Many pentagons in triple systems

Dhruv Mubayi* Jozsef Solymosi†

Abstract

We prove that every n vertex linear triple system with m edges has at least m^6/n^7 copies of a pentagon, provided $m > 100\,n^{3/2}$. This provides the first nontrivial bound for a question posed by Jiang and Yepremyan.

More generally, for each $\ell \geq 2$, we prove that there is a constant c such that if an n-vertex graph is ε-far from being triangle-free, with $\varepsilon \gg n^{-1/3\ell}$, then it has at least $c\,\varepsilon^{3\ell} n^{2\ell+1}$ copies of $C_{2\ell+1}$. This improves the previous best bound of $c\,\varepsilon^{4\ell+2} n^{2\ell+1}$ due to Gishboliner, Shapira and Wigderson.

Our result also yields some geometric theorems, including the following. For n large, every n-point set in the plane with at least $60\,n^{11/6}$ triangles similar to a given triangle T, contains two triangles sharing a special point, called the harmonic point. In the other direction, we give a construction showing that the exponent $11/6 \approx 1.83$ cannot be reduced to anything smaller than ≈ 1.726 and improve this further to ≈ 1.773 for a 3-partite version of the problem.

1 Introduction

We consider the supersaturation problem for odd cycles in linear 3-graphs (triple systems) and show some applications of this question. A (loose or linear) cycle C_k is the 3-graph containing k distinct vertices $v_1, \ldots, v_k$ and k distinct edges $e_1, \ldots, e_k$ such that e_i is obtained by enlarging $\{v_i, v_{i+1}\}$ by a new vertex w_i such that $w_1, \ldots, w_k$ are all distinct and distinct from all the v_js. In other words (taking indices modulo k),

$$V(C_k) = \{v_1, \ldots, v_k, w_1, \ldots, w_k\} \qquad \text{and} \qquad E(C_k) = \{v_i v_{i+1} w_i : i = 1 \ldots, k\}.$$

*Department of Mathematics, Statistics and Computer Science, University of Illinois, Chicago, IL 60607. Email: mubayi@uic.edu. Research partially supported by NSF Award DMS-2153576.

†University of British Columbia, Vancouver, Canada, and Obuda University, Budapest, Hungary Email: solymosi@math.ubc.ca Research supported by an NSERC Discovery grant and OTKA K grant no. 133819.

A triple system is *linear* if every pair of vertices lies in at most one edge. A natural extremal problem is to determine the Turán number $\mathrm{ex}_L(n, C_k)$, defined as the maximum number of edges in an n-vertex linear triple system that contains no copy of C_k as a (not necessarily induced) subgraph. We are especially interested in the case when $k = 2\ell + 1$ is odd. The case of C_3 is special as determining $\mathrm{ex}_L(n, C_3)$ is, apart from constant multiplicative factors, equivalent to the well-known $(6, 3)$-problem of Brown-Erdős-Sós. Here, famous results of Behrend [2] and Ruzsa-Szemerédi [17] show that $\mathrm{ex}_L(n, C_3) = n^{2-o(1)}$ with the $o(1)$ term being a function of intense study over the years.

We are mainly concerned with the case of C_5, henceforth called the pentagon. The first author, Kostochka, and Verstraëte proved that $\mathrm{ex}_L(n, C_5) = \Omega(n^{3/2})$ while writing the paper [13] in 2013. Theorem 1.2 in Collier-Cartaino, Graber, Jiang [4] refers to this result. This was later published in [6] by Ergemlidze, Győri, Methuku. More generally, the upper bound $\mathrm{ex}_L(n, C_{2\ell+1}) = O(n^{1+1/\ell})$ was proved in [4]; no corresponding lower bound is known for any $\ell > 2$.

Many extremal problems exhibit the property that when the underlying (typically large) discrete object is dense enough to contain a given forbidden subobject, it contains many of them. In our context, this means that n-vertex triple systems with m edges contain many pentagons when m is much larger than $n^{3/2}$. Indeed, our main result quantifies this dependence.

Theorem 1.1. *Let $n > 10$ and let H be an n-vertex linear triple system with $m > 100\, n^{3/2}$ edges. Then the number of copies of C_5 in H is at least m^6/n^7.*

Sidorenko's conjecture states that the homomorphism density of a graph G in a graph W is at least the edge density of W raised to the power e, where e is the number of edges in G. This is known to be false for some hypergraphs, but deciding if it is true for the pentagon when the underlying triple system is linear seems interesting. Namely, can the quantity m^6/n^7 in Theorem 1.1 be improved to m^5/n^5, which, if true, would be sharp in order of magnitude as shown by random triple systems? This problem was posed by Jiang and Yepremyan [10]. We believed that no such improvement is possible, and in fact, after our preprint was made public, this was shown to be true by Methuku (personal communication), who gave a construction where the number of C_5 is $O((m^5/n^5)^{1-\varepsilon})$. Further, we conjecture that the truth is $\Theta(m^6/n^7)$, but this remains open. The following result provides some motivation for our conjecture. Throughout this paper, we use standard asymptotic notation.

Proposition 1.2. *Suppose that for all n there exists a linear triple system H on n vertices and $m = \Theta(n^{3/2})$ edges, with maximum degree $O(n^{1/2})$, at most $O(n^{3/2})$ copies of C_3, and at most $O(n^2)$ copies of C_k for $k = 4, 5$. Then, the bound m^6/n^7 in Theorem 1.1 is tight in order of magnitude.*

We provide a construction of a linear H as in Proposition 1.2 with *no* copies of C_3 and C_5, but the number of copies of C_4 is $\Omega(n^{5/2})$. We remark that a linear 3-graph H as in

Proposition 1.2 with *no* of copies of C_k for each $3 \le k \le 5$ does not exist, as it was shown by Conlon, Fox, Sudakov, and Zhao [3] that any such H has $o(n^{3/2})$ edges.

A natural generalization of Theorem 1.1 has an application to a problem concerning quantitative aspects of removal lemmas in graphs, which are, in turn, connected to questions about property testing. Say that an n-vertex graph G is ε-far from being triangle-free if the minimum number of edges required to be removed from G to make G triangle-free is at least εn^2. Gishboliner, Shapira and Wigderson [9] proved that there is a constant c such that if $\varepsilon > 0$ and n is sufficiently large in terms of ε, and G is ε-far from being triangle-free, then G has at least $c\,\varepsilon^{4\ell+2}n^{2\ell+1}$ copies of $C_{2\ell+1}$. We improve this as follows. We write $a \gg b$ to denote that there is a sufficiently large constant $C > 0$ such that $a > C\,b$.

Theorem 1.3. *Fix $\ell \ge 2$. There is a constant c such that if an n-vertex graph G is ε-far from being triangle-free, with $\varepsilon \gg n^{-1/3\ell}$, then G has at least $c\,\varepsilon^{3\ell}n^{2\ell+1}$ copies of $C_{2\ell+1}$.*

The lower bound requirement on ε in Theorem 1.3 can be weakened, but we make no attempt to optimize its value (the optimal value would be $n^{-1+1/\ell}$). We note that the exponent 3ℓ of ε cannot be improved to anything smaller than $2\ell + 1$ as shown by random graphs.

This paper summarizes unpublished works by the authors previously presented in seminars and conferences like in [18, 19]. The following related results were published independently by others: a proof of a slight weakening of Theorem 1.1, and Theorem 1.3 for $\ell = 2$ was published recently in [8], and, as mentioned earlier, a construction similar to the one in Section 2.1, was published in [6].

1.1 An application in geometry

Theorem 1.1 provides a somewhat unexpected application to a problem in discrete geometry that we now describe.

Elekes and Erdős proved in [5] that for any triangle T, there are n-element planar point sets S with $\Omega(n^2)$ triangles similar to T. It was proved shortly after that if the number of equilateral triangles in S is least $(1/6 + \varepsilon)n^2$, then S contains large parts of a triangular lattice. On the other hand, no lattice is guaranteed if S contains at most $c\,n^2$ similar copies for $c < 1/6$. Weaker, local structural properties of point sets with a quadratic number of similar triangles were proved in [1]. We will prove that point sets with sub-quadratic (but still many) triangles, similar to a given T, are guaranteed to contain certain interesting local substructures.

We use complex numbers to represent points of the plane. A point P with coordinates (a, b) is represented by the complex number $z_P = a + ib$. A *cyclic quadrilateral* $ABCD$ is a quadrilateral that can be inscribed in a circle. A *harmonic quadrilateral* is a cyclic quadrilateral in which the product of one pair of opposite sides is equal to the product of the other pair of opposite sides [11, 7]. This property can also be described using complex numbers and the harmonic cross-ratio:

$$(z_A, z_B; z_C, z_D) = -1, \tag{1}$$

where the cross-ratio is defined as:

$$(z_A, z_B; z_C, z_D) = \frac{(z_A - z_C)(z_B - z_D)}{(z_A - z_D)(z_B - z_C)}.$$

The cross-ratio is important in analyzing point sets with a quadratic number of triangles and quadrangles in the plane. Laczkovich and Ruzsa proved that for a quadrilateral, Q, there exist arbitrarily large point sets with a quadratic number of quadrangles similar to Q if and only if the cross-ratio of Q is algebraic [14]. We will refer to this result as the Laczkovich-Ruzsa Theorem.

A simple calculation shows that given three points z_A, z_B, and z_C in the complex plane, the fourth point z_D such that the quadrilateral is harmonic can be expressed as

$$z_D = \frac{2z_A z_B - z_A z_C - z_B z_C}{z_A + z_B - 2z_C}. \tag{2}$$

A point D is a *harmonic point* of triangle ABC if $ABCD$ forms a harmonic quadrangle. By continuity, a triangle has three (distinct) harmonic points on its circumcircle, one in each of the three sectors of the circle between vertices of the triangle (See Figure 1 for some examples). Moreover, D is the harmonic point of ABC on the opposite side of AB as C iff z_D satisfies (1) or (2).

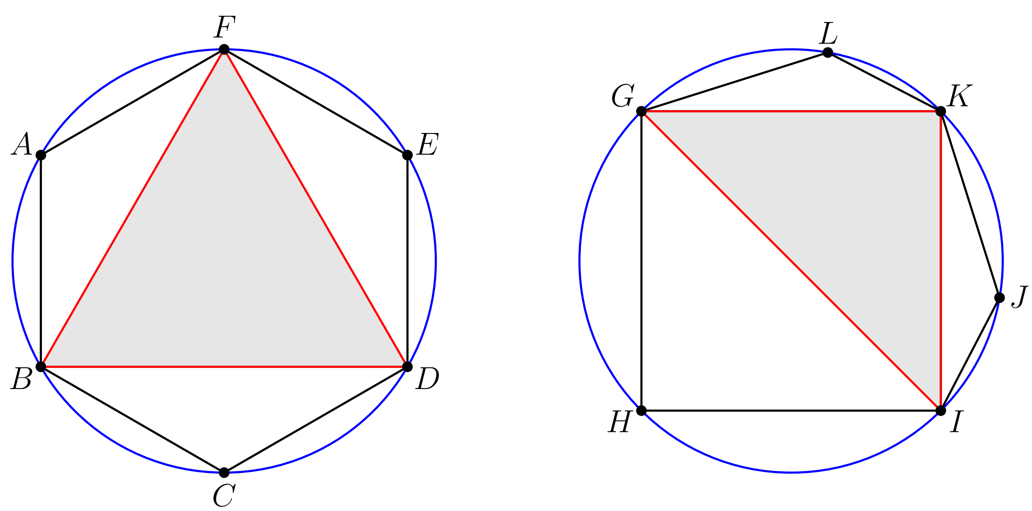

Figure 1: A, C, E are the harmonic points of the equilateral triangle BDF.
H, J, L are the harmonic points of the isosceles right triangle GIK.

By applying Theorem 1.1, we prove the following structural result about points sets with many similar copies of a given triangle T. Given a triangle ABC, call A, B, C, its vertices. Say that a point set *contains* a triangle if it contains the three vertices of the triangle.

Theorem 1.4. *Let T be a triangle and S be a set of $n > 10^6$ points in the plane such that S contains at least $m = 60\, n^{11/6}$ triangles similar to T. Then there are two triangles T_1, T_2 in S (similar to T), and a point P (not necessarily in S) such that P is the same*[1] *harmonic point of both T_1 and T_2. Moreover, the exponent $11/6 = 1.8\dot{3}$ cannot be reduced to anything less than $\approx 1.773...$.*

Our proof actually gives $\Omega(m^6/n^{11})$ pairwise vertex disjoint triangles that all share a common harmonic point. This observation leads to a stronger structural result as the number of triangles m increases. According to the Laczkovich-Ruzsa theorem, for any quadrangle with algebraic cross-ratio, there are point sets with quadratically many copies of quadrangles similar to it. Such sets also have many triangles similar to a triple of the four points of the quadrangle. We show a reverse statement that any set with quadratically many triangles similar to a given triangle T is part of a point set with many quadrangles.

Theorem 1.5. *For every $c > 0$ and $\varepsilon > 0$, there is a $D > 0$ such that the following holds for large enough n. Let T be a triangle and S be a set of n points in the plane such that S contains at least cn^2 triangles similar to T. Then, there is a quadrangle Q and a set U of at most Dn points such that U contains at least $(c-\varepsilon)n^2$ quadrangles similar to Q and $S \subset U$.*

The exponent 11/6 in Theorem 1.4 also appears in a seemingly unrelated problem investigated by Katz and Tao [12]. This is no accident, as their question (at least for real or complex numbers) can also be translated to the geometric problem above using the identity

$$\left(a, b; \frac{1-i}{2}a+\frac{1+i}{2}b, i\left(\frac{1-i}{2}a-\frac{1+i}{2}b\right)\right) = -1$$

for complex numbers $a \neq b$. The four points represented by the complex numbers on the left side form a square in the plane. We omit the discussion of the arithmetic question of Katz and Tao here but note that improvements in Theorem 1.4 would imply improvements in their bound.

2 Proof of Theorem 1.1

Given a hypergraph H, write $e(H)$ and $d(H)$ for the number of edges and average degree of H, respectively.

Proof. For each vertex $u \in V(H)$, define the graph G_u as follows: $V(G_u) = V(G) \setminus \{u\}$ and

$$E(G_u) = \{yz : \exists w, x \text{ such that } uwx, xyz \in E(H)\}.$$

Note that the linearity of H implies that the vertices u, w, x, y, z above are all distinct. Another way to define $E(G_u)$ is that it is the set of pairs $yz \in \partial H$ such that there exist

[1] two harmonic points of similar triangles are the same when there is a similarity transformation that moves one triangle into the other and also moves the harmonic points into the same point

distinct edges $e, f \in E(H)$ with $|e \cap f| = 1$, $\{y, z\} \subset f \setminus e$, and $u \in e \setminus f$. Put differently, $yz \in E(G_u)$ iff there is a linear two-edge path with edges e, f in H starting at u and ending at yz (see Figure 1).

Write $d := d(H) = 3m/n > 300\sqrt{n}$ for the average degree in H. Observe that

$$\sum_{v \in V(H)} e(G_v) = 4 \sum_{x \in V(H)} \binom{d(x)}{2} \geq 4n \binom{d}{2} \geq 10^6 n^2. \tag{3}$$

To see the equality, note that $\sum_{x \in V(H)} \binom{d(x)}{2}$ is the number of pairs of edges e, f in H with $|e \cap f| = 1$. Writing $e = abx$ and $f = a'b'x$ we see that $ab \in E(G_{a'}) \cap E(G_{b'})$ and $a'b' \in E(G_a) \cap E(G_b)$. Hence the pair $\{e, f\}$ contributes 4 to $\sum_{v \in V(H)} e(G_v)$ and this yields the equality in (3). The first inequality in (3) follows from the convexity of binomial coefficients, and the last inequality follows from $d > 300\sqrt{n}$.

Say that a path $wxyz$ in G_u with edges wx, xy, yz is a *good path* if there are distinct vertices a, b, c, a', c' in $V(H)$, such that

$$\{a, b, c, a', c'\} \cap \{u, w, x, y, z\} = \emptyset \tag{4}$$

and the following six edges all lie in $E(H)$:

$$uaa', awx, xyb, ubb', yzc, ucc'. \tag{5}$$

We note that if we exclude ubb', the remaining five edges above form a C_5. Indeed, this C_5 is an expansion of $uaxyc$ (see Figure 2).

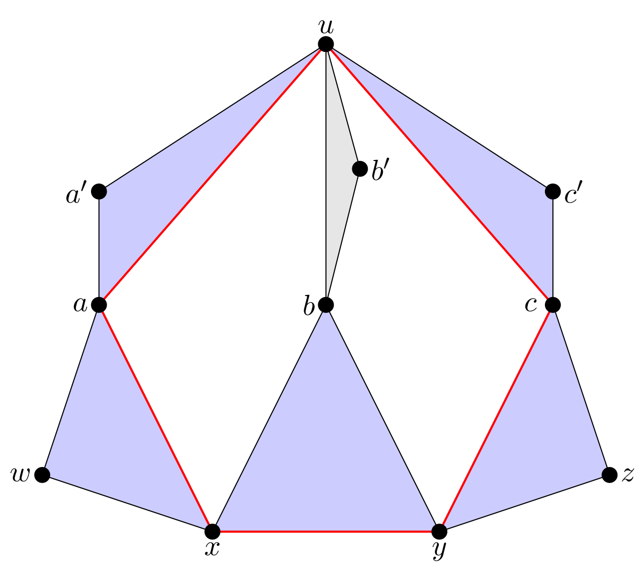

Figure 2: A good path $wxyz$ in G_u and expansion of $uaxyc$

Write p_v for the number of good paths in G_v. Since each good path gives rise to a C_5 and each C_5 is counted at most five times, we conclude that the number of C_5s in H is at least $\sum_v p_v/5$. Next, we will obtain a lower bound on $\sum_v e(G_v)$, which will in turn give a lower bound on $\sum_v p_v$.

If $e(G_v)$ is small, then $p_v = 0$ is possible, and this is not helpful for us, so we say that v is *useful* if $e(G_v) > 1000n$ and v is *useless* if $e(G_v) \le 1000n$. Note that (3) shows

$$\sum_{v \, useless} e(G_v) \le 1000n^2 < 10^{-3} \sum_{v \in V(H)} e(G_v) \tag{6}$$

so most of the contribution to $\sum_{v \in V(H)} e(G_v)$ comes from useful v and our plan is to lower bound $\sum_v p_v$ where the sum is over all useful v. To this end, let us fix a useful v and consider G_v and p_v. First, it is necessary to pass to a subgraph of G_v with a large minimum degree, so let G'_v be the subgraph of G_v that remains after iteratively deleting vertices of degree at most 100. As $e(G_v) > 1000n$, we have

$$e(G'_v) \ge e(G_v) - 100n > 0.9e(G_v). \tag{7}$$

Call a 3-edge path in G'_v a *bad path* if it is not a good path and let b'_v be the number of bad paths in G'_v. Our main claim is the following.

Claim.

$$b'_v \le \sum_{xy \in E(G'_v)} 12(d_{G'_v}(x) + d_{G'_v}(y) - 2).$$

Proof of Claim. We count bad paths of the form $wxyz$ in G'_v as follows: first, we choose the middle edge xy and then vertices w and z such that wx and yz are both in $E(G_v)$. The definition of G_v gives us (not necessarily distinct) vertices a, b, c, a', b', c' and (not necessarily distinct) edges as in (5) (see Figure 2 for an example where the vertices and edges are distinct). So we must upper bound the number of w, z such that a, b, c, a', c' are not all distinct or that (4) fails. First, we upper bound the number of $\{w, z\}$ such that $b \in \{c, c'\}$. Given any edge $wx \in E(G_v)$, the number of z such that $b \in \{c, c'\}$ is at most two due to linearity of H. Indeed, if we have three such distinct vertices, z, z', z'' then for two of them, say z and z', vertex b coincides with c or b coincides with c'. If b and c coincide, then the pair $yb = yc$ lies in two distinct edges ybz and ybz', and if b and c' coincide, then the pair $vb = vc'$ lies in two distinct edges; in either case, this contradicts the linearity of H. Hence, the number of such bad paths is at most $2(d_{G_v}(x) - 1)$. Arguing similarly for x, we deduce that the number of bad paths such that $b \in \{a, c, a', c'\}$ is at most $2(d_{G_v}(x) + d_{G_v}(y) - 2)$.

Next we consider bad paths such that $\{a, a'\} \cap \{c, c'\} \ne \emptyset$ and $b \notin \{a, c, a', c'\}$. For each choice of w, the number of choices of z such that the corresponding vertex c lies in $\{a, a'\}$ is at most two by the linearity of H. Hence the number of bad paths with $c \in \{a, a'\}$ is at most $2(d_{G_v}(x) - 1)$. We argue similarly if c is replaced by c' and if the roles of z and w are interchanged. We conclude that the number of bad paths with a, c, a', c' not all distinct is at most $4(d_{G_v}(x) + d_{G_v}(y) - 2)$. As we have assumed $b \notin \{a, c, a', c'\}$, the number of bad paths with a, b, c, a', c' not all distinct is at most $6(d_{G_v}(x) + d_{G_v}(y) - 2)$.

We now consider bad paths containing xy for which a, b, c, a', c' are all distinct that fail (4). Given a choice of w, and hence of distinct a, a', b, the number of $z \in \{w, a, a', b\}$ is at most four, since if there are five such distinct z, then two of them coincide with one of $\{w, a, a', b\}$, which is impossible. Hence the number of bad paths such that a, b, c, a', c' are all distinct and $z \in \{w, a, a', b\}$ is at most $4(d_{G_v}(x) - 1)$. Similarly, the number of z such that $w \in \{c, c'\}$ is at most $2(d_{G_v}(x) - 1)$. Reversing the roles of w and x we obtain that the number of bad paths containing xy for which a, b, c, a', c' are all distinct that fail (4) is at most $6(d_{G_v}(x) + d_{G_v}(y) - 2)$. Altogether, the number of bad paths containing xy is at most $12(d_{G'_v}(x) + d_{G'_v}(y) - 2)$ and the proof of the claim is complete. □

Let s'_v be the number of 3-edge paths in G'_v. Then

$$s'_v \geq \sum_{xy \in E(G'_v)} (d_{G'_v}(x) - 2)(d_{G'_v}(y) - 2)$$

as we count paths by picking a neighbor w of x that is not y and then a neighbor z of y that is not w or x. Assume by symmetry that $d_{G'_v}(x) \geq d_{G'_v}(y)$. As the minimum degree in G'_v is at least 100,

$$\begin{aligned}(d_{G'_v}(x) - 2)(d_{G'_v}(y) - 2) &\geq \frac{d_{G'_v}(x) + d_{G'_v}(y) - 4}{2}(d_{G'_v}(y) - 2) \\ &\geq 49(d_{G'_v}(x) + d_{G'_v}(y) - 4) \\ &> 48(d_{G'_v}(x) + d_{G'_v}(y) - 2).\end{aligned}$$

Consequently, the Claim implies that

$$s'_v \geq \sum_{xy \in E(G'_v)} (d_{G'_v}(x) - 2)(d_{G'_v}(y) - 2) > \sum_{xy \in E(G'_v)} 48\,(d_{G'_v}(x) + d_{G'_v}(y) - 2) \geq 4b'_v.$$

Write p'_v for the number of good paths in G'_v. Then $s'_v = p'_v + b'_v$, so

$$p_v \geq p'_v = s'_v - b'_v \geq (0.75)s'_v. \tag{8}$$

The number of 3-edge paths in an n' vertex graph with e' edges and average degree $d' = 2e'/n' > 100$ is at least

$$\frac{(e')^3}{10n'^2}. \tag{9}$$

Indeed, to see this, first iteratively delete vertices of degree at most $d'/4$ until no such vertices remain. The remaining graph has at least $e' - n'd'/4 = e'/2$ edges. Now pick an edge and then a neighbour of each of its endpoints to get at least $(e'/2)(d'/4 - 2)^2 > (e')^3/(10n'^2)$ paths.

Recall from (7) that G'_v is a graph with $n' \leq n$ vertices and at least $(0.9)e(G_v) \geq 900n$ edges, where the last inequality holds because v is useful. Hence by (8) and (9),

$$p_v \geq (0.75)s'_v \geq (0.75)\frac{e(G'_v)^3}{10n^2} \geq (0.75)\frac{(0.9)^3 e(G_v)^3}{10n^2} > \frac{e(G_v)^3}{20n^2}.$$

From (6), (3) and $d = 3m/n$, we obtain

$$\sum_{v\, useful} e(G_v) \geq (0.99) \sum_{v \in V(H)} e(G_v) \geq (0.99)nd^2 > \frac{8m^2}{n}.$$

Hence, by convexity,

$$\sum_{v\in V(H)} p_v \ge \sum_{v\,usefull} p_v \ge \sum_{v\,usefull} \frac{e(G_v)^3}{20n^2} \ge \frac{1}{20n}\left(\frac{\sum_{v\,usefull} e(G_v)}{n}\right)^3 > \frac{5m^6}{n^7}.$$

The number of C_5 in H is at least $(1/5)\sum_v p_v$, so the proof is complete. □

2.1 Constructions of pentagon-free triple systems

As mentioned earlier, the bound $m \gg n^{3/2}$ in Theorem 1.1 is sharp. Indeed, Kostochka, the first author and Verstraëte constructed a linear n-vertex 3-graph with $\Omega(n^{3/2})$ edges and no C_5. We present this construction below as it has not been published before.

Construction: Let $T_3(n)$ be the complete 3-partite graph with parts X, Y, Z each of size n. Form the 3-partite linear triple system H of $T_3(n)$ as follows:

$$V(H) = (X \times Y) \cup (Y \times Z) \cup (X \times Z)$$

$$E(H) = \{\{xy, yz, xz\} : (x, y, z) \in X \times Y \times Z\}.$$

Observe that $N := |V(H)| = 3n^2$ and $|E(H)| = n^3 = (N/3)^{3/2}$. Clearly, H is linear as any two vertices of H that lie in an edge e of H uniquely determine the third vertex of e. For example, xy and yz uniquely determine xz.

Next, we prove that H contains no C_5. Here, it is convenient to view $E(H)$ as a set of vectors in $\mathbb{R}^3$ of the form (x, y, z) (rather than sets of the form $\{xy, yz, xz\}$) and use geometric arguments. Now suppose that there is a C_5 with edges $e_1, e_2, \ldots, e_5$ in cyclic order. This means that $e_i \cap e_{i+1}$ are all distinct of size one, and there are no other intersections among the e_is. The list $e_1, \ldots, e_5$ gives rise to a closed walk W of length five in the 3-dimensional grid $\mathbb{Z}^3$. If some two vertices v, w of W differ in all three coordinates, then the distance between them on W is at least three, so it is impossible to go from v and w and then back in five steps. Hence, we may assume that W is planar and no two consecutive edges of W are in the same axis (as this corresponds to three edges e_i, e_{i+1}, e_{i+2} that all share the same vertex). However, any such closed walk in a planar grid must have an even length. We conclude that H contains no C_5. □

We conjecture below that Theorem 1.1 is tight.

Conjecture 2.1. *For $n^{3/2} \ll m \ll n^2$, there exists an n-vertex linear 3-graph in which the number of copies of C_5 is $O(m^6/n^7)$.*

As mentioned in the introduction, Proposition 1.2 provides some evidence for Conjecture 2.1.

Proof of Proposition 1.2. Recall that we are given a linear H on n vertices and $m = \Theta(n^{3/2})$ edges, with maximum degree $O(n^{1/2})$, the number of C_3 in H is $O(n^{3/2})$, and for $4 \le k \le 5$, the number of C_k in H is $O(n^2)$. We let $H(t)$ be the 3-graph obtained from H by replacing each vertex of H by a t-set of vertices and by replacing each edge of H by a linear 3-partite 3-graph with t vertices in each part and t^2 edges. Then $H(t)$ has $N = nt$ vertices and $M = mt^2$ edges. Moreover, it is a short exercise to see that the degree two vertices of each C_5 in $H(t)$ come from the following three structures in H:

1. degree two vertices of C_3's in H together with an additional edge intersecting the C_3
2. degree two vertices of C_k's in H for $k = 4$ or $k = 5$
3. paths of length at most two in H.

The number of C_5s that arise from C_3 plus edges is by hypothesis $O(n^{3/2}n^{1/2}t^5) = O(n^2t^5)$, the number of C_5s that arise from C_k for $k = 4, 5$ is $O(n^2t^5)$ and the number of C_5s arising from single edges and two edge paths is $O(mt^5) + O(n^2t^5) = O(n^2t^5)$. So the total number of C_5 in $H(t)$ is at most $O(n^2t^5) = O(M^6/N^7)$. □

We point out that our construction $T_3(n)$ has the required properties in the hypothesis of Proposition 1.2 except that the number of C_4 is $\Theta(n^{5/2})$. Indeed, $T_3(n)$ is linear with no copies of C_3 and C_5.

3 Generalization to longer cycles

In this section, we generalize Theorem 1.1 to longer cycles and show a connection to removal lemmas. The *shadow graph* G of a triple system H is defined as follows:

$$V(G) = V(H) \qquad \text{and} \qquad E(G) = \partial H = \{yz : \exists x \text{ with } xyz \in E(H)\}.$$

A linear path is a 3-graph obtained from a linear cycle by deleting exactly one edge. Given a linear path P in a 3-graph, say that a vertex is an endpoint of P if it lies in the first or last edge of P and it has degree one on P. In the theorem below, we use asymptotic notation and assume, wherever needed, that n is sufficiently large. In particular, $a \gg b$ means that $a > Cb$ for some sufficiently large constant $C > 0$.

Theorem 3.1. *Let $k \geq 2$ be an integer and let H be an n-vertex linear triple system with $m \gg n^{2-1/3k}$ edges. Then the shadow graph G of H contains at least m^{3k}/n^{4k-1} copies of C_{2k+1}.*

Proof. For each vertex $u \in V(H)$, define the multigraph G_u as follows. Let $V(G_u) = V(H) \setminus \{u\}$. Next, let P be a k-edge linear path $e_1, \ldots, e_k$ in H with endpoint $u \in e_1$ and $e_k = xyz$ where y, z are endpoints of P. Then the edge $e_u(P) = yz$ is an edge of G_u. We emphasize that G_u is a multigraph; indeed, the pair yz can arise many times in $E(G_u)$ due to many paths P from u, and we distinguish the edges comprising the pair depending on the path (see Figure 3).

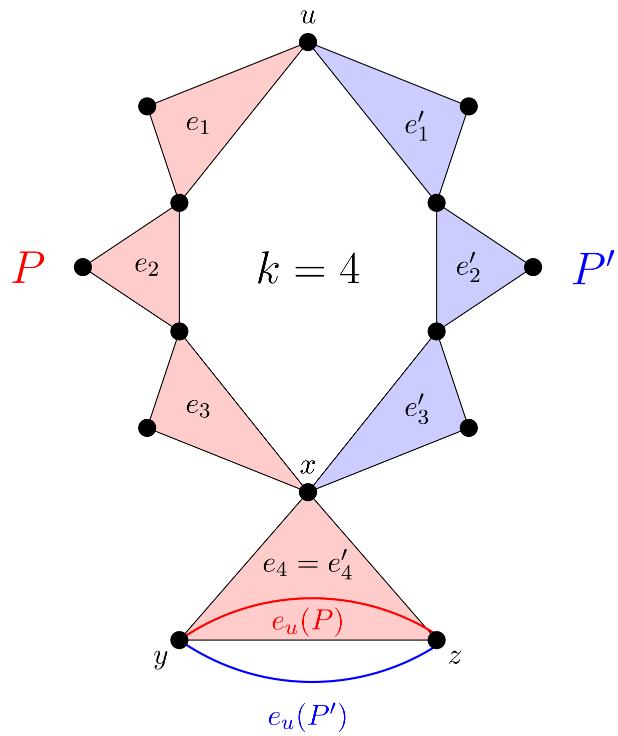

Figure 3: The multigraph G_u

Write $e_u = |E(G_u)|$ and d for the average degree of H, so that $d \gg n^{1-1/3k}$. The number of k-edge linear paths in H is at least $\Omega(nd^k)$; to see this, let $H' \subset H$ be the 3-graph that remains after iteratively removing vertices of degree at most $d/2$ from H, and then build paths by starting with any edge e_1 of H' and then greedily choosing edges $e_2, \ldots, e_k$, where we have at least $d/2 - 2i > d/4$ choices for each e_i.

The quantity $\sum_u e_u$ is the number of pairs $(u, e_u(P))$ where u is a vertex and P is a k-edge path with endpoint u. The number of k-edge paths P is $\Omega(nd^k)$ and each such P gives rise to four pairs $(u, e_u(P))$. A given pair $(u, e_u(P))$ cannot arise from a k-edge path different from P. As $d \gg n^{1-1/3k}$,

$$\sum_{u \in V(H)} e_u = \Omega(nd^k) \gg n^2. \tag{10}$$

As before, say that u is *useless* if $e_u < 100n$ and useful otherwise. Then $\sum_{u\,useless} e_u < 100n^2 \ll \sum_{u \in V(H)} e_u$, so $\sum_{u\,useful} e_u = \Omega(nd^k)$.

A 3-edge path in the multigraph G_v is a set of four (not necessarily distinct) vertices v_1, v_2, v_3, v_4 and three distinct edges e_1, e_2, e_3 such that $e_i = v_i v_{i+1}$ for $i = 1, 2, 3$ (the pair might appear with multiplicity greater than one, but the edges are distinct). For every useful v the number of 3-edge paths p_v in G_v is at least $\Omega(e_v^3/n^2)$. Indeed, since v is useful, $e_v \geq 100n$ and G_v has average degree $d_v \geq 300$, so we restrict to a subgraph of minimum degree at least $d_v/4$ and then build a 3-edge path greedily. There are at least $e_v/4$ choices for the middle edge and at least $d_v/4 - 2 > d_v/5 > (3e_v/5n)$ choices for each of the other two edges. Consequently, by (10),

$$\sum_{v\in V(H)} p_v = \Omega\left(\frac{\sum_v e_v^3}{n^2}\right) = \Omega\left(\frac{(\sum_v e_v)^3}{n^4}\right) = \Omega\left(\frac{(nd^k)^3}{n^4}\right) = \Omega\left(\frac{d^{3k}}{n}\right) = \Omega\left(\frac{m^{3k}}{n^{3k+1}}\right). \quad (11)$$

An ℓ-*pseudocycle* is a homomorphic image of an ℓ-cycle in G. Suppose that $wxyz$ is a 3-edge path in G_v with edges $e_v(P^1) = wx, e_v(P^2) = xy, e_v(P^3) = yz$. Let $P^i = e_1^i, \dots, e_k^i$ denote the k-path in H from v to the edge $e_v(P^i) \in G_v$ for $i = 1, 2, 3$. For $1 \le j \le k-1$, let $v_j^i = e_j^i \cap e_{j+1}^i$. Each 3-edge path P in G_v with vertices w, x, y, z as above gives rise to the following $(2k+1)$-pseudocycle C_v in G (see Figure 4).

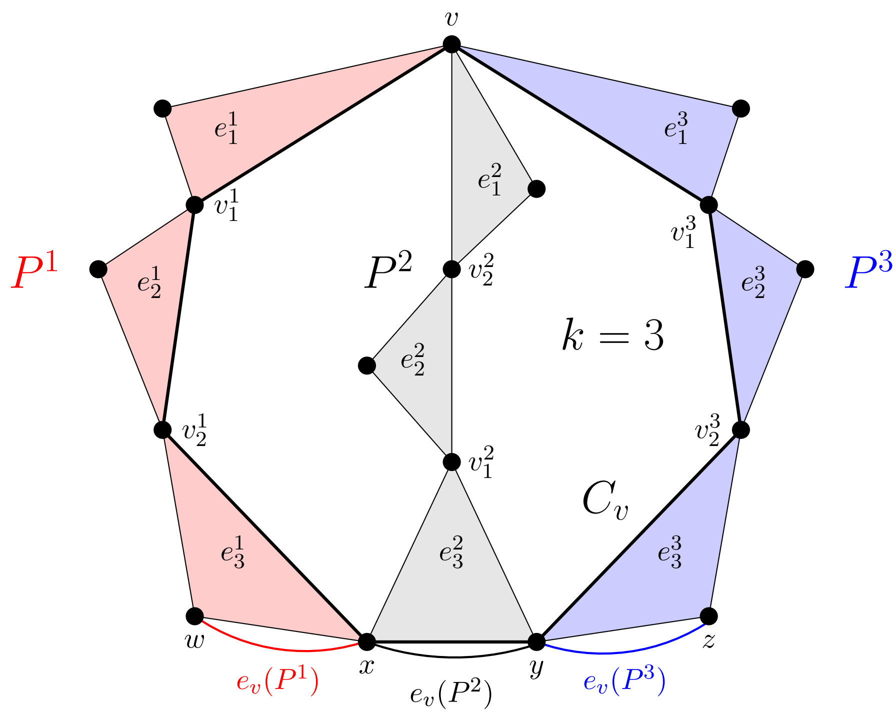

Figure 4: A 7-pseudocycle C_v

The vertices of C_v, in cyclic order, are

$$v, v_1^1, v_2^1, \dots, v_{k-1}^1, x, y, v_{k-1}^3, v_{k-2}^3, \dots, v_1^3.$$

We emphasize that C_v is a $(2k+1)$-pseudocycle as vertices can be repeated.

Given v and $e = xy$, the number of k-edge paths P^2 in H starting at v and ending at $e = e_v(P^2)$ is at most n^{k-2}. This is because there is at most one choice for the vertex v_{k-1}^2 as H is linear, there are at most $k-2$ other vertices of degree two on P^2, and once these are chosen, the path P^2 is determined again due to linearity of H. Hence each $(2k+1)$-pseudocycle C_v obtained from $wxyz$ is counted at most n^{k-2} times. Consequently, by (11), the number of $(2k+1)$-pseudocycles in G is at least

$$\frac{\sum_{v\in V(H)} p_v}{n^{k-2}} = \Omega\left(\frac{m^{3k}}{n^{4k-1}}\right).$$

The number of these $(2k+1)$-pseudocycles with fewer than $2k+1$ vertices is at most $n^{2k} \ll m^{3k}/n^{4k-1}$ as $m \gg n^{2-1/3k}$. Hence the number of copies of C_{2k+1} in G is at least m^{3k}/n^{4k-1} by adjusting the constant in the hypothesis $m \gg n^{2-1/3k}$. $\square$

We remark that with more care, Theorem 3.1 can be extended to find Berge cycles in H instead of just cycles in the shadow graph G (the additional requirement is that the edges of the cycle are distinct). We wrote the technically simpler argument that finds only cycles in the shadow graph as it suffices for the application below, which restates Theorem 1.3.

Corollary 3.2. *Fix $k \geq 2$. There is a constant c such that if an n-vertex graph G is ε-far from being triangle-free, with $\varepsilon \gg n^{-1/3k}$, then G has at least $c\,\varepsilon^{3k}n^{2k+1}$ copies of C_{2k+1}.*

Proof. Let H be a maximal collection of edge-disjoint triangles in G. View H as a 3-graph whose edges are the triangles. Because the triangles in G are edge-disjoint, H is linear. Moreover, if H has m (hyper)edges, then by maximality, we can delete $3m$ edges in G so that the resulting graph is triangle-free. As G is ε-far from being triangle-free, $m \geq \varepsilon n^2/3 \gg n^{2-1/3k}$. Since G contains the shadow graph of H, by Theorem 3.1, the number of C_{2k+1} in G is at least $\Omega(m^{3k}/n^{4k-1}) = \Omega(\varepsilon^{3k}n^{2k+1})$. $\square$

4 Proof of Theorem 1.4

In this section, we use Theorem 1.1 to prove Theorem 1.4. Say that a triangle lies in a set if its three vertices are in the set. Suppose $n > 10^6$ and S is a set of n points and there are $m > 60n^{11/6}$ triangles in S similar to a given triangle $T = (A, B, C)$. Partition S randomly into three sets, V_A, V_B, V_C, where we place each point of S into one of the sets with equal probability $1/3$. The expected number of triangles $A'B'C'$ in S similar to T with $A' \in V_A, B' \in V_B, C' \in V_C$ such that there is a similarity transformation $A'B'C' \to ABC$ with $A \to A', B \to B', C \to C'$ is $m/27$. Therefore, there is a particular choice of V_A, V_B, V_C such that the number of triangles $A'B'C'$ as above is at least $m/27$. We will also need the family of similar triangles to have the same orientation. There are at least

$$m' \geq m/54 > n^{11/6} > 100n^{3/2}$$

such triangles.

Let H be the 3-partite 3-graph where $V(H) = S$ and $E(H)$ is the set of triangles in S similar to T. Then H is linear with n vertices and $m' > 100n^{3/2}$ edges, so by Theorem 1.1, the number of linear C_5's (henceforth pentagons) in H is at least

$$\frac{m'^6}{n^7} > \frac{n^{11}}{n^7} = n^4. \tag{12}$$

The cycle of a pentagon is the (unique) graph cycle in the shadow graph of the pentagon. Every pentagon P in H has one degree-two vertex of its cycle C in one of the three vertex classes and two degree-two vertices in each of the remaining two vertex classes. For a given pentagon P, suppose that V_A and V_B have two degree-two vertices and V_C has one degree-two vertex (See Figure 5).

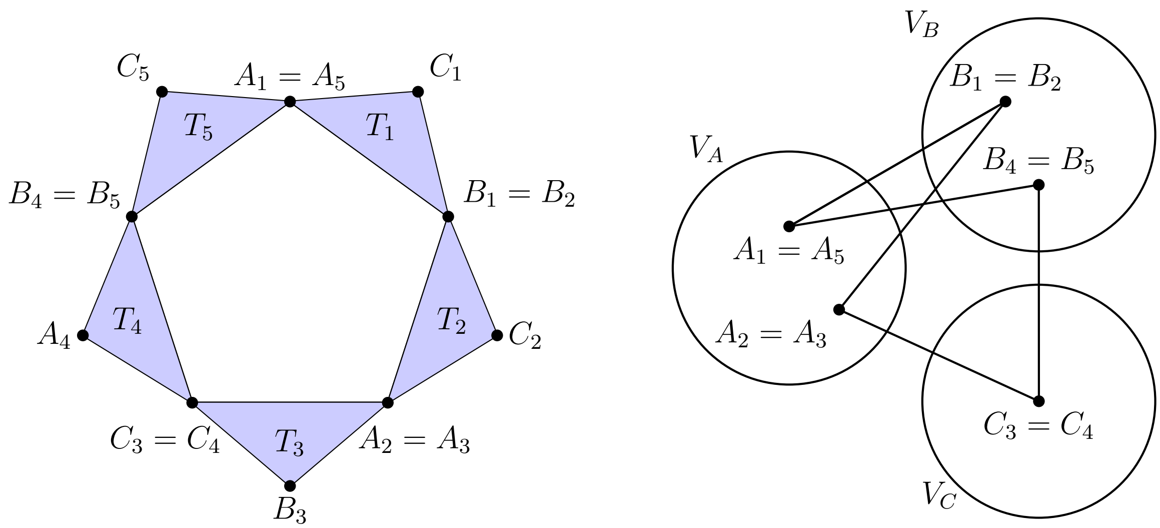

Figure 5: Triangles forming a pentagon

Denote the five triangles of the pentagon by $T_1, \ldots, T_5$, in cyclic order, and the vertices of T_j by A_j, B_j, C_j. Then, the vertices of the cycle C of the pentagon P in cyclic order are

$$A_1(= A_5), B_1(= B_2), A_2(= A_3), C_3(= C_4), B_4(= B_5). \tag{13}$$

Note that the five degree-one points of P, in cyclic order, are

$$C_1, C_2, B_3, A_4, C_5.$$

The first four of theses, C_1, C_2, B_3, A_4, are vertices of $T_1, \ldots T_4$, respectively, and C_5 is a vertex of T_5. To prove the theorem, we show the following lemma.

Lemma 4.1. *The four points C_1, C_2, B_3, A_4 in P determine a harmonic point of the fifth triangle T_5.*

This will complete our proof of Theorem 1.4 since we may associate to each pentagon P its four points as in the claim. By pigeonhole, using (12), there are two pentagons P, P' that are associated to the same four points C_1, C_2, B_3, A_4. The fifth triangles T_5 of P and T_5' of P' then have the same harmonic points. Moreover, T_5 and T_5' have distinct points in V_A and in V_B, as any one of these points determines the pentagon if we are also given C_1, C_2, B_3, A_4 as the degree one points. Further, the two quadrangles given by the two triangles T_5 and T_5' and their common harmonic point are similar, so if they share two vertices (with the same labels), they are the same. Therefore, T_5 and T_5' are, in fact, vertex disjoint.

Proof of Lemma 4.1. For the sake of simplicity, the complex number z_P is denoted by the point P in the following calculations. The triangles $T_1, \ldots, T_5$ are similar, so the vertex

C_j can be expressed as the following linear combination of A_j, B_j where $z = z(T)$ depends only on T:

$$C_j = \frac{A_j + B_j}{2} + \frac{z(A_j - B_j)}{2}, \quad \text{where} \quad z = re^{i\theta} \tag{14}$$

To see that (14) holds, note that

$$z = \frac{2C_j - (A_j + B_j)}{A_j - B_j}.$$

Multiplying each of A_j, B_j, C_j by $re^{i\alpha}$ clearly leaves z unchanged, which means that dilating and rotating $A_jB_jC_j$ by a factor r and an angle α preserves z. Adding $w = s + ti$ to each of A_j, B_j, C_j also leaves z unchanged. Since every triangle similar to $A_jB_jC_j$ with the same orientation is obtained by dilating, rotating and shifting, z indeed depends only on T.

Using parameter z, we can express a harmonic point of $A_5B_5C_5$ as

$$D_5 = \frac{A_5 + B_5}{2} + \frac{A_5 - B_5}{2z}.$$

We will use this expression for the calculations, but first, let us confirm that the expression of the harmonic point above agrees with the definition. Given

$$C = \frac{A + B}{2} + \frac{z(A - B)}{2}, \quad D = \frac{A + B}{2} + \frac{A - B}{2z},$$

let us show that the cross-ratio is -1. The differences are

$$A - C = \frac{(1 - z)(A - B)}{2}, \quad B - D = \frac{B - A}{2}\left(1 + \frac{1}{z}\right),$$

$$A - D = \frac{(1 - \frac{1}{z})(A - B)}{2}, \quad B - C = \frac{(1 + z)(B - A)}{2}.$$

The cross-ratio becomes:

$$(A, B; C, D) = \frac{\frac{(1-z)\left(1+\frac{1}{z}\right)(A-B)(B-A)}{4}}{\frac{\left(1-\frac{1}{z}\right)(1+z)(A-B)(B-A)}{4}} = -1.$$

We now show that the points A_4, B_3, C_1, C_2 determine a harmonic point of the fifth triangle T_5 by proving

$$D_5 = \frac{A_4 + B_3}{2} + \frac{A_4 - B_3}{2z} + C_1 - C_2.$$

Using $C_3 = C_4$, we obtain

$$
\begin{aligned}
&\frac{A_3+B_3}{2}+\frac{z(A_3-B_3)}{2}=\frac{A_4+B_4}{2}+\frac{z(A_4-B_4)}{2}\\
\iff &A_3\left(\frac{1+z}{2}\right)+B_3\left(\frac{1-z}{2}\right)=A_4\left(\frac{1+z}{2}\right)+B_4\left(\frac{1-z}{2}\right)\\
\iff &A_3\left(\frac{1+z}{2}\right)-B_4\left(\frac{1-z}{2}\right)=A_4\left(\frac{1+z}{2}\right)-B_3\left(\frac{1-z}{2}\right)\\
\iff &A_3\left(\frac{\frac{1}{z}+1}{2}\right)-B_4\left(\frac{\frac{1}{z}-1}{2}\right)=A_4\left(\frac{\frac{1}{z}+1}{2}\right)-B_3\left(\frac{\frac{1}{z}-1}{2}\right)\\
\iff &\frac{A_3+B_4}{2}+\frac{A_3-B_4}{2z}=\frac{A_4+B_3}{2}+\frac{A_4-B_3}{2z}.
\end{aligned}
$$

By considering the difference C_1-C_2, the other equation is

$$
\begin{aligned}
C_1-C_2&=\frac{A_1+B_1}{2}+\frac{z(A_1-B_1)}{2}-\left(\frac{A_2+B_2}{2}+\frac{z(A_2-B_2)}{2}\right)\\
&=\frac{A_1}{2}-\frac{A_2}{2}+\frac{A_1}{2z}-\frac{A_2}{2z}=\frac{A_5}{2}-\frac{A_3}{2}+\frac{A_5}{2z}-\frac{A_3}{2z}.
\end{aligned}
$$

Putting the two calculations together, we obtained the required equality

$$
\begin{aligned}
\frac{A_4+B_3}{2}+\frac{A_4-B_3}{2z}+C_1-C_2&=\frac{A_3+B_4}{2}+\frac{A_3-B_4}{2z}+\frac{A_5}{2}-\frac{A_3}{2}+\frac{A_5}{2z}-\frac{A_3}{2z}\\
&=\frac{A_5+B_4}{2}+\frac{A_5-B_4}{2z}\\
&=\frac{A_5+B_5}{2}+\frac{A_5-B_5}{2z}=D_5. \qquad \square
\end{aligned}
$$

4.1 Two geometric constructions

First, we give a simple arrangement of $n^{1.726\ldots}$ isosceles right triangles on n points without a pair sharing their harmonic point (points H, J, L in Figure 1). Our construction is based on Ruzsa's trick "*much-more-differences-than-sums*" [16] (see [12] for another application of this method). We also provide a modified construction with $n^{1.773\ldots}$ triangles without two sharing a selected harmonic point (points L or J in Figure 1). It proves the second part of Theorem 1.4.

Both exponents can be improved using more advanced construction with similar techniques like in [15], but the improvements are minor so we keep the simpler ones.

The bases of the triangles are spanned between two point sets, A and B, along the axes. The following sums of complex numbers define the elements of the sets ($s \neq 1$ is a constant we will specify later; in fact, for concreteness, we will take $s=2$ though we leave the variable s in the proof for clarity of presentation):

$$
A=\left\{\sum_{k=1}^{3m} a_k 13^k : a_k \in \{1,s\}, |\{k : a_k=1\}|=2m\right\},
$$

$$B = \left\{ \sum_{k=1}^{3m} b_k 13^k : b_k \in \{i, is\}, |\{k : b_k = i\}| = m \right\}.$$

Note that elements of A and B are determined uniquely by the coefficients a_k, b_k. With this definition, $|A| = |B| = \binom{3m}{m}$.

In our construction, two points, $\alpha = \sum_{k=1}^{3m} a_k 13^k \in A$ and $\beta = \sum_{k=1}^{3m} b_k 13^k \in B$ form the base of a triangle if

$$(a_k, b_k) \neq (s, i) \text{ for all } k \in [3m].$$

Each $\alpha \in A$ forms a base with $\binom{2m}{m}$ distinct $\beta \in B$. Indeed, there are $2m$ coordinates where $a_k = 1$ and m coordinates where $a_k = s$. In these latter m coordinates $b_k = si$, so in the former $2m$ coordinates, $b_k = i$ for exactly m of them.

The third point of the triangle, denoted γ, is uniquely determined by α and β as

$$\gamma = \frac{\alpha + \beta}{2} + i\frac{\beta - \alpha}{2} = \sum_{k=1}^{3m} \frac{a_k(1-i) + b_k(1+i)}{2} 13^k = \sum_{k=1}^{3m} g_k 13^k.$$

The angle at γ is the right angle of triangle $\alpha\beta\gamma$, and γ is below the base $\alpha\beta$. The set of these γ's is denoted by C. The possible values of the g_k's are 0 and $\frac{(1-s)(1-i)}{2}$, as indicated in the tableau below. Moreover, exactly $2m$ values are 0.

$a_k \backslash b_k$	i	is
1	0	$\frac{(1-s)(1-i)}{2}$
s	nil	0

With these definitions we have $|A| = |B| = |C| = \binom{3m}{m}$. We noted earlier that any $\alpha \in A$ is the vertex of $\binom{2m}{m}$ triangles, so the total number of triangles is

$$\binom{3m}{m}\binom{2m}{m}.$$

It remains to prove that the harmonic points of those selected triangles are distinct. Given triangle $T = \alpha\beta\gamma$, write δ_α for the harmonic point of T that lies on the opposite side of segment $\beta\gamma$ as α, write δ_β for the harmonic point of T that lies on the opposite side of segment $\alpha\gamma$ as β, and write δ_γ for the harmonic point of T that lies on the opposite side of segment $\alpha\beta$ as γ.

The cross-ratio conditions for these points are the following:

$$(\alpha, \beta; \gamma, \delta_\gamma) = -1 \qquad (\gamma, \alpha; \beta, \delta_\beta) = -1 \qquad (\beta, \gamma; \alpha, \delta_\alpha) = -1.$$

Let us first analyze the case $\delta = \delta_\gamma$. In this case, $(\alpha, \beta; \gamma, \delta) = -1$ yields

$$\begin{aligned}\delta &= \frac{2\alpha\beta - \alpha\gamma - \beta\gamma}{\alpha + \beta - 2\,\gamma} \\ &= \frac{\alpha+\beta}{2} + \frac{\beta-\alpha}{2i} \\ &= \frac{\alpha+\beta}{2} - i\frac{\beta-\alpha}{2} \\ &= \sum_{k=1}^{3m} \frac{a_k(1+i) + b_k(1-i)}{2} 13^k \\ &= \sum_{k=1}^{3m} d_k 13^k.\end{aligned}$$

Note that we could immediately have obtained the third display $\delta = (\alpha+\beta)/2 - i(\beta-\alpha)/2$ by observing that $\alpha, \beta, \gamma, \delta$ form the corners of a square with diagonal $\alpha\beta$ so we obtain δ from the midpoint of the segment $\alpha\beta$ by moving in the direction opposite to that of γ. The possible values of the d_k's are $1+i$, $(1+s)(1+i)/2$, $s+is$, as indicated in the tableau below.

$a_k \setminus b_k$	i	is
1	$1+i$	$\frac{(1+s)(1+i)}{2}$
s	nil	$s+is$

For any

$$\delta \in \left\{ \sum_{k=1}^{3m} d_k 13^k : d_k \in \left\{ 1+i, \frac{(1+s)(1+i)}{2}, s+is \right\} \right\},$$

there is a triangle $\alpha\beta\gamma$ with harmonic point δ. As $s \neq 1$, from the digits of a harmonic point, we can uniquely recover the α, β base points of the triangles so no two triangles share their harmonic points. All of these harmonic points δ_γ are in the positive quadrant, and the others are outside this quadrant, so they do not overlap. To see that, note that the circumcircle of triangle $\alpha\beta\gamma$ goes through the origin and the points $\delta_\alpha, \delta_\gamma$ lie in the arc of this circle between $\beta\gamma$ and between $\alpha\gamma$. Both these arcs are outside the first quadrant (see Figure 6).

There are two more harmonic points to consider for each triangle.

Recall $\gamma = (\alpha+\beta)/2 + i(\beta-\alpha)/2$, and $(\gamma, \alpha; \beta, \delta_\beta) = -1$. Consequently,

$$\begin{aligned}\delta_\beta &= \frac{2\alpha\gamma - \beta\gamma - \beta\alpha}{\alpha + \gamma - 2\beta} \\ &= \frac{2\alpha^2 - \alpha\beta - \beta^2 + i(3\alpha\beta - 2\alpha^2 - \beta^2)}{(\alpha-\beta)(3-i)} \\ &= \frac{(2-2i)\alpha + (1+i)\beta}{3-i} \\ &= \frac{4-2i}{5}\alpha + \frac{1+2i}{5}\beta.\end{aligned}$$

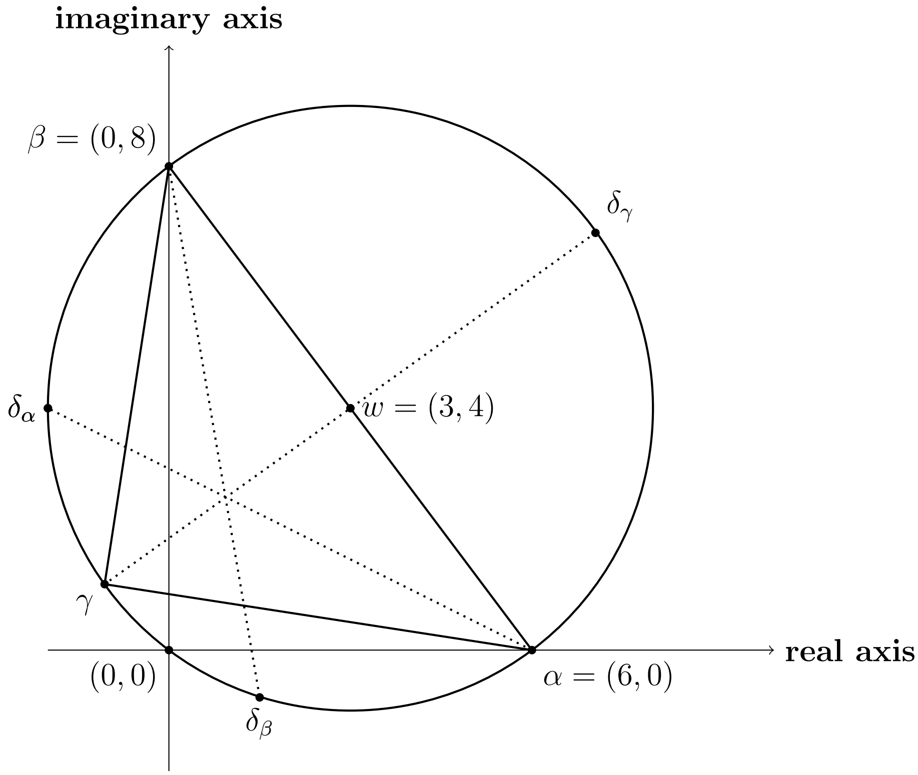

Figure 6: Example of triangle $\alpha\beta\gamma$ and its harmonic points

For the sake of simplicity, we will count the number of different $\delta' = 5\delta_\beta$ values. As before, we check the results digit-wise of $\delta' = (4-2i)\alpha + (1+2i)\beta$ for the possible α, β combinations. The results are summarized in the tableau below.

$a_k \setminus b_k$	i	is
1	$2-i$	$(2-s)(2-i)$
s	nil	$s(2-i)$

As before, by the digits of δ' we can recover the base of the triangle uniquely.

The harmonic point δ_α is obtained by reflecting δ_β about the line segment $w\gamma$, where $w = (\alpha+\beta)/2$ is the midpoint of the base, and $\gamma = (\alpha+\beta)/2 + i(\beta-\alpha)/2$ is the third point of the triangle (see Figure 6). An easy calculation now yields

$$\delta_\alpha = \frac{1-2i}{5}\alpha + \frac{4+2i}{5}\beta.$$

We set $\delta'' = 5\delta_\alpha = (1-2i)\alpha + (4+2i)\beta$ for the remaining harmonic point. The possible digit-wise entries of δ'' are

$a_k \setminus b_k$	i	is
1	$-1+2i$	$(1-2s)(1-2i)$
s	nil	$s(2i-1)$

We want to choose s such that the sets of points of the two harmonic points are disjoint. It can be achieved for example by setting $s = 2$, when the digits of δ' are $2 - i, 0, 4 - 2i$ and of δ'' are $-1 + 2i, -3 + 6i, 4i - 2$.

In the construction there are $n = 3\binom{3m}{m}$ points and $\binom{3m}{m}\binom{2m}{m}$ triangles with disjoint harmonic points. Define x as

$$\left(3\binom{3m}{m}\right)^x = \binom{3m}{m}\binom{2m}{m}.$$

Taking logarithms and letting $m \to \infty$ leads to

$$x = 1 + \frac{\log_2 \binom{2m}{m}}{\log_2 \binom{3m}{m}} + o(1) = 1 + \frac{2m + o(m)}{3H(1/3)m + o(m)}$$

where $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$ is the binary entropy function. For large m, we obtain

$$x \sim 1 + \frac{2}{3H(1/3)} = 1 + \frac{2}{3\log_2 3 - 2} \approx 1.726.$$

Now, we modify the previous construction to one where the number of isosceles right triangles is more, it is $n^{1.773\ldots}$, and their δ_β harmonic points are all distinct. In the upper bound, in Theorem 1.4, we proved that many similar triangles guarantee that they share any selected harmonic points.

The bases of the triangles are spanned between the two point sets, A and B, along the axes. The following sums of complex numbers define the elements of the sets. :

$$A = \left\{\sum_{k=1}^{2m} a_k 13^k : a_k \in \{1,2\}, |\{k : a_k = 1\}| = m\right\},$$

$$B = \left\{\sum_{k=1}^{2m} b_k 13^k : b_k \in \{i,2i\}, |\{k : b_k = i\}| = m\right\}.$$

With this definition, $|A| = |B| = \binom{2m}{m}$.

Some pairs from A and B will form the bases of the isosceles right triangles. We need a parameter, ν, to select the pairs, which comes from a simple optimization problem. We set $\nu = 0.773$. In this construction, two points, $\alpha = \sum_{k=1}^{2m} a_k 13^k \in A$ and $\beta = \sum_{k=1}^{2m} b_k 13^k \in B$ form the base of a triangle if the number of $(a_k, b_k) = (1, 2i)$ pairs equals to the number of $(a_k, b_k) = (2, i)$ pairs which is $(1-\nu)m$.

Each $\alpha \in A$ forms a base with $\binom{m}{(1-\nu)m}^2$ distinct $\beta \in B$.

As before, the third point of the triangle is determined by α and β as

$$\gamma = \frac{\alpha+\beta}{2} + i\frac{\beta-\alpha}{2} = \sum_{k=1}^{2m} \frac{a_k(1-i) + b_k(1+i)}{2} 13^k = \sum_{k=1}^{2m} g_k 13^k.$$

The possible values of the g_k's are 0 and $\pm(1-i)/2$, as indicated in the tableau below.

$a_k \backslash b_k$	i	$2i$
1	0	$\frac{i-1}{2}$
2	$\frac{1-i}{2}$	0

We have

$$|C| = \binom{2m}{2\nu m}\binom{(1-\nu)2m}{(1-\nu)m} \approx \binom{2m}{m} = |A| = |B|.$$

We noted earlier that any $\alpha \in A$ is the vertex of $\binom{m}{(1-\nu)m}^2$ triangles, so the total number of triangles is

$$\binom{2m}{m}\binom{m}{(1-\nu)m}^2 \approx \binom{2m}{m}^{1.773}.$$

Now we have to check that the harmonic point, $\delta_\beta = \frac{4-2i}{5}\alpha + \frac{1+2i}{5}\beta$, is unique to the triangle.

As before, we check the results digit-wise of $\delta' = 5\delta_\beta = (4-2i)\alpha + (1+2i)\beta$ for the possible α, β combinations. The results are summarized in the tableau below.

$a_k \setminus b_k$	i	$2i$
1	$2-i$	0
2	$3(2-i)$	$2(2-i)$

As before, by the digits of δ' we can uniquely recover the triangle's base. Let us check the other harmonic point, δ_α.

We set $\delta'' = 5\delta_\alpha = (1-2i)\alpha + (4+2i)\beta$ for the remaining harmonic point. The possible digit-wise entries of δ'' are

$a_k \setminus b_k$	i	$2i$
1	$-1+2i$	$3(-1+2i)$
2	0	$2(2i-1)$

There is one pair of triangles where $\delta_\alpha = \delta_\beta$, in which case both points are in the origin. Let us remove one of these triangles. The remaining triangles have distinct harmonic points with the possible exception of their δ_γ points.

5 Proof of Theorem 1.5

Recall that we are to prove the following: For every $c > 0$ and $\varepsilon > 0$, the following holds for large enough n. Let T be a triangle and S be a set of n points in the plane such that

S contains cn^2 triangles similar to T. Then, there is a quadrangle Q and a set of at most $(2c/\varepsilon^6)n$ points, denoted by U, such that U contains at least $(c-\varepsilon)n^2$ quadrangles similar to Q and $S \subset U$.

We are going to use the counting methods from the proof of Theorem 1.4. The proof follows a simple algorithm. Select one of the harmonic points of T. These four points will give Q. For any triangle similar to T, we will only consider the harmonic point, which gives a quadrangle similar to Q. The set of the selected harmonic points is denoted by H. Set $\delta = \varepsilon^6$.

1. Let us begin with $U = S$.
2. Select a point $h \in H$ which is not in U and the harmonic point of at least δn triangles.
3. If there is no such point, then stop.
4. Set $U = U \cup h$ and repeat from step 2.

The proof of Theorem 1.4 shows that there exist m^6/n^{11} triangles sharing the same harmonic point, hence for any $c > 0$, as n is sufficiently large and the number of triangles on n points similar to T is $m = cn^2 > 60n^{11/6}$, there are at least $m^6/n^{11} = c^6 n$ triangles sharing a given harmonic point. In every step we selected at least δn triangles and no triangle was selected multiple times. Hence the number of iterations in the algorithm is at most $cn^2/(\delta n)$ and $|U| \leq |S| + cn^2/(\delta n) = (c/\delta + 1)n \leq (2c/\varepsilon^6)n$. Also, this selection of δ guarantees that all but at most εn^2 triangles have their harmonic points in U. For if there are more than $\varepsilon n^2 > 60n^{11/6}$ triangles with harmonic point not in U, then by Theorem 1.4, there are at least $(\varepsilon n)^6/n^{11} = \varepsilon^6 n = \delta n$ triangles that share a common harmonic point and the algorithm would not have terminated.

References

[1] Ábrego, S., Fernández-Merchant, S., and Elekes, Gy. *Structural results for planar sets with many similar subsets*, Combinatorica 24(4) (2004), 541–554.

[2] Behrend, F. A, *On sets of integers which contain no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. U.S.A. 32 (1946), 331–332.

[3] Conlon, D., Fox, J., Sudakov, B., Zhao, Y. *The regularity method for graphs with few 4-cycles* J. London Math. Soc. (2) 104 (2021) 2376—2401.

[4] Collier-Cartaino, C., Graber, N., and Jiang, T. *Linear Turán Numbers of Linear Cycles and Cycle-Complete Ramsey Numbers*, Combinatorics, Probability and Computing (2018) 27, 358–386.

[5] Elekes, Gy. and Erdős, P. *Similar configurations and pseudo-grids*. Böröczky, K., et al. (Eds.), *Intuitive Geometry*, Colloq. Math. Soc. János Bolyai, vol. 63, North-Holland (1991).

[6] Ergemlidze, B., Győri, E., and Methuku, A. *Asymptotics for Turán numbers of cycles in 3-uniform linear hypergraphs*, Journal of Combinatorial Theory, Series A (163), 2019, 163–181.

[7] Gallatly, W. *The Harmonic Quadrilateral*. §124 in *The Modern Geometry of the Triangle*, 2nd ed., London: Hodgson, pp. 90 and 92, 1913.

[8] Girao, António; Hurley, Eoin; Illingworth, Freddie; Michel, Lukas *J. Lond. Math. Soc.* (2) 110 (2024), no. 5, Paper No. e70015, 26 pp.

[9] Gishboliner, L., Shapira, A., and Wigderson, Y. *An efficient asymmetric removal lemma and its limitations*, Forum Math. Sigma, to appear.

[10] Jiang, T. and Yepremyan, L. *Supersaturation of even linear cycles in linear hypergraphs*, Combinatorics, Probability and Computing (2020), 29, 698–721.

[11] Johnson, Roger A. *Advanced Euclidean Geometry*, Dover (2007) [1929], pp. 100.

[12] Katz, N. and Tao, T. *Bounds on arithmetic projections, and applications to the Kakeya conjecture*, Math. Res. Lett. 6 (1999), no. 5-6, 625–630.

[13] Kostochka, A., Mubayi, D., and Verstraëte, J. *Hypergraph Ramsey numbers: triangles versus cliques*, J. Combin. Th. Ser. A, 120 (2013), 1491-1507.

[14] Laczkovich, M. and Ruzsa, I.Z. *The Number of Homothetic Subsets*. In: Graham, R.L., and Nešetřil, J. (Eds.), *The Mathematics of Paul Erdős II*. Algorithms and Combinatorics, vol. 14, Springer, Berlin, Heidelberg (1997).

[15] Lemm, M. *New Counterexamples for Sums-Differences*. Proceedings of the American Mathematical Society 143, no. 9 (2015): 3863–3868.

[16] Ruzsa, I. Z. *Sums of finite sets*, In: Number Theory: New York Seminar; Springer-Verlag (1996), D.V. Chudnovsky, G.V. Chudnovsky and M.B. Nathanson editors.

[17] Ruzsa, I. Z. and Szemerédi, E, *Triple systems with no six points carrying three triangles*, in Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, Colloq. Math. Soc. János Bolyai, vol. 18, North-Holland, Amsterdam-New York, 1978, 939–945.

[18] Solymosi, J, *Berge Pentagons in 3-uniform Hypergraphs* Combinatorics Colloquium, Rényi Institute, Budapest. (November 8, 2018)

[19] Mubayi, D. and Solymosi, J., *Planar point sets with many similar triangles*, Discrete Geometry Days 2024 Conference, Budapest. (July 5, 2024) http://dgd.math.bme.hu/program.html