the degrees of these polynomials. This might be useful in the setting of very sparse graphs. Below, we show how to do this in the most

basic case of the triangle-abundance of C_5. First, we show that we can find many paths with four vertices of a certain type in a graph that is far from triangle-free.

Proposition 2.18. *Let $p \geqslant 100/n$ and let G be an n-vertex tripartite graph with tripartition $V(G) = A \cup B \cup C$ that contains at least pn^2 edges-disjoint triangles. Then, G contains at least $\Omega(p^3 n^4)$ paths of length 3 of the form $x_1 x_2 x_3 x_4$ with $x_1, x_4 \in A$, $x_2 \in B$, and $x_3 \in C$.*

Proof. We may assume that $E(G)$ is the disjoint union of pn^2 edge-disjoint triangles. Observe that for every $x \in A$, $d_B(x) = d_C(x)$ (and analogously for B and C).

Simple counting gives that the number of P_3's with middle edge in $G[B, C]$ is at least

$$\sum_{(x,y) \in E(B,C)} d_A(x)(d_A(y) - 1) \geqslant \sum_{(x,y) \in E(B,C)} (d_C(x) - 1)(d_B(y) - 1) = \#P_3\text{'s in } G[B, C].$$

Using the fact that P_3 is Sidorenko (see e.g. [Zha23, Theorem 5.5.10]) and $E(G[B, C]) = pn^2$, there must be at least $\Omega(p^3 n^4)$ paths of length 3 as required. $\qquad\square$

We now use a strategy similar to Theorem 2.7 for extending this result to show that a graph that is far from triangle-free contains many copies of C_5.

Lemma 2.19. *Let $100 n^{-1/2} \leqslant p \leqslant n^{-\gamma}$ for some $\gamma > 0$. Then, for every $\varepsilon' > 0$, there exists $C(\varepsilon') > 0$ such that every n-vertex graph G which is p-far from being triangle-free contains $C(\varepsilon') p^{6+\varepsilon'} n^5$ copies of C_5.*

Proof. Let $\varepsilon = \varepsilon' \gamma$ and choose $\delta > 0$ sufficiently small. We randomly partition the vertices of G into three parts of equal size. The probability that a triangle of G has one vertex in each part of the partition is $2/9 > 1/5$, and so there exists a partition $V(G) = A \cup B \cup C$ such that G has $(p/5)n^2$ edge-disjoint triangles with one vertex in each part. We delete all other edges from the graph. For a graph H, denote by $m(H)$ the number of edge-disjoint triangles in H.

We now perform the following iterative process. We start with $A_0 = A$ and $p_0 = m(G)/(n|A|)$. Then, given $A_i \subset A$ and p_i, let $A_{i+1} \subset A_i$ be the set of those vertices that are contained in at least $\delta p_i n$ edge-disjoint triangles, and let $p_{i+1} = m(G[A_{i+1} \cup B \cup C])/(n|A_{i+1}|)$. If $|A_{i+1}| \geqslant \delta|A_i|$, we stop at this step, and otherwise we continue.

Define $G_i = G[A_i \cup B \cup C]$, so $p_i = m(G_i)/(n|A_i|)$. Note that $|A_i| \leqslant \delta^i n$. Moreover,

$$m(G_{i+1}) \geqslant m(G_i) - \delta p_i n |A_i| = (1 - \delta)m(G_i)$$

and so $m(G_i) \geqslant (1 - \delta)^i m(G) \geqslant (1 - \delta)^i (p/5)n^2$.

Claim 2.20. *The process stops after $t \leqslant \log(n)/\log(1/\delta)$ steps.*

Proof. Indeed, after $i > \log(n)/\log(1/\delta)$ steps, we have $|A_i| = \delta^i n < 1$ and so $|A_i| = 0$. However, $m(G_i) \geqslant (1 - \delta)^i (p/5)n^2 > 0$ which gives a contradiction. $\qquad\square$

14

Note that if δ is sufficiently small, then $m(G_t) \geqslant (1-\delta)^t (p/5)n^2 \geqslant (p/5)n^{2-\varepsilon/9}$. In particular, $p_t \geqslant (p/5)n^{1-\varepsilon/9}/|A_t|$, and so every vertex in A_{t+1} is contained in at least $(\delta p/5)n^{2-\varepsilon/9}/|A_t|$ edge-disjoint triangles. Now, we apply the exact same argument to G_{t+1} but with B instead of A. This yields a graph $G' = G[A_{t+1} \cup B_{s+1} \cup C]$ for some subset $B_{s+1} \subset B$ with $m(G') \geqslant (p/5)n^{2-2\varepsilon/9}$ such that every vertex in B_{s+1} is contained in at least $(\delta p/5)n^{2-2\varepsilon/9}/|B_s|$ edge-disjoint triangles and $|B_{s+1}| \geqslant \delta|B_s|$.

Now, for each $x \in B_{s+1}$, consider the graph $G'' = G[(N_G(x) \cap A_{t+1}) \cup B \cup C]$. Note that $|N_G(x) \cap A_{t+1}| \geqslant (\delta p/5)n^{2-2\varepsilon/9}/|B_s|$ since x is contained in at least that many edge-disjoint triangles. So, by construction, we get $m(G'') \geqslant (\delta p/5)n^{2-2\varepsilon/9}/|B_s| \cdot (\delta p/5)n^{2-\varepsilon/9}/|A_t| = qn^2$ where $q = (\delta p/5)^2 n^{2-\varepsilon/3}/(|A_t||B_s|)$. Applying Proposition 2.18 shows that there are at least $\Omega(q^3 n^4)$ paths in G'', each of which can be extended by x to a copy of C_5. Since $x \in B_{s+1}$ was arbitrary, it follows that in total there are at least

$$\Omega(q^3 n^4 |B_{s+1}|) = \Omega\left(\frac{\delta^6 p^6 n^{10-\varepsilon} |B_{s+1}|}{|A_t|^3 |B_s|^3}\right) \geqslant \Omega(\delta^7 p^6 n^{5-\varepsilon}) \geqslant \Omega(\delta^6 p^{6+\varepsilon'} n^5) = C(\varepsilon') p^{6+\varepsilon'} n^5$$

copies of C_5 in G. $\qquad\square$

We remark that in the special case that p is of the order $n^{-1/2}$ the best results known gives on the order of n^4 copies of C_5, which is tighter by a factor of $n^{\varepsilon'/2}$. The tightest result is due to Conlon, Fox, Sudakov and Zhao [CFSZ21], and makes use of sparse regularity. They also give a construction of a graph containing $o(n^{2.442})$ copies of C_5 that cannot be made triangle free by deleting $o(n^{3/2})$ edges. This shows in particular that one cannot replace $p^{6+\varepsilon'}$ by p^5 in Lemma 2.19. It would be interesting to have similar constructions that work for larger p.

3. ADDITIVE RESULTS

In this section, we turn to a problem that is seemingly unrelated to abundance. We ask how large sets of integers can be without having solutions to certain equations, proving Theorem 1.2. To be precise, consider an equation E of the form

$$a_1 x_1 + \cdots + a_k x_k = 0$$

with $a_i \neq 0$ for $i \in [k]$ but $a_1 + \cdots + a_k = 0$. Ruzsa [Ruz93] studied the following two quantities:

- the size $R_E(N)$ of a largest set $\mathcal{A} \subset [N]$ such that equation E has no solutions with distinct integers $x_1, \ldots, x_k \in \mathcal{A}$, and
- the size $r_E(N)$ of a largest set $\mathcal{A} \subset [N]$ with only trivial solutions to equation E, where a solution $x_1, \ldots, x_k \in \mathcal{A}$ is called *trivial* if all maximal subsets $T \subset [k]$ with $x_i = x_j$ for all $i, j \in T$ satisfy $\sum_{i \in T} a_i = 0$.

Clearly, $r_E(N) \leqslant R_E(N)$. In Section 4, we explain why these quantities are important for abundance, but in short we want to know which equations satisfy $R_E(N) = N^{1-o(1)}$. In this context, the genus of an equation is important.

Definition 3.1. *The* genus *of equation E is the largest integer m such that there is a partition $\mathcal{T}_1 \cup \cdots \cup \mathcal{T}_m$ of the set of indices $[k]$ into m disjoint nonempty sets $\mathcal{T}_j$ such that, for every j,*

$$\sum_{i \in \mathcal{T}_j} a_i = 0.$$

Ruzsa proved the following.

Theorem 3.2. *For any equation E of genus m we have*

$$r_E(N) = \mathcal{O}(N^{1/m}).$$

Unfortunately, he did not prove the same result for $R_E(N)$. If the equation is *symmetric*, which means that it can be written as

$$\sum_{i \in [\ell]} a_i x_i = \sum_{i \in [\ell]} a_i x_{\ell+i},$$

then $r_E(N) = \mathcal{O}(N^{1/\ell})$, since symmetric equations in 2ℓ variables have genus ℓ. Moreover, Ruzsa showed the following.

Theorem 3.3. *Let $\ell \geqslant 2$. For a symmetric equation E in 2ℓ variables we have*

$$R_E(N) = \mathcal{O}(\sqrt{N}).$$

Ruzsa also showed that this bound on $R_E(N)$ is the best possible *general* bound for symmetric equations. Bukh [Buk08] later improved this bound for individual symmetric equations.

Based on these results, Ruzsa [Ruz93, § 9] asked whether $r_E(N) \geqslant N^{1/m-o(1)}$ holds for all equations of genus m. In that case, equations of genus one would satisfy $R_E(N) \geqslant r_E(N) \geqslant N^{1-o(1)}$. However, so far, the only equations known to have $R_E(N) = N^{1-o(1)}$ are *convex* equations, which are equations with exactly one negative coefficient. The corresponding set $\mathcal{A}$ can then be obtained by adapting Behrend's construction [Beh46] of a large set with only trivial arithmetic progressions.

Note that according to the above results, it was still possible that some nonsymmetric equation of genus at least two satisfied $R_E(N) \geqslant N^{1-o(1)}$. We make progress on this by proving that in fact $R_E(N) = \mathcal{O}(\sqrt{N})$ holds for all equations of genus at least two; this is Theorem 1.2.

We first sketch the idea of this proof for the equation

$$x_1 - x_2 + y_1 + y_2 - 2y_3 = 0.$$

Assume that $\mathcal{A} \subset [N]$ is a subset of size $|\mathcal{A}| \geqslant C\sqrt{N}$. To find a solution to the above equation with distinct integers from $\mathcal{A}$, we will construct an auxilliary graph. This will be a 4-partite graph whose parts are $\mathcal{A}$, $\mathcal{A} + \mathcal{A}$, $\mathcal{A} + 2\mathcal{A}$, and $2\mathcal{A}$, and for all $x, y \in \mathcal{A}$ we will add the edges of the following path to our graph:

$$x, \quad x+y, \quad x+2y, \quad 2y.$$

In total, this adds C^2N many paths and edges to the graph. So, if C is sufficiently large, the average degree of this graph will be larger than any constant that we want, and we may pass to a subgraph where the minimum degree is larger than any constant that we want.

Now, by construction, any path across the four parts of this subgraph will be of the form

$$x_1, \quad x_1 + y_1, \quad x_1 + y_1 + y_2, \quad x_1 - x_2 + y_1 + y_2$$

where $x_1, x_2, y_1, y_2 \in \mathcal{A}$. Moreover, since the vertex $x_1 - x_2 + y_1 + y_2$ is in the last part of the graph, we must have $x_1 - x_2 + y_1 + y_2 \in 2\mathcal{A}$ and so $x_1 - x_2 + y_1 + y_2 = 2y_3$ for some $y_3 \in \mathcal{A}$. But this means that $x_1 - x_2 + y_1 + y_2 - 2y_3 = 0$, which implies that such a path corresponds to a solution of the equation with integers from $\mathcal{A}$. It only remains to ensure that all integers are distinct. But this is not difficult: if we choose the path in the graph incrementally, every outgoing edge of a vertex corresponds to different value for one of the variables. Since the degree of every vertex is large, this means we can simply choose an edge where the value of the new variable is different from the values of all previous variables, which then yields a solution to the equation with distinct integers from $\mathcal{A}$. This concludes the sketch. In the general case, the proof is simply a generalisation of this argument with a few more details added.

Remark. This argument is essentially a translation of the proof of Theorem 2.9 to the additive setting. One could view it as replacing an edge of a C_4 by path on 3 vertices. The essential difference is that the additive structure allows one to find the optimal number of edges to be replaced by disjoint paths.

Theorem 3.4. *Let E be an equation of genus at least two. Then, $R_E(N) = \mathcal{O}(\sqrt{N})$.*

Proof. Since E is an equation of genus at least two, we may write E as

$$\sum_{i=1}^{s} a_i x_i + \sum_{j=1}^{t} b_j y_j = 0$$

where $\sum_{i=1}^{s} a_i = \sum_{j=1}^{t} b_j = 0$. Let $\mathcal{A} \subset [N]$ and write $M = |\mathcal{A}|$. We want to show that if $M > C\sqrt{N}$ for some sufficiently large constant C depending only on E, then there exists a solution to E with distinct integers from $\mathcal{A}$.

For this, we define an auxiliary $(s + t - 1)$-partite graph G as follows. Its vertex sets $V_1, \ldots, V_{s+t-1}$ are defined as

$$V_k = a_1 \mathcal{A} + \sum_{j=1}^{k-1} b_j \mathcal{A} \text{ for } k \in [t-1] \quad \text{and} \quad V_{k+t-1} = \sum_{i=k+1}^{s} (-a_i \mathcal{A}) - b_t \mathcal{A} \text{ for } k \in [s].$$

For every ordered pair (x, y) of distinct $x, y \in \mathcal{A}$, we add a $V_1 V_2 \ldots V_{s+t-1}$ path to G whose vertices $v_1 v_2 \ldots v_{s+t-1}$ are given by

$$v_k = a_1 x + \sum_{j=1}^{k-1} b_j y \text{ for } k \in [t-1] \quad \text{and} \quad v_{k+t-1} = \sum_{i=k+1}^{s} (-a_i x) - b_t y \text{ for } k \in [s],$$
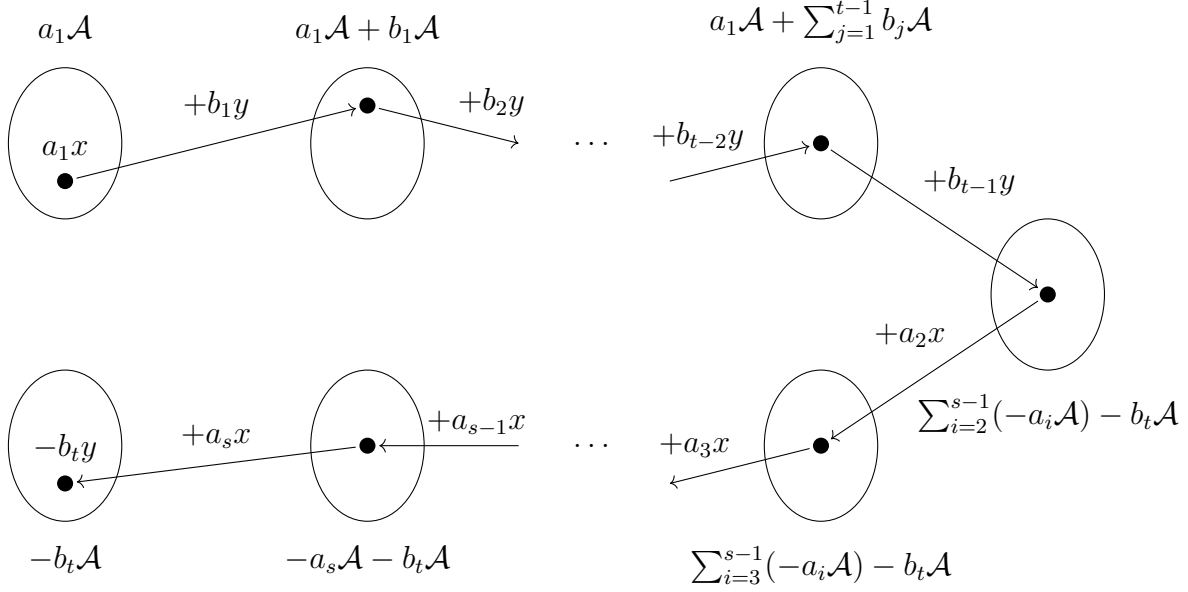
as shown in Figure 3.

FIGURE 3. The path for the ordered pair (x, y) of distinct $x, y \in \mathcal{A}$.

We observe that $v_1 = a_1 x$, $v_{k+1} = v_k + b_k y$ for $k \in [t-1]$, and $v_{k+t} = v_{k+t-1} + a_{k+1} x$ for $k \in [s-1]$. This is obvious except for $v_t = v_{t-1} + b_{t-1} y$, but there we have

$$v_t = \sum_{i=2}^{s} (-a_i x) - b_t y = a_1 x + \sum_{j=2}^{t-1} b_j y = v_{t-1} + b_{t-1} y,$$

where we used $\sum_{i=1}^{s} a_i = \sum_{j=1}^{t} b_j = 0$. In particular, the difference between the two endpoints of any edge in this path uniquely determines either x or y. Moreover, once we have one of these two values, we can determine the other value by looking at either endpoint of the edge. So, from any edge we can determine both x and y, which implies that all of these paths are edge-disjoint. We denote the collection of these paths by $\mathcal{P}$, so we have $|\mathcal{P}| = M(M-1) \geqslant M^2/2$.

We claim that any $V_1 V_2 \ldots V_{s+t-1}$ path in G yields a solution to the equation E. Indeed, if $u_1 u_2 \ldots u_{s+t-1}$ is such a path, then by the observation from above there are $x_i, y_j \in \mathcal{A}$ with $u_1 = a_1 x_1$, $u_{k+1} = u_k + b_k y_k$ for $k \in [t-1]$, and $u_{k+t} = u_{k+t-1} + a_{k+1} x_{k+1}$ for $k \in [s-1]$. Moreover, $u_{s+t-1} = -b_t y_t$ for some $y_t \in \mathcal{A}$ by definition of V_{s+t-1}. Therefore,

$$\sum_{i=1}^{s} a_i x_i + \sum_{j=1}^{t} b_j y_j = a_1 x_1 + \sum_{j=1}^{t-1} b_j y_j + \sum_{i=2}^{s} a_i x_i + b_t y_t = u_{s+t-1} + b_t y_t = 0,$$

as claimed. It remains to show that there exists such a path where the corresponding variables are distinct.

The graph G has at most $n = cN$ many vertices where $c = (s+t-1)(\sum_{i=1}^{s} |a_i| + \sum_{j=1}^{t} |b_j|)$. We pass to a subgraph H of G as follows. While there exists a vertex v of G that is contained in less than $|\mathcal{P}|/2n$ many paths of $\mathcal{P}$, we delete all paths containing v. Since this deletes at most $n \cdot (|\mathcal{P}|/2n) = |\mathcal{P}|/2$ many paths, we are left with at least $|\mathcal{P}|/2$ many paths $\mathcal{P}' \subset \mathcal{P}$ and every vertex is contained in either none or at least $|\mathcal{P}|/2n$ many of these paths. Let H

be the (non-empty) subgraph of G induced by the vertices in $\mathcal{P}'$. In particular, every vertex $v \in V(H) \cap V_k$ for $k \in [s+t-2]$ has at least $|\mathcal{P}|/2n \geqslant M^2/4cN \geqslant C^2/4c$ many neighbours in $V(H) \cap V_{k+1}$. If C is sufficiently large, this is at least $2(s+t)$.

We now construct the path $u_1 u_2 \ldots u_{s+t-1}$. We start with any vertex $u_1 = a_1 x_1 \in V(H) \cap V_1$. Then, for $k \in [t-1]$, we pick a neighbour $u_{k+1} = u_k + b_k y_k \in V(H) \cap V_{k+1}$ of u_k such that y_k is distinct from $x_1, y_1, \ldots, y_{k-1}$. Since u_k has more than k neighbours in $V(H) \cap V_{k+1}$ and every neighbour corresponds to a different value of y_k, this is always possible. Afterwards, for $k \in [s-2]$, we pick a neighbour $u_{k+t} = u_{k+t-1} + a_{k+1} x_{k+1} \in V(H) \cap V_{k+t}$ of u_{k+t-1} such that x_{k+1} is distinct from $x_1, \ldots, x_k, y_1, \ldots, y_{t-1}$. Again, u_{k+t-1} has more than $k+t-1$ neighbours in $V(H) \cap V_{k+t}$, so this is possible. Finally, we pick a neighbour $u_{s+t-1} = u_{s+t-2} + a_s x_s = -b_t y_t \in V(H) \cap V_{s+t-1}$ of u_{s+t-2} such that x_s and y_t are both distinct from $x_1, \ldots, x_{s-1}, y_1, \ldots, y_{t-1}$. As u_{s+t-2} has more than $2(s+t-2)$ many neighbours in $V(H) \cap V_{s+t-1}$, this is possible. Also, x_s and y_t are distinct: this is true for all edges between V_{s+t-2} and V_{s+t-1} in G since all paths in $\mathcal{P}$ were constructed with x and y distinct. Thus, $x_1, \ldots, x_s, y_1, \ldots, y_t$ are all distinct, and we know from above that

$$\sum_{i=1}^{s} a_i x_i + \sum_{j=1}^{t} b_j y_j = 0. \qquad \square$$

In fact, the above proof immediately gives something stronger. We say an equation E in ℓ variables is *abundant* if there exists constants $\gamma, C > 0$, depending only on E, such that for all $\varepsilon \in (0, 1)$ and $N \in \mathbb{N}$, any subset $\mathcal{A} \subset [N]$ of order at least εN contains at least $\gamma \varepsilon^C N^{\ell-1}$ solutions to E. Note that abundance implies unavoidability. The reverse implication is an interesting open question (see Section 5).

Corollary 3.5. *All equations of genus at least* 2 *are abundant.*

Proof. Suppose $\mathcal{A} \subset [N]$ and $|\mathcal{A}| = M = \varepsilon N$, and let E be an equation of genus at least 2. Then we proceed as in the above proof to find an $(s+t-1) = (\ell-1)$-partite subgraph $(V_1, \ldots, V_{\ell-1})$, with the property that each $v \in V_i$ has at least $M^2/4cN > \varepsilon^2 N/4c$ neighbours in V_{i-1} and V_{i+1}. Thus, V_1 has at least $\varepsilon^2 N/4c$ vertices and choosing the first vertex of the path arbitrarily and then extending arbitrarily yields $(\varepsilon^2 N/4c)^{\ell-1}$ paths via $V_1 V_2 \ldots V_{\ell-1}$. As c only depends on E, this completes the proof. $\qquad \square$

4. FROM GRAPHS TO EQUATIONS AND BACK

Thus far we discussed abundance and the problem of integer solutions to equations largely in isolation. In reality, these two subjects are intimately connected. This has been known since at least the work of Ruzsa and Szemerédi [RS78] who used Behrend's construction [Beh46] of a large set with only trivial arithmetic progressions to show that triangles are not triangle-abundant. At the moment, the Ruzsa-Szemerédi construction is the only technique we have for showing that a graph H with $H \to F$ is not F-abundant. So far, however, it remained possible that this is simply a convenient way of proving non-abundance and nothing more. In this section, we

explore this connection in more depth and show that in some cases the avoidability of equations associated to H can imply abundance and even characterise it. We also show that a strategy for proving non-abundance suggested by Gishboliner, Shapira and Wigderson [GSW23] cannot succeed, and that their random construction of a strongly genus one graph can be replaced by a small explicit graph.

We begin by explaining the Ruzsa-Szemerédi construction in full generality. Fix an injection $c\colon V(F) \to [n]$. Then, for any integer N and any subset $\mathcal{A} \subset [N]$, the *Ruzsa-Szemerédi graph* $\mathrm{RS}(F, N, \mathcal{A})$ is the $|F|$-partite graph with vertex sets $V_v = [nN]$ for $v \in V(F)$ whose edges between V_u and V_v for $uv \in E(F)$ are those pairs $xy \in V_u \times V_v$ that satisfy $y - x \in (c(v) - c(u))\mathcal{A}$.

It is easy to see that this graph contains many edge-disjoint copies of F. Indeed, let $x \in [N]$ and $a \in \mathcal{A}$ be arbitrary. Then, for all $v \in V(F)$, let $x_v = x + (c(v) - 1)a \in V_v$. This set of vertices induces a copy of F, and every copy so generated is edge-disjoint. Therefore, this graph is far from F-free if $|\mathcal{A}|$ is sufficiently large. If we want to show that H is not F-abundant, we need to ensure that this graph also contains few (not necessarily edge-disjoint) copies of H. For that, consider an F-colouring σ of H and associate to any cycle $v_1 \ldots v_\ell$ of H its *cycle-equation*

$$(c(\sigma(v_2)) - c(\sigma(v_1)))x_{v_1 v_2} + \cdots + (c(\sigma(v_\ell)) - c(\sigma(v_{\ell-1})))x_{v_{\ell-1} v_\ell} + (c(\sigma(v_1)) - c(\sigma(v_\ell)))x_{v_\ell v_1} = 0.$$

We denote this system of equations by $\mathcal{E}(H, \sigma, c)$. In every embedding of (H, σ) in the Ruzsa-Szemerédi graph, the embedding of a cycle corresponds to a solution of its cycle-equation with integers from $\mathcal{A}$. So, if $\mathcal{A}$ has few solutions to $\mathcal{E}(H, \sigma, c)$, the graph contains few copies of (H, σ).

With some additional work, it can be shown that the Ruzsa-Szemerédi construction implies:

> If (H, σ) is F-abundant, then $\mathcal{E}(H, \sigma, c)$ is abundant.

Conversely, if $\mathcal{E}(H, \sigma, c)$ is not abundant, this shows that (H, σ) is not F-abundant. Thus, Corollary 3.5 is directly relevant to our investigations of graph abundance. It implies that a single equation of genus at least two does not suffice for the Ruzsa-Szemerédi construction.

4.1. **From graphs to equations.** Ruzsa's question, Question 1.3, asks whether all equations of genus one satisfy $R_E(N) = N^{1-o(1)}$. Under the assumption that the answer is yes, Gishboliner, Shapira, and Wigderson [GSW23] showed that every K_3-colouring of an appropriate triangle-free random graph G has an associated system of equations which can be linearly combined to obtain an equation with genus one, implying that G is not K_3-abundant (see [GSW23, Section 3.2]).

We explicitly construct a graph with this property. Let G_n be the 3-partite graph with vertex sets $A = [n]$, $B = 2^{[n]}$, and $C = 2^{[n]}$. Then, join $a \in A$ to exactly those $X \in B$ and $Y \in C$ that satisfy $a \in X$ and $a \in Y$. Moreover, join $X \in B$ to $Y \in C$ if and only if $X \cap Y = \varnothing$. It is easy to see that this graph is triangle-free. Moreover, it can be shown that each tripartite triangle-free graph is a subgraph of G_n for n sufficiently large. Thus, if any tripartite triangle-free graph is not K_3-abundant, then G_n will not be K_3-abundant for some sufficiently large n.

In fact, a computer can check that the subgraph of G_3 depicted in Figure 4 has the property that every K_3-colouring has a linear combination with genus one. So, this subgraph (and therefore G_3) is not K_3-abundant if Ruzsa's question has a positive answer.
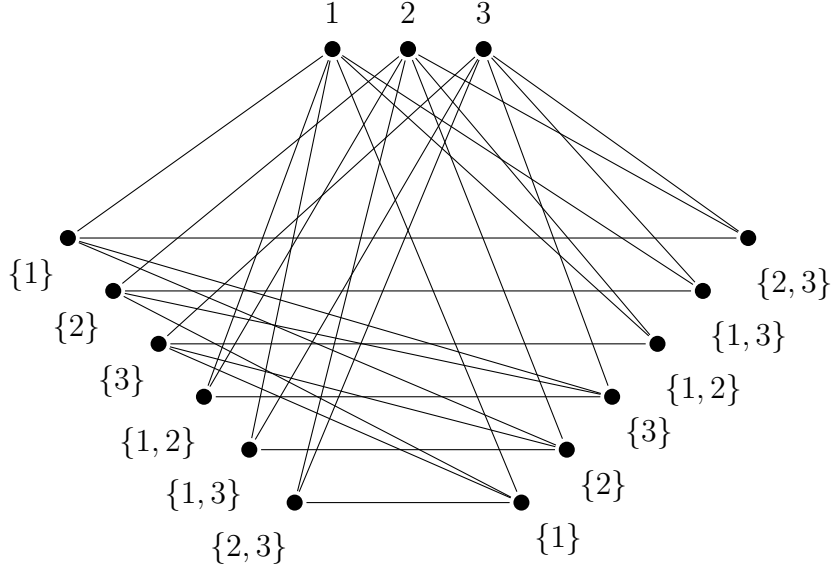

FIGURE 4. A subgraph of G_3 obtained by removing the empty and complete sets.

It would be interesting to prove that some triangle-free graph is not K_3-abundant without relying on the assumption that Ruzsa's question has a positive answer. Unfortunately, as mentioned in Section 3, the only equations known to have $R_E(N) = N^{1-o(1)}$ are convex equations, which are equations with exactly one negative coefficient. Gishboliner, Shapira, and Wigderson therefore suggested trying to find a triangle-free graph where every K_3-colouring has an associated system that can be linearly combined to give a convex equation. This would unconditionally prove that such a graph is not K_3-abundant. They tried to find such a graph with a computer search, but were unsuccessful.

We show that this cannot work. Indeed, we prove that if some linear combination of the equations from $\mathcal{E}(H, \sigma, c)$ is convex, then there is a single cycle that yields a convex equation. In H, the colours associated to such a cycle by $c \circ \sigma$ are in strictly increasing order, so the cycle has length at most 3 and must therefore be a triangle. In general, this means that we can currently only show that a graph H is not F-abundant if every F-colouring of H has an increasing cycle, a property that is called *increasing-cycle-unavoidable* (see [GSW23, Section 1.5]).

Lemma 4.1. *Let (H, σ) be an F-coloured graph and $c\colon V(F) \to [n]$ be an injection. If some linear combination of the equations from $\mathcal{E}(H, \sigma, c)$ is convex, then there is a single cycle-equation of H that is convex.*

Proof. Let

$$\sum_{e \in E(H)} a_e x_e = 0$$

be a linear combination of the cycle-equations from $\mathcal{E}(H, \sigma, c)$ that is convex. Let $v \in V(H)$ be any vertex. Then, we claim that

$$S(v) = \sum_{u \in N(v)} \frac{a_{vu}}{c(\sigma(u)) - c(\sigma(v))} = 0.$$

Indeed, this holds for any single cycle-equation: if v appears in the cycle, the contribution to $S(v)$ of the two edges incident to v in the cycle cancel each other out which means that $S(v)$ is zero as claimed. So this holds for a single cycle-equation, and therefore also for any linear combination of them.

Claim 4.2. *Let v be such that $a_{vu} \geqslant 0$ for all $u \in N(v)$. Then, if some $u \in N(v)$ satisfies $c(\sigma(u)) < c(\sigma(v))$ and $a_{vu} > 0$, then some $u' \in N(v)$ satisfies $c(\sigma(u')) > c(\sigma(v))$ and $a_{vu'} > 0$.*

Proof. The contribution of u to $S(v)$ is negative. Since $S(v) = 0$, there must therefore exist some $u' \in N(v)$ such that the contribution of u' to $S(v)$ is positive. Given that $a_{vu'} \geqslant 0$, it must hold that $a_{vu'} > 0$ and thus also $c(\sigma(u')) - c(\sigma(v)) > 0$. This proves the claim. □

Now, let $vu \in E(H)$ be such that $c(\sigma(v)) < c(\sigma(u))$ and $a_{vu} < 0$. Note that the contribution of u to $S(v)$ is negative. Since $S(v) = 0$, there must therefore exist some $w \in N(v) \setminus \{u\}$ whose contribution to $S(v)$ is positive. Since the linear combination of the cycle-equations is convex, a_{vu} is the only coefficient that is negative, and so we must have $a_{vw} > 0$. In order for the contribution of w to $S(v)$ to be positive, it then follows that $c(\sigma(w)) > c(\sigma(v))$.

Pick a maximal path $v_1 \ldots v_\ell$ with $v_1 = v$, $c(\sigma(v_1)) < \cdots < c(\sigma(v_\ell))$, and $a_{v_i v_{i+1}} > 0$ for all $i \in [\ell - 1]$. The path vw shows that $\ell \geqslant 2$. If $v_\ell \neq u$, then we know that $a_{v_\ell u'} \geqslant 0$ for all $u' \in N(v_\ell)$. Moreover, note that $v_{\ell-1} \in N(v_\ell)$ satisfies $c(\sigma(v_{\ell-1})) < c(\sigma(v_\ell))$ and $a_{v_{\ell-1} v_\ell} > 0$ by the choice of the path. Thus, Claim 4.2 implies that there exists some $v_{\ell+1} \in N(v_\ell)$ with $c(\sigma(v_{\ell+1})) > c(\sigma(v_\ell))$ and $a_{v_\ell v_{\ell+1}} > 0$. This contradicts the maximality of the path.

Therefore, it must hold that $v_\ell = u$. But then $v_1 \ldots v_\ell$ is a cycle with $c(\sigma(v_1)) < \cdots < c(\sigma(v_\ell))$, which means that the associated cycle-equation is convex as required. □

4.2. **From equations to graphs.** Thus far we have only discussed cases where non-abundance of $\mathcal{E}(H, \sigma, c)$ determines non-abundance of (H, σ). But in general, we may ask whether certain properties of the equations $\mathcal{E}(H, \sigma) = \bigcup_c \{\mathcal{E}(H, \sigma, c)\}$ associated to an F-coloured graph (H, σ) determine whether this graph is F-abundant or not. First, we consider the special case where the graph H is a cycle C. In that case, each $\mathcal{E}(C, \sigma, c)$ is a single equation, and we show that the properties of $\mathcal{E}(H, \sigma)$ determine F-abundance.

Theorem 4.3. *Let (C, σ) be an F-coloured cycle. The following are equivalent*:

    (1) (C, σ) *is F-abundant*,

(2) *No equation from $\mathcal{E}(C, \sigma)$ is convex, and*

(3) *Every equation from $\mathcal{E}(C, \sigma)$ has genus at least two.*

Proof. (1) $\Rightarrow$ (2): If $\mathcal{E}(C, \sigma, c)$ is convex, we know from above that this would allow us to form a Ruzsa-Szemerédi graph with too few copies of (C, σ), contradicting (1).

(2) $\Rightarrow$ (1), (3): Let $C = v_1 \ldots v_\ell$. We claim that there exist $i \neq j$ with $\sigma(v_i) = \sigma(v_j)$. Indeed, otherwise we could choose c in such a way that $c(\sigma(v_i)) = i$ for all $i \in [\ell]$, but then the equation $\mathcal{E}(C, \sigma, c)$ is $\sum_{i=1}^{\ell-1} x_i - (\ell - 1)x_\ell = 0$ which is a convex equation, a contradiction. Note that v_i and v_j cannot be adjacent because $\sigma(v_i) = \sigma(v_j)$, and so C consists of two paths P and Q between v_i and v_j, each of length at least 2.

Given any injection $c \colon V(F) \to [n]$, we now conclude that $\mathcal{E}(C, \sigma, c)$ has genus at least two by partitioning the variables into the two sets of variables corresponding to the edges in P and Q respectively. This proves (3).

To obtain (1), we use Theorem 2.9. Without loss of generality we may assume that $v_i v_{i+1} \in P$ and $v_{i-1} v_i \in Q$. Observe that $(v_i v_{i+1} v_j v_{i-1}, \sigma)$ is an F-coloured C_4 that is F-abundant by Theorem 2.7 (or indeed by Cauchy-Schwarz if one is so inclined). Furthermore, $(P \setminus v_i, \sigma)$ and $(Q \setminus v_i, \sigma)$ are F-abundant. We now replace the edge $v_{i+1} v_j$ by $(P \setminus v_i, \sigma)$ and the edge $v_{i-1} v_j$ by $(Q \setminus v_i, \sigma)$. The resulting graph (C, σ) is then F-abundant by Theorem 2.9, as required.

(3) $\Rightarrow$ (2): This is true because an equation with genus at least two cannot be convex. $\quad\square$

In the case of triangle-abundance, we show that even if H is not a cycle, we can sometimes derive triangle-abundance (and in fact something much stronger) from properties of $\mathcal{E}(H, \sigma)$. In this setting, an F-colouring σ of H simply corresponds to a K_3-colouring of H with colours 1, 2, and 3. Let a *colour homomorphism* be a homomorphism between two coloured graphs (H, σ) and (H', σ') that preserves colours. Let $P_\infty^3 = (P_\infty, \phi)$ denotes the infinite path P_∞ with a cyclic 3-colouring ϕ of its vertices. Note that if (H, σ) is colour homomorphic to P_∞^3, then (H, σ) is triangle-abundant by Theorem 2.7. We prove the following.

Theorem 4.4. *Let (H, σ) be a K_3-coloured graph. Then $\mathcal{E}(H, \sigma)$ consists solely of symmetric equations if and only if (H, σ) is colour homomorphic to P_∞^3.*

Proof. A *walk* W in a graph is a sequence of vertices $v_0, \ldots, v_\ell$ such that $v_{i-1} v_i$ is an edge for every $i \in [\ell]$. We define the *wrap* of an oriented edge uv to be 1 if $\sigma(u) + 1 \equiv \sigma(v) \pmod 3$ and -1 otherwise. Let $\mathrm{wrap}(W) = \sum_{i=1}^{\ell} \mathrm{wrap}(v_{i-1} v_i)$. A cycle C is *wrapped*[3] if $\mathrm{wrap}(C) \neq 0$.

Claim 4.5. *If (W, σ) is a K_3-coloured walk, then there is at most one colour homomorphism $\gamma \colon (W, \sigma) \to P_\infty^3$ (up to shifts of P_∞^3). Moreover, if γ exists, then $\mathrm{wrap}(W) = \mathrm{wrap}(\gamma(W))$.*

Proof. Note that for every colour c, every vertex in P_∞^3 has at most one neighbour of colour c. Thus, once we pick $\gamma(v_0)$ (this choice is unique up to shifts of P_∞^3), this determines $\gamma(v_i)$ for all

---

[3]Any homotopically inclined reader is free to reintepret this.

$i \in [\ell]$ as the unique neighbour of $\gamma(v_{i-1})$ with colour $\sigma(v_i)$. Hence, there can be at most one colour homomorphism $\gamma \colon (W, \sigma) \to P_\infty^3$ (up to shifts of P_∞^3).

Moreover, if γ exists, we have $\sigma(v_i) = \phi(\gamma(v_i))$ for all i, so $\mathrm{wrap}(v_{i-1}v_i) = \mathrm{wrap}(\gamma(v_{i-1})\gamma(v_i))$ and thus $\mathrm{wrap}(W) = \mathrm{wrap}(\gamma(W))$. $\qquad\square$

Claim 4.6. *Let (H, σ) be a K_3-coloured graph. Then, (H, σ) is colour homomorphic to P_∞^3 if and only if no cycle in H is wrapped.*

Proof. Suppose that some cycle C in H is wrapped. If (H, σ) is colour homomorphic to P_∞^3, this yields a colour homomorphism $\gamma \colon (C, \sigma) \to P_\infty^3$. By Claim 4.5, we know that $\mathrm{wrap}(C) = \mathrm{wrap}(\gamma(C))$. However, $\gamma(C)$ is a walk from a vertex back to itself, and such walks have wrap 0 in P_∞^3, contradicting $\mathrm{wrap}(C) \neq 0$.

Conversely, suppose that no cycle in H is wrapped. Let $v \in V(H)$ be arbitrary. We may identify the vertices of P_∞^3 with $\mathbb{Z}$ such that $\phi(n) \equiv \phi(n-1) + 1 \pmod{3}$ for all $n \in \mathbb{Z}$, and we may assume that $\sigma(v) = \phi(0)$. Define γ by $\gamma(u) = \mathrm{wrap}(W)$ where W is an arbitrary walk in H from v to u. This is well defined because two distinct walks from v to u with different wrap would imply the existence of a wrapped cycle. It is straightforward to check that γ is a colour homomorphism. $\qquad\square$

Claim 4.7. *Let (C, σ) be a K_3-coloured cycle. Then, (C, σ) is wrapped if and only if some equation in $\mathcal{E}(C, \sigma)$ is not symmetric.*

Proof. Suppose that (C, σ) is wrapped. Let $c : [3] \to [n]$ be the identity and consider the equation $\mathcal{E}(C, \sigma, c)$. If $\mathrm{wrap}(C) > 0$, then the coefficients $c(2) - c(1)$, $c(3) - c(2)$, and $c(1) - c(3)$ appear more often than their inverses in $\mathcal{E}(C, \sigma, c)$, and so there are more positive than negative coefficients. The converse holds if $\mathrm{wrap}(C) < 0$. Therefore, the equation cannot be symmetric since symmetric equations have an equal number of positive and negative coefficients.

Conversely, suppose that (C, σ) is not wrapped. Let $c : [3] \to [n]$ be an arbitrary injection. Since $\mathrm{wrap}(C) = 0$, the coefficients $c(2) - c(1)$, $c(3) - c(2)$, and $c(1) - c(3)$ appear exactly as often as their inverses in $\mathcal{E}(C, \sigma, c)$. Hence, $\mathcal{E}(C, \sigma, c)$ is symmetric, and because c was arbitrary, every equation in $\mathcal{E}(C, \sigma)$ must be symmetric. $\qquad\square$

By Claim 4.7 we know that every equation of $\mathcal{E}(H, \sigma)$ is symmetric if and only if no cycle in H is wrapped. However, Claim 4.6 shows that this is equivalent to (H, σ) being colour homomorphic to P_∞^3. This concludes the proof. $\qquad\square$

Remark. Interestingly in the case of K_3-coloured graphs, the only graphs for which the Turán *degree* density[4] is zero are those colour homomorphic to P_∞^3.[5]

---

[4]For (G, σ) this is defined as the infimum $\delta > 0$ such that any K_3-coloured graph H of sufficiently large order with minimum degree (between each pair of parts) at least $\delta|H|$ contains a copy of (G, σ).

[5]To see it is positive for other graphs, let H be the tripartite graph with vertex set $([N], [N], [N])$, and connect two vertices x, y if $y - x \mod N$ is less than εN. The resulting graph has no short wrapped cycles.

This concludes what we can prove, but two central directions remain. The first is to find an example of a tripartite triangle-free graph that is not K_3-abundant. We have shown that one cannot do this through the Ruzsa-Szemerédi construction with convex equations, but the graph G_3 would be a good candidate for investigation. The second question is understanding in what cases the abundance of $\mathcal{E}(H, \sigma)$ determines the abundance of (H, σ) (the opposite implication is given by the Ruzsa-Szemerédi construction).

In both cases a new method of constructing non-abundant graphs would be hugely helpful. However, it is unclear how difficult this is. Another approach is to extend Ruzsa's work, either by understanding $R_E(N)$ for a larger class of equations or by understanding better systems of equations. This latter undertaking was begun by Shapira [Sha06]. Concretely, in an attempt to extend Theorem 4.4 beyond symmetric equations, one could ask whether or not there are genus-type properties of *systems* of equations that, if satisfied by all elements of $\mathcal{E}(G, \sigma)$, would imply that (G, σ) is F-abundant. See the following section for all open questions.

5. Open questions

Thus far we can only show that H is F-abundant by showing that some F-coloured (H, σ) is F-abundant. While this is clearly sufficient, we conjecture that is also necessary.

Conjecture 5.1. *A graph H is F-abundant if and only if there is an F-colouring σ of H such that (H, σ) is F-abundant.*

We say a *family $\mathcal{F}$ of graphs is triangle-abundant* if there exists $C > 0$ such that in any graph G that contains at least εn^2 disjoint triangles, some $F \in \mathcal{F}$ has density at least ε^C. The following is much stronger than Conjecture 5.1, but we are not confident enough to call it a conjecture. Perhaps examples of non-compactness in generalised Turán numbers[6] [AS16] could inspire counter-examples.

Question 5.2. *Is it true that a family $\mathcal{F}$ of (coloured) graphs is triangle-abundant if only if $\mathcal{F}$ contains a triangle-abundant (coloured) graph.*

Returning to the classification of abundance, a major open question is whether or not you can glue graphs on monochromatic vertex sets (if so then all *pinchable* graphs are K_3-abundant where pinchable means every subgraph has a monochromatic vertex cut set). The most basic open case is for cut sets of size two.

Question 5.3. *Suppose (F, σ) and (F', σ') are triangle-abundant. Let $u, v \in F$ and $u', v' \in F'$ with $\sigma(u) = \sigma(v) = \sigma(u') = \sigma(v')$. If we identify u with u' and v with v' is the resulting graph triangle-abundant?*

Remark. It is essential that $\sigma(u) = \sigma(v) = \sigma(u') = \sigma(v')$. For example, (C_4, σ) is not K_4-abundant where $C_4 = x_1 x_2 x_3 x_4$ and $\sigma(x_i) = i$.

---

[6]The maximum number of copies of G in an n-vertex F-free graph.

On the connection between equations and graphs, we re-iterate the (admittedly informal) challenge to extend Theorem 4.4 to the case where $\mathcal{E}$ satisfies some genus condition, rather than a symmetry condition. The work of Shapira [Sha06] on systems of equations may be useful. The following remarkable possibility also remains.

Question 5.4. *Is it true that an F-coloured graph (G, σ) is F-abundant if and only if all systems $S \in \mathcal{E}(G, \sigma)$ are abundant.*

If it is the case that the only unavoidable equations are equations of genus at least 2 (i.e. if there is a positive answer to Ruzsa's question), then it is plausible that the methods contained in this paper yield all abundant graphs. If on the other hand the only avoidable equations are convex equations, then it seems likely that there are many more (kinds of) abundant graphs, and it may even be that the only non-abundant graphs are increasing-cycle-unavoidable graphs. If the latter holds then all tripartite triangle-free graphs are triangle-abundant. We thus reiterate the question of Gishboliner, Shapira, and Wigderson [GSW23] which asks for a single tripartite triangle-free graph that can be (unconditionally) shown not to be triangle-abundant.

In the additive context, determining $R_E(N)$ for any non-convex equation E of genus one remains wide open. Any progress would be exciting. For example, the following equation has genus one and is not convex. Its solutions are sets of three numbers and two numbers whose averages are equal.

Question 5.5. *Let E be given by $2x_1 + 2x_2 + 2x_3 = 3x_4 + 3x_5$. Is E avoidable?*

Finally, the following additive question is both relevant and interesting in its own right, and may be solvable without classifying any abundant and avoidable equations. It has a super-saturation flavour to it. We recall that an equation E in ℓ variables is *abundant* if there exists constants $\gamma, C > 0$ such that any subset $\mathcal{A} \subset [N]$ of order εN contains at least $\gamma \varepsilon^C N^{\ell-1}$ solutions to E, and that it is *avoidable* if $R_E(N) = N^{1-o(1)}$.

Question 5.6. *For an equation, does unavoidability imply abundance?*

Acknowledgements

References

[ABE20] NOGA ALON and OMRI BEN-ELIEZER (2020). Efficient removal lemmas for matrices. *Order* **37**(1), 83–101. 2

[AF15] NOGA ALON and JACOB FOX (Feb. 2015). Easily testable graph properties. *Combinatorics, Probability and Computing* **24**(4), 646–657. 2

[AFN07] NOGA ALON, ELDAR FISCHER, and ILAN NEWMAN (Aug. 2007). Efficient testing of bipartite graphs for forbidden induced subgraphs. *SIAM Journal on Computing* **37**(3), 959–976. 2

[Alo02] NOGA ALON (Oct. 2002). Testing subgraphs in large graphs. *Random Structures & Algorithms* **21**(3-4), 359–370. 2, 6

[AS04] NOGA ALON and ASAF SHAPIRA (Nov. 2004). Testing subgraphs in directed graphs. *Journal of Computer and System Sciences* **69**(3), 353–382. 2

[AS06] NOGA ALON and ASAF SHAPIRA (Aug. 2006). A characterization of easily testable induced subgraphs. *Combinatorics, Probability and Computing* **15**(6), 791–805. 2

[AS16] NOGA ALON and CLARA SHIKHELMAN (Nov. 2016). Many T copies in H-free graphs. *Journal of Combinatorial Theory, Series B* **121**, 146–172. 25

[Beh46] FELIX ADALBERT BEHREND (1946). On sets of integers which contain no three terms in arithmetical progression. *Proceedings of the National Academy of Sciences* **32**(12), 331–332. 1, 16, 19

[Buk08] BORIS BUKH (2008). Non-trivial solutions to a linear equation in integers. *Acta Arithmetica* **131**(1), 51–55. 5, 16

[CFSZ21] DAVID CONLON, JACOB FOX, BENNY SUDAKOV, and YUFEI ZHAO (2021). The regularity method for graphs with few 4-cycles. *Journal of the London Mathematical Society* **104**(5), 2376–2401. 15

[Csa21] BÉLA CSABA (Sep. 2021). Regular decomposition of the edge set of graphs with applications. arXiv:2109.12394. 2

[EFR86] PAUL ERDŐS, PÉTER FRANKL, and VOJTĚCH RÖDL (Dec. 1986). The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs and Combinatorics* **2**(2), 113–121. 2

[Erd64] PAUL ERDŐS (Sep. 1964). On extremal problems of graphs and generalized graphs. *Israel Journal of Mathematics* **2**(3), 183–190. 7

[FL17] JACOB FOX and LÁSZLÓ MIKLÓS LOVÁSZ (Dec. 2017). A tight bound for Green's arithmetic triangle removal lemma in vector spaces. *Advances in Mathematics* **321**, 287–297. 2

[FLS18] JACOB FOX, LÁSZLÓ MIKLÓS LOVÁSZ, and LISA SAUERMANN (Nov. 2018). A polynomial bound for the arithmetic k-cycle removal lemma in vector spaces. *Journal of Combinatorial Theory, Series A* **160**, 186–201. 2

[GS21] LIOR GISHBOLINER and ASAF SHAPIRA (Nov. 2021). Removal lemmas with polynomial bounds. *International Mathematics Research Notices* **2021**(19), 14409–14444. 2

[GSW23] LIOR GISHBOLINER, ASAF SHAPIRA, and YUVAL WIGDERSON (Jan. 2023). An efficient asymmetric removal lemma and its limitations. arXiv:2301.07693. 1, 2, 3, 5, 6, 20, 21, 26

[GT22] LIOR GISHBOLINER and ISTVÁN TOMON (Oct. 2022). Polynomial removal lemmas for ordered graphs. *Combinatorial Theory* **2**(3), Paper No. 3. 2

[Kee11] PETER KEEVASH (2011). Hypergraph Turán problems. *Surveys in Combinatorics* 2011 (edited by ROBIN CHAPMAN), London Mathematical Society Lecture Note Series, 83–140. 7

[KF23] GABOR KUN and PANNA TIMEA FEKETE (2023). A polynomial removal lemma for posets. *Proceedings of EUROCOMB* 2023, 695–701. 2

[NZ04] JAROSLAV NEŠETŘIL and XUDING ZHU (Jan. 2004). On sparse graphs with given colorings and homomorphisms. *Journal of Combinatorial Theory, Series B* **90**(1), 161–172. 13

[RS78] IMRE Z. RUZSA and ENDRE SZEMERÉDI (1978). Triple systems with no six points carrying three triangles. *Combinatorics, Volume II*, no. 18 in Colloquia Mathematica Societatis János Bolyai, 939–945. 1, 4, 19

[Ruz93] IMRE Z. RUZSA (1993). Solving a linear equation in a set of integers. I. *Acta Arithmetica* **65**(3), 259–282. 1, 2, 4, 5, 15, 16

[Sha06] Asaf Shapira (2006). Behrend-type constructions for sets of linear equations. *Acta Arithmetica* **122**(1), 17–33. 25, 26

[Sid32] S. Sidon (1932). Ein Satz über trigonometrische Polynome und seine Anwendung in der Theorie der Fourier-Reihen. *Mathematische Annalen* **106**, 536–539. 3

[Zha23] Yufei Zhao (2023). Graph Theory and Additive Combinatorics (Cambridge University Press). 14