Paley-like graphs for the Ramsey number $r(C_4, K_t)$

Yuval Wigderson*

Abstract

A famous conjecture of Erdős asserts that the Ramsey number $r(C_4, K_t)$ is upper-bounded by $O(t^{2-\varepsilon})$ for some $\varepsilon > 0$, but the best known upper bound is $r(C_4, K_t) = O(t^2/\log^2 t)$, due to Szemerédi and Caro–Li–Rousseau–Zhang. We present a new explicit construction which we think disproves Erdős's conjecture. In fact, we believe our construction shows that $r(C_4, K_t) = \Theta(t^2/\log^2 t)$.

Our construction is closely related to Paley graphs, and proving that it works is at least as hard as proving a good upper bound on the independence number of Paley graphs, which is a major open problem. As such, we are not able to refute Erdős's conjecture. Nonetheless, we present heuristics and computational data which suggest our construction should indeed perform well.

1 Introduction

Given two graphs H_1, H_2, their *Ramsey number* $r(H_1, H_2)$ is defined as the least integer n such that every two-coloring of the edges of the complete graph K_n contains a red copy of H_1 or a blue copy of H_2. For the rest of this paper, we focus on the case where $H_2 = K_t$ is a clique and $H_1 = H$ is any fixed graph. In this case, $r(H, K_t)$ equals the least n such that every H-free graph on n vertices has independence number at least t. Because of this, we will usually think of the Ramsey problem in terms of the minimum independence number of an H-free graph on n vertices.

One of the greatest successes of graph Ramsey theory over the past century concerns the Ramsey number $r(C_3, K_t)$. Equivalently, this is the question of the minimum independence number of a triangle-free graph on n vertices. As a special case of their general bound on Ramsey numbers, Erdős and Szekeres [21] proved that every triangle-free graph on n vertices contains an independent set of order at least $\sqrt{n}$. The argument is simple: if there is a vertex of degree at least $\sqrt{n}$, then its neighborhood is the requisite independent set. If not, then we may greedily select low-degree vertices and delete their neighborhoods to construct an independent set of order at least $n/\sqrt{n} = \sqrt{n}$.

This greedy argument was improved by Ajtai, Komlós, and Szemerédi [1], who showed that every n-vertex triangle-free graph has an independent set of order at least $\Omega(\sqrt{n \log n})$.

*Department of Mathematics, Stanford University, Stanford, CA 94305, USA. Email: yuvalwig@stanford.edu. Research supported by NSF GRFP Grant DGE-1656518.

After much partial progress [15, 20, 29, 37] Kim [27] showed the existence of a triangle-free graph with independence number $O(\sqrt{n \log n})$, matching the lower bound up to a constant factor. Thanks to improvements by Shearer [35] to the lower bound and by Bohman–Keevash [7] and Fiz Pontiveros–Griffiths–Morris [23] to the upper bound, it is now known that the minimum independence number of an n-vertex triangle-free graph lies between $(\frac{1}{\sqrt{2}} - o(1))\sqrt{n \ln n}$ and $(\sqrt{2} + o(1))\sqrt{n \ln n}$. For more on the history of this problem, see the excellent survey of Spencer [38].

There are two natural extensions of the problem discussed above. The first is to view C_3 as a complete graph, and to study the minimum independence number of K_s-free graphs for fixed s. This question is very well-studied, see e.g. [6, 32]. However, we may also view C_3 as the shortest cycle, and study $r(C_s, K_t)$ for fixed $s \geq 4$.

The first case to consider is $r(C_4, K_t)$, i.e. the minimum independence number of C_4-free graphs. Kővári, Sós, and Turán [28] proved that an n-vertex C_4-free graph has $O(n^{\frac{3}{2}})$ edges, and thus contains a vertex of degree $O(\sqrt{n})$. By greedily selecting such vertices and deleting their neighbors, we can find an independent set of order $\Omega(\sqrt{n})$. This simple greedy argument was improved by Szemerédi, who showed that an n-vertex C_4-free graph must have independence number at least $\Omega(\sqrt{n} \log n)$; however, Szemerédi never published this result, and the first written proof is due to Caro, Li, Rousseau, and Zhang [9].

Unlike the case of triangles, no matching upper bound has yet been found. For many years, the best upper bound was due to Spencer [37], who proved the existence of an n-vertex C_4-free graph with independence number $O(n^{\frac{2}{3}} \log n)$. This was improved slightly by Bohman and Keevash [6] to $O(n^{\frac{2}{3}} \log^{\frac{2}{3}} n)$, and a different construction for the same bound was found by Mubayi and Verstraëte [32], but there remains a polynomial gap between the $n^{\frac{1}{2}+o(1)}$ lower bound and the $n^{\frac{2}{3}+o(1)}$ upper bound.

Moreover, Erdős conjectured that, in contrast to the case of triangles, the greedy argument gives a lower bound that is far from the truth.

Conjecture 1.1 (Erdős)**.** *There exists some $\varepsilon > 0$ such that every n-vertex C_4-free graph has independence number $\Omega(n^{\frac{1}{2}+\varepsilon})$.*

This was evidently one of Erdős's favorite conjectures, and he stated it in many problem papers, e.g. [16, 17, 18, 19].

The purpose of this paper is to present two closely related constructions which we believe disprove Conjecture 1.1. Both constructions yield an n-vertex C_4-free graph, which we expect to have independence number $n^{\frac{1}{2}+o(1)}$.

One of our constructions is as follows. Given an odd prime p, we define a C_4-free graph Λ_p whose vertex set is $\mathbb{F}_p^2$ and where two vertices (x, y) and (z, w) are adjacent if and only if $y \neq w$ and $(x + z)^2 = y + w$. Λ_p is closely related to the Paley sum graph P_p, whose vertex set is $\mathbb{F}_p$ and where vertices y and w are adjacent if and only if $y + w$ is a square in $\mathbb{F}_p$. In particular, it is easy to check that any independent set I in P_p can be lifted to an independent set in Λ_p of order $p|I|$. Conjecture 3.5, stated formally in Section 3, posits that this is the "best" way of building independent sets in Λ_p, and that the independence number of Λ_p is exactly p times the independence number of P_p. Another of our conjectures, Conjecture 3.9, is a strong form of a well-known number-theoretic conjecture, and says that the independence number of P_p is at most $(2+o(1))\log p$ for very many primes p; again, we defer the precise statement to Section 3.

Assuming these two conjectures, we are able to determine $r(C_4, K_t)$ up to a factor of $4\log^2 e + o(1) \approx 8.325 + o(1)$.

Theorem 1.2. *Assume Conjectures 3.5 and 3.9 hold. Then the minimum independence number of an n-vertex C_4-free graph is between $(\frac{1}{2\log e} - o(1))\sqrt{n}\log n$ and $(1+o(1))\sqrt{n}\log n$. In other words,*

$$\left(\frac{1}{4} - o(1)\right)\frac{t^2}{\log^2 t} \le r(C_4, K_t) \le (\log^2 e + o(1))\frac{t^2}{\log^2 t}.$$

Conjectures 3.5 and 3.9 are among the strongest conjectures we pose, and weaker assumptions lead to weaker versions of Theorem 1.2. However, all of these weaker assumptions still refute Erdős's Conjecture 1.1, by yielding a family of n-vertex C_4-free graphs with independence number $n^{\frac{1}{2}+o(1)}$.

The fact that we cannot prove any of these conjectures is, in some sense, to be expected. As indicated above, our constructions are closely related to the Paley sum graphs, which are widely believed have very small clique and independence numbers. Proving that our construction works implies that P_p has independence number $p^{o(1)}$, which is a major open problem. Conversely, we present heuristics suggesting that if one could prove that P_p has independence number $p^{o(1)}$, it might be possible to control the independence number of Λ_p, and thus refute Erdős's Conjecture 1.1.

The rest of this paper is structured as follows. In Section 2, we present our first construction of a C_4-free graph which we expect to have small independence number. In Section 3, we present the second construction, describe its connection to the Paley sum graph, formulate some more refined conjectures about their independence numbers, and prove Theorem 1.2. In Section 4, we present a simple random model which we expect to capture important properties of our constructions, and show that this random model has independence number $n^{\frac{1}{2}+o(1)}$ with high probability. Finally, in Section 5, we present computational data in support of our conjectures.

As usual, $\alpha(G)$ denotes the independence number of a graph G. We use ln to denote the natural logarithm and log to denote the base-two logarithm. For the sake of clarity of presentation, we systematically omit floor and ceiling signs whenever they are not crucial.

2 A new construction of C_4-free graphs

To describe our construction, we recall the notion of a Sidon set.

Definition 2.1. Given an abelian group Γ, a set $S \subseteq \Gamma$ is called a *Sidon set* if the only solutions to the equation $a + b = c + d$ with $a, b, c, d \in S$ are the solutions where $a = c$ or $a = d$.

The connection between Sidon sets and C_4-free graphs is given by the following lemma. Although we have not seen this precise lemma in the literature, it is essentially identical to other well-known facts about graphs arising from Sidon sets, e.g. [14, Proposition 2.1]. Recall that given a finite abelian group Γ and a set $S \subseteq \Gamma$, the *Cayley sum graph* $\mathrm{Cay}^+(\Gamma, S)$ is the graph whose vertex set is Γ, and where two distinct[1] vertices u, v are joined by an edge if and only if $u + v \in S$.

Lemma 2.2. *Let Γ be a finite abelian group. If $S \subseteq \Gamma$ is a Sidon set, then $\mathrm{Cay}^+(\Gamma, S)$ is a C_4-free graph.*

Proof. Suppose for contradiction that distinct vertices $v_1, \ldots, v_4$ form a C_4 in $\mathrm{Cay}^+(\Gamma, S)$. Let

$$a = v_1 + v_2 \qquad b = v_3 + v_4 \qquad c = v_2 + v_3 \qquad d = v_1 + v_4,$$

which are all elements of S by assumption. Then $a + b = v_1 + v_2 + v_3 + v_4 = c + d$, so by the Sidon condition, we must have $a = c$ or $a = d$. If $a = c$ then $v_1 = v_3$, which contradicts our assumption of distinct vertices. Similarly, if $a = d$ then $v_2 = v_4$, which yields the same contradiction. □

Remark. Lemma 2.2 is false for Cayley graphs in place of Cayley sum graphs. Indeed, any Cayley graph with at least two generators will contain many copies of C_4, corresponding to degenerate solutions of the Sidon equation arising from commutativity.

The Sidon set that we will use is the parabola construction of Erdős and Turán[2] [22].

Lemma 2.3 (Erdős and Turán)**.** *Let p be an odd prime. Then the parabola $P = \{(t, t^2) : t \in \mathbb{F}_p\}$ is a Sidon subset of the abelian group $\mathbb{F}_p^2$.*

Proof. Suppose that $(a, a^2) + (b, b^2) = (c, c^2) + (d, d^2)$. By squaring the equality of the first coordinates and subtracting the equality of the second coordinates, we find that $2ab = 2cd$, and thus that $ab = cd$ since p is odd. So the polynomials $(x - a)(x - b)$ and $(x - c)(x - d)$ are equal, and thus $a = c$ or $a = d$. □

With these ingredients, we are now ready to present our main construction.

Definition 2.4. Let p be an odd prime. The *parabola graph* Π_p is the Cayley sum graph $\mathrm{Cay}^+(\mathbb{F}_p^2, P)$, where $P = \{(t, t^2) : t \in \mathbb{F}_p\}$. Concretely, the vertex set of Π_p is $\mathbb{F}_p^2$, and vertices (x, y) and (z, w) are joined by an edge if and only if $(x + z)^2 = y + w$.

As a consequence of Lemmas 2.2 and 2.3, we see that for every odd prime p, the graph Π_p is a C_4-free graph on p^2 vertices. Our main conjecture is that the independence number of Π_p is small.

[1] We insist on distinct vertices so that the resulting graph has no loops.

[2] The construction is somewhat implicit in [22], since Erdős and Turán were interested in constructing a Sidon set in $\mathbb{Z}$. See e.g. [14, Construction 1] for an explicit exposition of this construction in $\mathbb{F}_p^2$.

Conjecture 2.5. *The independence number of Π_p is $p^{1+o(1)}$, where the $o(1)$ term tends to 0 as $p \to \infty$. In particular, we have that $r(C_4, K_t) = t^{2-o(1)}$ as $t \to \infty$.*

Remark. Since Π_p is a Cayley sum graph, its eigenvalues are determined by character sums on the parabola $P \subseteq \mathbb{F}_p^2$ (see e.g. [3, Section 3.1]). A simple computation involving Gauss sums (done in e.g. [8, Lemma 13.6]) shows that every non-trivial eigenvalue of Π_p has absolute value $O(\sqrt{p})$. Thus, by the Hoffman bound (see [25]), or by the expander mixing lemma, one can conclude that $\alpha(\Pi_p) = O(p^{\frac{3}{2}})$. Letting $n = p^2$ be the number of vertices of Π_p, this says that $\alpha(\Pi_p) = O(n^{\frac{3}{4}})$.

A nearly identical argument shows that several other explicit C_4-free graphs, such as the projective norm graph and the projective plane polarity graph, have independence number $O(n^{\frac{3}{4}})$. In these cases, it was shown by Mubayi and Williford [33] that this spectral upper bound is tight up to the implicit constant. Thus, Conjecture 2.5 asserts that in contrast to known constructions of C_4-free graphs, the independence number of Π_p is significantly smaller than the bound given by its spectrum.

In the subsequent sections, we will describe various heuristics and computational data in support of Conjecture 2.5, and will formulate certain stronger conjectures which we believe explain why Π_p should have such a small independence number.

3 The graph Λ_p, the Paley sum graph, and more refined conjectures

The parabola graph Π_p is a natural construction of a C_4-free graph, since it is the C_4-free graph arising from applying Lemma 2.2 to one of the most basic Sidon sets we know of, the parabola in $\mathbb{F}_p^2$. However, the following large subgraph of Π_p, already mentioned in the Introduction, seems more well-behaved for certain purposes.

Definition 3.1. Given an odd prime p, the *limited parabola graph* Λ_p is the subgraph of Π_p obtained by deleting all edges between vertices with equal second coordinate. Concretely, the vertex set of Λ_p is $\mathbb{F}_p^2$, and vertices (x, y) and (z, w) are joined by an edge if and only if $y \neq w$ and $(x + z)^2 = y + w$.

Since Λ_p is a subgraph of Π_p, it is certainly C_4-free. Moreover, the following simple lemma shows that, as far as independence numbers are concerned, there is no significant difference between Π_p and Λ_p.

Lemma 3.2. *For every odd prime p, we have that $\alpha(\Pi_p) \leq \alpha(\Lambda_p) \leq 3\alpha(\Pi_p)$.*

Proof. Since Λ_p is a subgraph of Π_p, we immediately have that $\alpha(\Pi_p) \leq \alpha(\Lambda_p)$. Moreover, we see that the edges deleted to obtain Λ_p from Π_p form a graph of maximum degree 2, which implies that $\alpha(\Pi_p) \geq \alpha(\Lambda_p)/3$; indeed, any independent set in Λ_p can be partitioned into three independent sets in Π_p. □

Thanks to Lemma 3.2, in order to prove Conjecture 2.5, it suffices to prove an upper bound on the independence number of Λ_p. The advantage of considering Λ_p is that it has a very rigid structure, which allows one to make more precise conjectures about its independence number.

First, note that the vertex set of Λ_p is a disjoint union of p independent sets, each of size p. Indeed, since vertices with the same second coordinate are always non-adjacent in Λ_p, this partition into independent sets is obtained by partitioning the vertex set into "horizontal lines", i.e. sets of vertices with the same second coordinate. Moreover, in this partition, we see that two such independent sets have edges between them only if the sum of their second coordinates is a square in $\mathbb{F}_p$. In other words, this partition yields a homomorphism from Λ_p to the *Paley sum graph* of order p, whose definition is as follows.

Definition 3.3. Given a prime p, the *Paley sum graph* P_p is the graph $\mathrm{Cay}^+(\mathbb{F}_p, Q)$, where $Q = \{t^2 : t \in \mathbb{F}_p\}$ is the set of perfect squares in $\mathbb{F}_p$.

This perspective gives us another description of Λ_p. We begin with the Paley sum graph P_p, and replace every vertex by an independent set of order p. For every edge wy of P_p, we put some edges between the corresponding independent sets. Namely, if $y \neq -w$, then we put a 2-regular bipartite graph between the independent sets corresponding to w and y; we connect a vertex (x, y) to (z_1, w) and (z_2, w), where $x + z_1$ and $x + z_2$ are the two square roots of $w + y$. Similarly, if $y = -w$, then $w + y$ has a unique square root (namely 0), in which case we connect (x, y) to $(-x, w)$; thus, we place a matching between the independent sets corresponding to w and y.

Thus, we find that Λ_p is "essentially" a lift of the Paley sum graph P_p. Recall that a graph G is an *ℓ-lift* of a graph H if the vertex set of G is obtained from the vertex set of H by replacing each vertex by an independent set of size ℓ, and replacing every edge of H by a matching between the corresponding independent sets in G. For more on lifts, see e.g. [4]. In our case, Λ_p is obtained from P_p by replacing each vertex by an independent set of order p, but rather than lifting every edge of P_p to a matching, we lift most of them to 2-regular graphs, while exactly $\frac{p-1}{2}$ of them are lifted to matchings.

This perspective immediately yields a relationship between the independence numbers of the Paley sum graph and the limited parabola graph.

Lemma 3.4. *For every odd prime p, we have that $\alpha(\Lambda_p) \geq p\alpha(P_p)$.*

Proof. Suppose we are given an independent set I in P_p of size $\alpha(P_p)$. Taking the union of the "horizontal lines" in Λ_p corresponding to every vertex $v \in I$, we get an independent set in Λ_p of size $p\alpha(P_p)$, as claimed. □

Moreover, we conjecture that this simple lower bound is *tight*.

Conjecture 3.5. *For every odd prime p, we have that $\alpha(\Lambda_p) = p\alpha(P_p)$.*

This conjecture is rather strong, and it may well be false: any p-lift G of P_p will also have $\alpha(G) \geq p\alpha(P_p)$, but in general, this inequality is not tight. That said, we computationally verified Conjecture 3.5 for $p \leq 17$ (see Section 5). Roughly speaking, Conjecture 3.5 says that the *only* way to generate a maximum independent set in Λ_p is to begin with an independent set in P_p, to lift it to Λ_p by taking the union of the corresponding "horizontal lines", and then to potentially perturb it by shifting some vertices from occupied horizontal lines to unoccupied ones.

One reason for formulating Conjecture 3.5 is that the independence number of the Paley sum graph[3] has received a great deal of attention. Indeed, it is widely expected that the Paley sum graph is a good Ramsey graph, in the sense that its clique number and independence number should both be very small. In particular, the following is a special case of a well-known number-theoretic conjecture, often called the Paley graph conjecture (see e.g. [10, 11, 36]).

Conjecture 3.6. *$\alpha(P_p) = p^{o(1)}$ as $p \to \infty$.*

Note that Conjectures 3.5 and 3.6 together imply Conjecture 2.5. Conversely, Lemmas 3.2 and 3.4 imply that

$$\alpha(P_p) \leq \frac{1}{p}\alpha(\Lambda_p) \leq \frac{3}{p}\alpha(\Pi_p).$$

Thus, Conjecture 2.5 implies Conjecture 3.6.

This suggests that proving Conjecture 2.5 is very difficult, since Conjecture 3.6 is a famously intractable open problem. Indeed, the best known bound on the independence number of P_p is $\alpha(P_p) \leq (1 + o(1))\sqrt{p/2}$, proved recently by Hanson–Petridis [26] and by Di Benedetto–Solymosi–White [13]. This only improves by a constant factor the simple, classical bound of $\alpha(P_p) \leq \sqrt{p}$, and it remains a major open problem to even prove $\alpha(P_p) = o(\sqrt{p})$.

Despite these difficulties, one can formulate natural conjectures that go beyond Conjecture 3.6. In many ways, P_p closely resembles the Erdős–Rényi random graph $G(p, \frac{1}{2})$; see e.g. [8, Section 13.2]. Because of this, one might expect that $\alpha(P_p) = O(\log p)$. This turns out to be false; Graham and Ringrose [24] showed that for infinitely many primes p, one has $\alpha(P_p) = \Omega(\log p \log \log \log p)$, and Montgomery [30, Theorem 13.5] proved the stronger bound $\alpha(P_p) = \Omega(\log p \log \log p)$ for infinitely many p, assuming the generalized Riemann hypothesis. Nonetheless, heuristics suggest that the primes p for which $\alpha(P_p) = \omega(\log p)$ should be very rare. In particular, the following conjecture seems reasonable; it asserts that $\alpha(P_p) = O(\log p)$ for "many" primes p.

Conjecture 3.7. *There exists a constant $C > 0$ such that for all sufficiently large m, there exists a prime p between m and Cm with $\alpha(P_p) \leq C \log p$.*

Suppose that Conjectures 3.5 and 3.7 are true, and let n be sufficiently large. By Conjecture 3.7, we can find some prime $\sqrt{n} \leq p \leq C\sqrt{n}$ such that $\alpha(P_p) \leq C \log p$. By Conjecture 3.5, this implies that $\alpha(\Lambda_p) \leq Cp \log p \leq 2C^2\sqrt{n} \log n$. Finally, by deleting some arbitrary vertices from Λ_p, we can find an n-vertex C_4-free graph whose independence

[3] As a matter of fact, most of the attention has been paid to the Paley graph, rather than the Paley *sum* graph. But the two are closely related, and every result and heuristic mentioned here applies equally well to the Paley graph and the Paley sum graph.

number is at most $2C^2\sqrt{n}\log n$. Combining this with the result of Szemerédi and Caro–Li–Rousseau–Zhang [9] mentioned in the Introduction, we find that we have proved the following theorem.

Theorem 3.8. *Assume Conjectures 3.5 and 3.7 hold. Then the minimum independence number of an n-vertex C_4-free graph is $\Theta(\sqrt{n}\log n)$. In other words,*

$$r(C_4, K_t) = \Theta\left(\frac{t^2}{\log^2 t}\right).$$

In fact, something much stronger than Conjecture 3.7 is likely true. Mrazović [31] recently introduced a random model for the Paley sum graph, and proved that its independence number is $\Omega(\log p \log\log p)$ infinitely often, but for almost all primes, its independence number is $(2+o(1))\log p$. Such a model supports the following conjecture, which strengthens Conjecture 3.7.

Conjecture 3.9. *For every integer m, there exists a prime p between m and $(1+o(1))m$ with $\alpha(P_p) \le (2+o(1))\log p$.*

Indeed, assuming Mrazović's model accurately captures the behavior of the Paley sum graphs, it is likely the case that $\alpha(P_p) = (2+o(1))\log p$ for a $1-o(1)$ fraction of the primes p. By standard results on primes in short intervals (e.g. [30, Theorem 14.1]), this suggests that one should always be able to find such a "good" prime between m and $(1+o(1))m$.

Using Conjecture 3.9, we are able to prove the following refinement of Theorem 3.8, which was stated in the Introduction.

Theorem 1.2. *Assume Conjectures 3.5 and 3.9 hold. Then the minimum independence number of an n-vertex C_4-free graph is between $(\frac{1}{2\log e} - o(1))\sqrt{n}\log n$ and $(1+o(1))\sqrt{n}\log n$. In other words,*

$$\left(\frac{1}{4} - o(1)\right)\frac{t^2}{\log^2 t} \le r(C_4, K_t) \le (\log^2 e + o(1))\frac{t^2}{\log^2 t}.$$

Proof. The upper bound $r(C_4, K_t) \le (1+o(1))t^2/\ln^2 t = (\log^2 e + o(1))t^2/\log^2 t$ is given in [9, Corollary 3]. This is equivalent to the assertion that every n-vertex C_4-free graph has independence number at least $(\frac{1}{2\log e} - o(1))\sqrt{n}\log n$.

For the other direction, by Conjecture 3.9, we can find a prime $\sqrt{n} \le p \le (1+o(1))\sqrt{n}$ with $\alpha(P_p) \le (2+o(1))\log p$. By Conjecture 3.5, this implies that $\alpha(\Lambda_p) \le (2+o(1))p\log p = (1+o(1))\sqrt{n}\log n$. By deleting some vertices from Λ_p to get an n-vertex C_4-free graph, we obtain the desired result. □

4 A probabilistic heuristic

In this section, we give a probabilistic heuristic in support of Conjecture 2.5. Our proof in this section is inspired by the work of Amit, Linial, and Matoušek [4], who analyzed the independence number of a random lift of a graph. However, we cannot apply their results directly, since they deal with large lifts of a fixed graph, whereas we wish to consider lifts where the number of vertices in each independent set equals the number of vertices in the graph being lifted.

Our random model is as follows. Given the Paley sum graph P_p, we form a random graph $\widetilde{P_p}$ by replacing each vertex of P_p by an independent set of order p, and by replacing each edge of P_p by a uniformly random matching between the two corresponding independent sets. Recall that in Λ_p, we actually replace most edges of P_p by a 2-regular graph between the two corresponding independent sets, so one can think of $\widetilde{P_p}$ as a random model for a subgraph of Λ_p, containing roughly half its edges. But since we are interested in *upper*-bounding the independence number of Λ_p, we may freely pass to a subgraph.

Before stating the results, we should consider whether $\widetilde{P_p}$ really is a good random model for Λ_p. Recall that the edges of Λ_p are obtained by taking square roots in $\mathbb{F}_p$: given two independent sets, corresponding to $w, y \in \mathbb{F}_p$, a vertex (x, y) in the first independent set is joined to (z_1, w) and (z_2, w), where $x + z_1$ and $x + z_2$ are the two square roots of $y + w$ in $\mathbb{F}_p$. Much of the intuition driving the study of Paley sum graphs and related objects is that the quadratic residues in $\mathbb{F}_p$ behave like a random set. In particular, squaring should behave like a random two-to-one function on $\mathbb{F}_p$, and thus taking square roots should also be a random-like operation. All of this suggests that a random 2-regular graph between independent sets is a good model for the edges of Λ_p, and therefore that $\widetilde{P_p}$ is a good random model for a subgraph of Λ_p. However, we note that there is a key difference between the explicit graph Λ_p and its random model $\widetilde{P_p}$, namely that the former is C_4-free, while the latter contains many copies of C_4 with high probability. This is akin to many existing algebraic constructions of graphs, e.g. the pseudorandom triangle-free graphs of Alon [2]. Such graphs are in many ways very similar to random graphs of the same density, but differ greatly in some local way; for instance, Alon's graphs are triangle-free, though a random graph of the same density contains very many triangles with high probability.

We can now state our main result on the random model for Λ_p.

Theorem 4.1. *Assume that Conjecture 3.6 holds, and let $\widetilde{P_p}$ be the random lift of the Paley sum graph, defined above. For any $\varepsilon > 0$, and any sufficiently large prime p, we have that*

$$\Pr[\alpha(\widetilde{P_p}) \leq p^{1+\varepsilon}] \geq 1 - \varepsilon.$$

In particular, sending $\varepsilon \to 0$, we conclude that with high probability as $p \to \infty$, we have that $\alpha(\widetilde{P_p}) = p^{1+o(1)}$.

In order to prove Theorem 4.1, we need the following simple lemma.

Lemma 4.2. *Let G be a graph with independence number α. Then every $Y \subseteq V(G)$ with $|Y| \geq 6\alpha^2$ spans at least $|Y|^2/(6\alpha)$ edges.*

Proof. Suppose that this is false for some set Y. Since Y spans fewer than $|Y|^2/(6\alpha)$ edges, the average degree of the induced subgraph $G[Y]$ is less than $d = |Y|/(3\alpha)$. By greedily selecting vertices of lowest degree and deleting their neighbors (or by applying Turán's theorem to the complement graph), we may find an independent subset of Y of size

$$\frac{|Y|}{d+1} \geq \frac{|Y|}{2d} = \frac{3}{2}\alpha,$$

where the first inequality uses the fact that $d \geq 1$ since $|Y| \geq 6\alpha^2$. This contradicts the assumption that G has independence number α. □

With this lemma in hand, we can prove Theorem 4.1.

Proof of Theorem 4.1. Fix $\varepsilon > 0$, and let $\delta = \varepsilon/4$. We assume for simplicity that $1/\delta$ is an integer, which we may do since proving the result for any fixed $\varepsilon > 0$ implies it for all $\varepsilon' > \varepsilon$.

For every sufficiently large prime p, we have that $\frac{1}{\delta}p^{1+3\delta} \leq p^{1+\varepsilon}$. Recall that we assume that Conjecture 3.6 holds, so that $\alpha(P_p) = p^{o(1)}$. Therefore, we may also pick p sufficiently large so that $\alpha(P_p) \leq \frac{1}{6}p^{\delta/2}$. By Lemma 4.2, this implies that every $Y \subseteq V(P_p)$ with $|Y| \geq p^\delta$ spans at least $p^{-\delta/2}|Y|^2$ edges.

For every $y \in \mathbb{F}_p$, let $V_y \subseteq V(\widetilde{P_p})$ be the independent set corresponding to y in the random lift. Fix a set $T \subseteq V(\widetilde{P_p})$ with $|T| = \frac{1}{\delta}p^{1+3\delta}$. For every integer $0 \leq i \leq \frac{1}{\delta} - 1$, let

$$Y_i = \left\{y \in \mathbb{F}_p : p^{i\delta} \leq |T \cap V_y| \leq p^{(i+1)\delta}\right\}.$$

Note that we have

$$\frac{1}{\delta}p^{1+3\delta} = |T| = \sum_{y \in \mathbb{F}_p} |T \cap V_y| \leq \sum_{i=0}^{\frac{1}{\delta}-1} p^{(i+1)\delta}|Y_i|,$$

which implies that there exists some i such that $|Y_i| \geq p^{1+2\delta-i\delta}$. Fix such an i. For every $y \in Y_i$, let $T_y = T \cap V_y$, so that $|T_y| \geq p^{i\delta}$.

Since Y_i is a subset of $\mathbb{F}_p$ with $|Y_i| \geq p^{1+2\delta-i\delta} \geq p^\delta$, we have by the above that Y_i spans at least $p^{-\delta/2}|Y_i|^2$ edges in P_p. Given an edge wy of $P_p[Y_i]$, there is a random matching between V_w and V_y in $\widetilde{P_p}$. Therefore, the probability that there is no edge between T_w and T_y in $\widetilde{P_p}$ is the probability that a random permutation of $[p]$ maps T_w to a set disjoint from T_y, which is

$$\frac{\binom{p-|T_y|}{|T_w|}}{\binom{p}{|T_w|}} \leq \left(\frac{p-|T_y|}{p}\right)^{|T_w|} \leq \exp\left(-\frac{|T_w||T_y|}{p}\right) \leq \exp\left(-p^{2i\delta-1}\right),$$

where the last step uses that $|T_w|, |T_y| \geq p^{i\delta}$, since $w, y \in Y_i$.

The probability that T is an independent set in $\widetilde{P_p}$ is at most the probability that there is no edge within $\bigcup_{y \in Y_i} T_y$, which equals the probability that there is no edge between T_w

and T_y for all edges $wy \in P_p[Y_i]$. Since there are at least $p^{-\delta/2}|Y_i|^2 \geq p^{2+7\delta/2-2i\delta}$ such edges, and since these events are independent over all distinct edges, we have that

$$\begin{aligned}\Pr[T \text{ is independent in } \widetilde{P_p}] &\leq \prod_{wy \in E(P_p[Y_i])} \Pr[\text{there is no edge between } T_w \text{ and } T_y] \\ &\leq \exp\left(-p^{2i\delta-1}\right)^{p^{-\delta/2}|Y_i|^2} \\ &\leq \exp\left(-p^{2i\delta-1} \cdot p^{2+7\delta/2-2i\delta}\right) \\ &= \exp\left(-p^{1+7\delta/2}\right).\end{aligned}$$

On the other hand, the number of choices for T is

$$\binom{p^2}{\frac{1}{\delta}p^{1+3\delta}} \leq \left(p^2\right)^{\frac{1}{\delta}p^{1+3\delta}} = \exp\left(\frac{2}{\delta}\ln p \cdot p^{1+3\delta}\right)$$

which implies by the union bound that

$$\Pr\left(\alpha(\widetilde{P_p}) \geq p^{1+\varepsilon}\right) \leq \Pr\left(\alpha(\widetilde{P_p}) \geq \frac{1}{\delta}p^{1+3\delta}\right) \leq \exp\left(\frac{2}{\delta}\ln p \cdot p^{1+3\delta} - p^{1+7\delta/2}\right) = o(1),$$

since for fixed δ and p sufficiently large we have that $p^{\delta/2} > \frac{2}{\delta}\ln p$. In particular, by selecting p sufficiently large, we can ensure that this probability is at most ε, which proves the desired result. □

In Section 3, we pointed out that Conjecture 2.5 implies Conjecture 3.6, which is a major open problem. However, the results of this section suggest that some sort of converse may be true. Indeed, proving that $\alpha(P_p) = p^{o(1)}$ requires proving a strong quantitative form of the assertion that the quadratic residues in $\mathbb{F}_p$ are “random-like”. If one could develop the techniques to prove such a statement, then it is not unreasonable to expect that one could prove that the square root function on $\mathbb{F}_p$ is also “random-like”. If one could do that, then perhaps one could modify the argument above, which worked for a random lift, to prove that the deterministic but pseudorandom lift Λ_p also has independence number $p^{1+o(1)}$.

We remark that the random model proposed in this section gives a good heuristic justification for Conjecture 2.5, but not for the more refined Conjecture 3.5. For Conjecture 3.5, all our justification is from the numerical data, presented in Section 5. Nonetheless, it may be possible to prove that $\alpha(\widetilde{P_p}) = p\alpha(P_p)$ with high probability (possibly assuming some number-theoretic conjectures), which would provide such a heuristic justification for Conjecture 3.5. Indeed, one might be able to prove that with high probability, all maximum independent sets in $\widetilde{P_p}$ are “close” to a union of the basic independent sets V_y. If one could prove that, then potentially one could then modify any maximum independent set into one that genuinely is a union of V_y, at which point it is immediate that $\alpha(\widetilde{P_p}) = p\alpha(P_p)$. Results of this type have been proved using the container method (see e.g. the surveys [5, 34]), but we were not able to adapt such techniques to this problem, and thus leave it as a conjecture.

Conjecture 4.3. *$\alpha(\widetilde{P_p}) = p\alpha(P_p)$ with high probability as $p \to \infty$.*

5 Computational data

Table 1 below records the independence numbers of Π_p, Λ_p, and P_p for as many primes p as our computational power allowed. We used Sage [39] to define the graphs, and the `MaxCliquePara` algorithm developed in [12] to compute their independence numbers. The code used to generate this data is attached to the arXiv submission of this paper.

Unfortunately, we were not able to compute $\alpha(\Pi_p)$ and $\alpha(\Lambda_p)$ for very many primes p. Indeed, computing the independence number of a graph is NP-hard, and the graphs Π_p and Λ_p have p^2 vertices, which means that the complexity of the problem grows extremely quickly as p increases. For all primes $p \leq 13$, we were able to compute these independence numbers in under half a second. Yet already for $p = 17$, it took over seven hours to compute $\alpha(\Lambda_{17})$, and we could not compute $\alpha(\Pi_{17})$ even after running the program for seven days. At the moment, it appears that computing $\alpha(\Pi_{17})$ and $\alpha(\Lambda_{19})$ is out of our reach; that said, we would be very interested if someone else could push these computations further.

p	$\alpha(\Pi_p)$	$\alpha(\Lambda_p)$	$\alpha(P_p)$
3	4	6	2
5	9	10	2
7	14	21	3
11	28	33	3
13	37	52	4
17	?	68	4

Table 1: Numerical data for $\alpha(\Pi_p)$, $\alpha(\Lambda_p)$, and $\alpha(P_p)$ for odd primes $p \leq 17$.

There is not really enough data in Table 1 to support Conjecture 2.5, since it is basically impossible to determine the growth rate of a function from five or six data points. Nonetheless, we think that this data does provide reasonable evidence for Conjecture 3.5. Given how unstructured the values of $\alpha(\Pi_p)$ seem to be, it is rather striking that in all cases we could compute, we have that $\alpha(\Lambda_p)$ equals $p\alpha(P_p)$ *exactly.* Of course, this is a trend that could easily break down beyond $p = 17$, but it is notable and is the reason why we formulated Conjecture 3.5.

Acknowledgments I am indebted to Noga Alon, Jacob Fox, Xiaoyu He, and Avi Wigderson for helpful discussions on this topic, and for several suggestions that greatly improved this paper.

Some of the computing for this project was performed on the Sherlock cluster. I would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results. I am especially grateful to Randy White for his patience and his help on the computational aspects of this paper.

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi, A note on Ramsey numbers, *J. Combin. Theory Ser. A* **29** (1980), 354–360.

[2] N. Alon, Explicit Ramsey graphs and orthonormal labelings, *Electron. J. Combin.* **1** (1994), Research Paper 12, 8pp.

[3] N. Alon, Large sets in finite fields are sumsets, *J. Number Theory* **126** (2007), 110–118.

[4] A. Amit, N. Linial, and J. Matoušek, Random lifts of graphs: independence and chromatic number, *Random Structures Algorithms* **20** (2002), 1–22.

[5] J. Balogh, R. Morris, and W. Samotij, The method of hypergraph containers, in *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. IV. Invited lectures*, World Sci. Publ., Hackensack, NJ, 2018, 3059–3092.

[6] T. Bohman and P. Keevash, The early evolution of the H-free process, *Invent. Math.* **181** (2010), 291–336.

[7] T. Bohman and P. Keevash, Dynamic concentration of the triangle-free process, in *The Seventh European Conference on Combinatorics, Graph Theory and Applications*, *CRM Series*, vol. 16, Ed. Norm., Pisa, 2013, 489–495.

[8] B. Bollobás, *Random graphs*, *Cambridge Studies in Advanced Mathematics*, vol. 73, second ed., Cambridge University Press, Cambridge, 2001.

[9] Y. Caro, Y. Li, C. C. Rousseau, and Y. Zhang, Asymptotic bounds for some bipartite graph: complete graph Ramsey numbers, *Discrete Math.* **220** (2000), 51–56.

[10] M.-C. Chang, On a question of Davenport and Lewis and new character sum bounds in finite fields, *Duke Math. J.* **145** (2008), 409–442.

[11] B. Chor and O. Goldreich, Unbiased bits from sources of weak randomness and probabilistic communication complexity, *SIAM J. Comput.* **17** (1988), 230–261.

[12] M. Depolli, J. Konc, K. Rozman, R. Trobec, and D. Janežič, Exact parallel maximum clique algorithm for general and protein graphs, *J. Chem. Inf. Model.* **53** (2013), 2217–2228. Program available at `https://e6.ijs.si/~matjaz/maxclique/`.

[13] D. Di Benedetto, J. Solymosi, and E. P. White, On the directions determined by a Cartesian product in an affine Galois plane, 2020. Preprint available at arXiv:2001.06994.

[14] S. Eberhard and F. Manners, The apparent structure of dense Sidon sets, 2021. Preprint available at arXiv:2107.05744.

[15] P. Erdős, Graph theory and probability. II, *Canadian J. Math.* **13** (1961), 346–352.

[16] P. Erdős, On the combinatorial problems which I would most like to see solved, *Combinatorica* **1** (1981), 25–42.

[17] P. Erdős, Extremal problems in number theory, combinatorics and geometry, in *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983)*, PWN, Warsaw, 1984, 51–70.

[18] P. Erdős, On some problems in graph theory, combinatorial analysis and combinatorial number theory, in *Graph theory and combinatorics (Cambridge, 1983)*, Academic Press, London, 1984, 1–17.

[19] P. Erdős, Problems and results in combinatorial analysis and graph theory, in *Proceedings of the First Japan Conference on Graph Theory and Applications (Hakone, 1986)*, vol. 72, 1988, 81–92.

[20] P. Erdős, S. Suen, and P. Winkler, On the size of a random maximal graph, in *Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, "Random Graphs '93" (Poznań, 1993)*, vol. 6, 1995, 309–318.

[21] P. Erdös and G. Szekeres, A combinatorial problem in geometry, *Compositio Math.* **2** (1935), 463–470.

[22] P. Erdös and P. Turán, On a problem of Sidon in additive number theory, and on some related problems, *J. London Math. Soc.* **16** (1941), 212–215.

[23] G. Fiz Pontiveros, S. Griffiths, and R. Morris, The triangle-free process and the Ramsey number $R(3,k)$, *Mem. Amer. Math. Soc.* **263** (2020), v+125.

[24] S. W. Graham and C. J. Ringrose, Lower bounds for least quadratic nonresidues, in *Analytic number theory (Allerton Park, IL, 1989)*, *Progr. Math.*, vol. 85, Birkhäuser Boston, Boston, MA, 1990, 269–309.

[25] W. H. Haemers, Hoffman's ratio bound, *Linear Algebra Appl.* **617** (2021), 215–219.

[26] B. Hanson and G. Petridis, Refined estimates concerning sumsets contained in the roots of unity, *Proc. Lond. Math. Soc. (3)* **122** (2021), 353–358.

[27] J. H. Kim, The Ramsey number $R(3,t)$ has order of magnitude $t^2/\log t$, *Random Structures Algorithms* **7** (1995), 173–207.

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

[29] M. Krivelevich, Bounding Ramsey numbers through large deviation inequalities, *Random Structures Algorithms* **7** (1995), 145–155.

[30] H. L. Montgomery, *Topics in multiplicative number theory*, Lecture Notes in Mathematics, Vol. 227, Springer-Verlag, Berlin-New York, 1971.

[31] R. Mrazović, A random model for the Paley graph, *Q. J. Math.* **68** (2017), 193–206.

[32] D. Mubayi and J. Verstraëte, A note on pseudorandom Ramsey graphs, 2019. Preprint available at arXiv:1909.01461.

[33] D. Mubayi and J. Williford, On the independence number of the Erdős-Rényi and projective norm graphs and a related hypergraph, *J. Graph Theory* **56** (2007), 113–127.

[34] W. Samotij, Counting independent sets in graphs, *European J. Combin.* **48** (2015), 5–18.

[35] J. B. Shearer, A note on the independence number of triangle-free graphs, *Discrete Math.* **46** (1983), 83–87.

[36] I. D. Shkredov, Sumsets in quadratic residues, *Acta Arith.* **164** (2014), 221–243.

[37] J. Spencer, Asymptotic lower bounds for Ramsey functions, *Discrete Math.* **20** (1977/78), 69–76.

[38] J. Spencer, Eighty years of Ramsey $R(3,k)\ldots$ and counting!, in *Ramsey theory*, *Progr. Math.*, vol. 285, Birkhäuser/Springer, New York, 2011, 27–39.

[39] The Sage Developers, *SageMath, the Sage Mathematics Software System (Version 9.3)*, 2021. https://www.sagemath.org.