The Hypergraph Turán Densities of Tight Cycles Minus an Edge

Bernard Lidický* Connor Mattes† Florian Pfender‡

December 4, 2024

Abstract

A tight ℓ-cycle minus an edge C_ℓ^- is the 3-graph on the vertex set $[\ell]$, where any three consecutive vertices in the string $123\ldots\ell 1$ form an edge. We show that for every $\ell \geq 5$, ℓ not divisible by 3, the extremal number is

$$\mathrm{ex}\left(C_\ell^-, n\right) = \tfrac{1}{24}n^3 + O(n\ln n) = \left(\tfrac{1}{4} + o(1)\right)\binom{n}{3}.$$

We determine the extremal graph up to $O(n)$ edge edits.

1 Introduction

Maybe the oldest and most fundamental question in extremal hypergraph theory is to maximize the number of edges in an r-uniform hypergraph (aka r-graph) on n vertices, which does not contain a given hypergraph H as a subgraph. The resulting function is called the extremal number of H, and denoted by $\mathrm{ex}(H, n)$.

We can normalize this quantity, and an easy averaging argument shows that the limit

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

exists for every r-graph H. This limit is called the Turán density of H. For 2-graphs, a celebrated theorem by Erdős, Stone, and Simonovits relates the Turán density to the chromatic number $\chi(H)$.

*Department of Mathematics, Iowa State University, Ames, IA. E-mail: `lidicky@iastate.edu`. Research of this author is supported in part by NSF grant DMS-2152490 and Scott Hanna professorship.

†Sandia National Laboratories. Livermore, California, USA. E-mail: `clmatte@sandia.gov`. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. Research on this project was partially supported by NSF grant DMS-2152498.

‡Department of Mathematical and Statistical Sciences, University of Colorado Denver. E-mail: `Florian.Pfender@ucdenver.edu`. Research is partially supported by NSF grant DMS-2152498.

Theorem 1 ([9, 10])**.** *For every* 2*-graph* H*,* $\pi(H) = \frac{\chi(H)-2}{\chi(H)-1}$.

For $r \geq 3$, this question is wide open. Erdős [8] showed that $\pi(H) = 0$ if and only if H is an r-partite r-graph. Other than this, the answer is known for only very few r-graphs or classes of r-graphs.

Restricting ourselves to 3-graphs from now on, finding the Turán density for $K_4 = K_4^3$, the complete 3-graph on 4 vertices, may be the most famous open problem in extremal hypergraph theory. Even if we delete one edge from K_4 to create K_4^-, the smallest not 3-partite 3-graph, $\pi(K_4^-)$ is unknown. The conjectured values for these two densities are $\pi(K_4) = \frac{5}{9}$ [24] and $\pi(K_4^-) = \frac{2}{7}$ [20].

Three small 3-graphs we know the Turán density for are the Fano plane F on 7 vertices, and the books $B_{3,2}$ and $B_{3,3}$ on 5 vertices and edge sets $\{123, 124, 345\}$ and $\{123, 124, 125, 345\}$, respectively. We have $\pi(F) = \frac{3}{4}$ [6], $\pi(B_{3,2}) = \frac{2}{9}$ [11], and $\pi(B_{3,3}) = \frac{4}{9}$ [12]. For each of these three 3-graphs, the structure of the extremal construction is rather simple. The vertices are partitioned into two or three parts of appropriate sizes, and edges are completely determined by the parts the vertices belong to.

For known results and fundamental techniques, a survey by Keevash [16] is a good resource, for open problems see also a collection by Mubayi, Pikhurko and Sudakov [21]. For recent updates see Balogh, Clemen and Lidický [1].

Since these surveys, two recent results are exciting developments. For this, let us consider the tight ℓ-cycle C_ℓ, a 3-graph on vertex set $[\ell]$, and the edges encoded in the string $123\ldots\ell12$, where the edges are exactly all 3-sets which appear consecutively in the string. Similarly, we can define the tight ℓ-cycle minus an edge C_ℓ^- by the string $123\ldots\ell1$. Observe that $K_4 = C_4$ and $K_4^- = C_4^-$, so these two classes contain the two most notorious 3-graphs for which the Turán density is unknown. If ℓ is a multiple of 3, then C_ℓ and C_ℓ^- are 3-partite, so $\pi(C_\ell) = \pi(C_\ell^-) = 0$. For other large enough ℓ, we have the following two Theorems by Kamčev, Letzter, and Pokrovskiy [15], and by Balogh and Luo [4].

Theorem 2 ([15])**.** *There is a constant* L *such that* $\pi(C_\ell) = 2\sqrt{3} - 3$*, for every* $\ell \geq L$ *not divisible by* 3*.*

Theorem 3 ([4])**.** *There is a constant* L *such that* $\pi\left(C_\ell^-\right) = \frac{1}{4}$*, for every* $\ell \geq L$ *not divisible by* 3*.*

Both the previous statements are conjectured to be true already for $L = 5$ (they are false for $\ell = 4$), but the proofs only work for very large L. In this paper, we add to the very meager set of known results for small 3-graphs, and first determine the Turán density of C_5^-. The limit object is the same as in Theorem 3, and has a much more intricate structure, described later, than the limit objects for the results on other small 3-graphs mentioned above.

Theorem 4.

$$\pi\left(C_5^-\right) = \frac{1}{4}.$$

Using blow-up arguments from [4], this implies the Turán density for C_ℓ^- for all $\ell \geq 5$, completing the statement in Theorem 3.

Theorem 5. *For every $\ell \geq 5$,*

$$\pi\left(C_\ell^-\right) = \begin{cases} 0, & \text{if } 3 \mid \ell, \\ \frac{1}{4}, & \text{if } 3 \nmid \ell. \end{cases}$$

After proving the main lemma of this paper in Section 3, we prove Theorems 4 and 5 in Section 4.

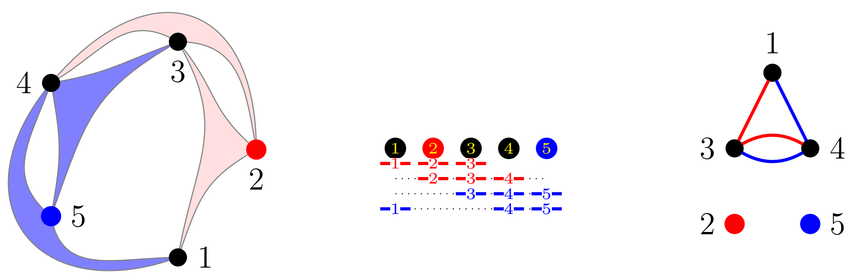


Figure 1: C_5^- drawn with hyperedges, as vertex-edge incidence, and as links of the two independent vertices

Furthermore, we show a stronger result, as we asymptotically determine the extremal 3-graphs for all C_ℓ^- with $\ell \geq 5$ and not divisible by 3. Let the 3-graph H_n be the iterated balanced blow-up of an edge on n vertices. In other words, for

$$i = \left\lfloor \tfrac{n}{3} \right\rfloor, \; j = \left\lfloor \tfrac{n+1}{3} \right\rfloor, \; k = \left\lfloor \tfrac{n+2}{3} \right\rfloor,$$

inductively start with the three 3-graphs H_i, H_j, and H_k, and add all edges spanning all three graphs. The start of the induction are the edge free graphs H_1 and H_2. Note that H_n has edge density $\frac{1}{4} + o(1)$. For any $\ell \geq 4$, C_ℓ^- is tightly connected, meaning that any two edges are connected by a tight walk, a sequence of at least three vertices where every three consecutive vertices span an edge. Thus, any copy of C_ℓ^- in H_n would have to be contained in the same tight block of H_n. The tight blocks of H_n are 3-partite, which shows that H_n contains no copy of C_ℓ^- unless ℓ is divisible by 3.

Let $F_{\ell,n}$ be an extremal graph for C_ℓ^-, a C_ℓ^--free 3-graph on n vertices with a maximum number of edges. We will show that, for sufficiently large n, $F_{\ell,n}$ is a balanced blow-up of an edge.

Theorem 6. *For $\ell \geq 5$ not divisible by 3, let $F_{\ell,n}$ be a C_ℓ^--free 3-graph on n vertices with a maximum number of edges. For n sufficiently large, there is a partition $V(F_{\ell,n}) = X_1 \cup X_2 \cup X_3$ such that $v_1v_2v_3 \in E(F_{\ell,n})$ for all $v_1 \in X_1, v_2 \in X_2, v_3 \in X_3$, and $|X_1|, |X_2|, |X_3| \in \left\{\left\lfloor \frac{n}{3} \right\rfloor, \left\lceil \frac{n}{3} \right\rceil\right\}$.*

Note that Theorem 6 implies that there are no edges in $F_{\ell,n}$ of the type $u_iv_iv_j$ with $u_i, v_i \in X_i$, $v_j \in X_j$, $i \neq j$, as this together with appropriately chosen $\ell - 3$ other vertices would induce a graph containing a copy of a C_ℓ^-.

This implies in a strong sense that $F_{\ell,n}$ and H_n converge to the same limit object, a hypergraphon. We discuss a bit of this perspective in Section 4, before we prove Theorem 6 in Section 5.

We can use the theorem inductively to show that for some large enough M, for every k and $3^{k+1}M > n \geq 3^k M$, $F_{\ell,n}$ agrees on the first k levels of H_n from the outside in. This shows that $F_{\ell,n}$ and H_n are isomorphic up to changing $3^k\binom{3M}{3} = O(n)$ edges. In particular, since we can determine the number of edges of H_n up to a small error, we have

Theorem 7.

$$\left| ex\left(C_\ell^-, n\right) - \frac{n^3}{24} \right| < \frac{1}{6} n \log_3 n + O(n).$$

In general, we can not expect that $F_{\ell,n} = H_n$, since modifying H_s for small s yields a few extra edges without creating a C_ℓ^-. For example for $\ell = 5$, H_4 is the 4-vertex graph with 2 edges, while $F_{5,4}$ is complete. Replacing every vertex of a single edge with a complete graph on 4 vertices creates a C_5^--free graph on 12 vertices with $64 + 12 = 76$ edges, while H_{12} contains only $64 + 6 = 70$ edges. Taking this further, we can construct a graph on $n = 4 \cdot 3^k$ vertices with $2 \cdot 3^k = \frac{n}{2}$ extra edges compared to H_n. For $5 \leq s \leq 8$, $F_{5,s}$ is the graph containing exactly all $\binom{s-1}{2}$ edges containing a single vertex, verified by computer by an exhaustive search. For $n \geq 9$, the best construction we know is as follows. Use the same recursion as in the construction of H_n, with the one difference that we use the known extremal graphs $F_{5,3}, F_{5,4}, \ldots, F_{5,8}$ instead of recursing all the way down to H_1 and H_2. We wonder if this is in fact the unique extremal construction for all n.

For $\ell \geq 7$, we observe that the extremal graph $F_{\ell,s}$ is complete for every $s < \ell$, but we have not explored cases with $s \geq \ell$. The resulting construction on $(\ell - 1)3^k$ vertices yields an extra $(\frac{\ell^2}{8} + O(\ell))n$ edges compared to $H_{(\ell-1)3^k}$, with the $O(\ell)$ in terms of growing ℓ.

2 Flag Algebra Methods

Our method relies on the theory of flag algebras developed by Razborov [22]. Flag algebras can be used as a general tool to attack problems from extremal combinatorics. In this paper, we use it to bound densities of 3-graphs, 2-graphs, and even 0-graphs (i.e. graphs with no edges), often with colors added to vertices and/or edges.

The plain flag algebra method by Razborov computationally automates many techniques from the theory of flag algebras. For a more thorough explanation of the method avoiding the technicalities of [22], one may look at Section 4 of [3]. A typical application of this method provides asymptotic bounds on densities of substructures subject to constraints on the densities of other substructures. For example, in Lemma 10, we give an upper bound on the maximum number of edges in a C_5^--free graph. To get accurate bounds, true inequalities and equalities involving the densities of substructures are combined using semidefinite programming. Often, one can create more accurate bounds by considering equalities and inequalities on larger substructures. However, the size of these semi-definite programs quickly reach the bounds of the capabilities of even the largest computers. Certificates for the truth of the statements proven by the plain flag algebra method appear in the form of sums of squares. Due to their sheer size required in this paper, they are impractical to print, and checking them involves a computer. We provide our programs and certificates at `http://lidicky.name/pub/c5-/`. Lemmas and Claims using flag algebras and other computer assistance are indicated by ⚙.

In some applications, the bounds from flag algebra are asymptotically sharp. Obtaining an exact description of the extremal structure from sharp bounds usually consists of first bounding the densities of some small substructures by $o(1)$. Often, these substructures can be read off from the flag algebra computation. Using a removal lemma applicable to the structure, one can get rid of all these substructures without changing the asymptotic densities of the extremal object. From this, one can extract a lot of information about the structure. Finally, stability arguments can sometimes be used to extract the precise extremal object. While the other known exact results on Turán densities of three small 3-graphs mentioned in Section 1 were not proven using flag algebras, the results for the two books can be re-proved with this near automatic approach. The corresponding computation for the Fano plane is too large, even with the very large computers we have access to.

For the result in this paper, bounds we get from the plain flag algebra method are not sharp. In our experience, this is typical when the extremal construction is more complicated than can be easily captured by the density of small substructures, as it is the case for most iterated constructions. We have encountered similar issues before on 2-graphs, see [2, 3, 17]. In these previous applications of the plain flag algebra method, we were able to determine the extremal graphs using stability arguments paired with bounds on the densities of a few subgraphs appearing with high density in the top level of the iterated extremal construction.

That same approach is not sufficient in this application. Looking merely at the top level does not give us good enough bounds. We instead also include densities of subgraphs in the next level. We then extensively explore the local structure of the link of a vertex, the 2-graph induced by the edges incident to one vertex, with the flag algebra method. The only time we stretched the limits of our available computation power is for Lemma 12, which took more than a month on a large cluster. All other computations are moderate in size and could be run on a personal computer in reasonable time.

3 The Main Lemma

For any two hypergraphs H on k vertices and G on $n > k$ vertices, let

$$p(H,G) = \frac{|\{X \subset V(G) : G[X] \simeq H\}|}{\binom{n}{k}}$$

denote the induced density of H in G. With a slight abuse of notation to improve readability, we will often just write H instead of $p(H,G)$ in equations of subgraph densities where the large graph G is clear from context. A first instance of this notation is the following proposition.

Proposition 8. *In any C_5^--free 3-graph G on n vertices, $K_4^- + K_4 \leq \frac{4}{n-3}$.*

Proof. Note that if we duplicate a vertex of degree 3 in a K_4^- or K_4, we create a graph containing C_5^-. In other words, any set of three vertices can have at most one vertex forming an edge with all three vertex pairs. Thus, we can have at most $\binom{n}{3} = \frac{4}{n-3}\binom{n}{4}$ copies of K_4^- and K_4 in G. $\square$

We can eliminate all copies of K_4^- via a removal lemma (see Section 4) without asymptotically changing the extremal graph. This new simplified graph is much easier to explore with the flag algebra method, as there are drastically[1] fewer 3-graphs which are $\{C_5^-, K_4^-\}$-free compared to all C_5^--free 3-graphs. For this reason, we will study $\{C_5^-, K_4^-\}$-free graphs for most of this paper before coming back to C_5^--free graphs. Similarly to $F_{5,n}$, define G_n to be extremal $\{C_5^-, K_4^-\}$-free graphs on n vertices.

The following main lemma falls slightly short of Theorem 6, but the proof of it contains most of the work.

Lemma 9 (Main Lemma). *For n sufficiently large, there is a partition $V(G_n) = X_1 \cup X_2 \cup X_3$ such that $v_1v_2v_3 \in E(G_n)$ for all $v_1 \in X_1, v_2 \in X_2, v_3 \in X_3$, and $|X_1|, |X_2|, |X_3| \in \{\lfloor \frac{n}{3} \rfloor, \lceil \frac{n}{3} \rceil\}$.*

In this case, we are not aware of any value of n where H_n and G_n differ, and it would not surprise us if $G_n = H_n$ for all n. In fact, Lemma 9 together with a blow-up argument can be used to show that $G_{3^k} = H_{3^k}$.

Proof of Lemma 9. The proof of this lemma is rather technical, and contains a number of lemmas and claims with their own separate proofs.

We apply the plain flag algebra method to get a first upper bound. It turns out that we never have to use this bound, but we provide it here to show what a simple application of the method achieves.

Lemma 10 (⚙). *G_n has edge density less than $0.2502175 + o(1)$.*

Proof. A straight forward application of the plain flag algebra method using a large computation gives an asymptotic upper bound of

$$\frac{14012179180700882058867945296422074007373856}{56000000000000000000000000000000000000000000} \approx 0.2502174853696586.$$

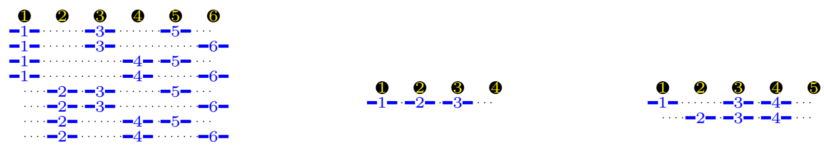
$\square$

The graph H_n has edge density $0.25 + o(1)$. The bound in Lemma 10 is a bit larger than that, and it will sometimes be useful to work with a graph with edge density $0.25 + o(1)$. For this, we randomly construct such a graph $\bar{G}_n$ from G_n by uniformly at random deleting an appropriate number of edges.

Note that $\bar{G}_n$ is $\{K_4^-, C_5^-\}$-free, since the class of $\{K_4^-, C_5^-\}$-free hypergraphs is closed under edge deletion, vertex deletion and vertex duplication. A standard symmetrization argument implies that G_n is asymptotically regular. Due to standard concentration results, $\bar{G}_n$ is also asymptotically regular with high probability.

Claim 11. *$\Delta(G_n) - \delta(G_n) < n$, and with probability $1 - o(1)$, $\Delta(\bar{G}_n) - \delta(\bar{G}_n) = o(n \log n)$.*

[1]On 8 vertices there are 161,023 $\{C_5^-, K_4^-\}$-free and 1,528,500 C_5^--free 3-graphs. Flag algebras on a supercomputer can run on up to $\approx 200,000$ graphs.



H_6 $\qquad$ H_3^T $\qquad$ H_4^T

Figure 2: Vertex-edge incidences of the hypergraphs H_6, H_3^T, and H_4^T.

Let $X_1, X_2, X_3 \subset V(G_n)$ be disjoint sets, and let $T = V(G) \setminus (X_1 \cup X_2 \cup X_3)$. Call a triple of vertices funky if it is an edge and contains two vertices in X_i and one vertex in X_j for $i \neq j$, or if it is a non-edge and contains a vertex from each of X_1, X_2, X_3. Let F denote the set of funky triples, with F_1 denoting the funky edges and F_2 denoting the funky non-edges.

If we can show that $F = T = \emptyset$, then we know that G_n is a blow-up of an edge. We now use a flag algebra argument to show that both F and T have relatively small cardinality on our way to show that G_n must indeed be the blow-up of an edge. Let $\binom{n}{3} f = |F|$, $\binom{n}{3} f_i = |F_i|$, $tn = |T|$, and $x_i n = |X_i|$. We will assume without loss of generality that $x_1 \geq x_2 \geq x_3$.

Recall that H_6 is the balanced blow-up of an edge on 6 vertices. Let H_3^T denote an edge with an additional isolated vertex. Finally, Let H_4^T denote the balanced blow-up of an edge on 4 vertices, plus an additional isolated vertex, see Figure 2.

Lemma 12 (⚙). *For both G_n and $\bar{G}_n$, we have*

$$6 \cdot \frac{1}{15} H_6 + 0.196 \cdot \frac{1}{4} H_3^T + 0.366 \cdot \frac{1}{10} H_4^T \geq 0.0552798 + o(1). \tag{1}$$

Proof. We express in flag algebra that G_n and $\bar{G}_n$ are $\{K_4^-, C_5^-\}$-free, with edge density at least $\frac{1}{4}$, and almost regular in the sense of Claim 11. Now we run a plain flag algebra computation on 8 vertices (this is the single very large computation used in our proof) to obtain

$$\begin{aligned} \frac{6}{15} H_6 + \frac{0.196}{4} H_3^T + \frac{0.366}{10} H_4^T &\geq \frac{232175476344384142902456040424939595692259 1}{42000000000000000000000000000000000000000} \\ &> 0.05527987532. \end{aligned}$$

$\square$

Lemma 13. *There exists some partition of the vertices of G_n into X_1, X_2, X_3, T such that*

$$6x_1x_2x_3 - f_2 + 0.196t + 0.366 \cdot t(1-t) \geq 0.221119 + o(1). \tag{2}$$

Proof. For a slightly improved bound, we use $\bar{G}_n$ to partition the vertices. Note that F_2 in G_n is contained in the corresponding set in $\bar{G}_n$. For any graph G and any edge $e \in E(G_n)$, let $N_e(G)$ denote the number of subgraphs in $\bar{G}_n$ isomorphic to G containing e. Then,

$$\begin{aligned} &\sum_e 6N_e(H_6) + 0.196n^2 N_e(H_3^T) + 0.366n N_e(H_4^T) \\ &\qquad = 48\binom{n}{6} H_6 + 0.196n^2 \binom{n}{4} H_3^T + 2 \cdot 0.366n \binom{n}{5} H_4^T. \end{aligned}$$

Thus, by an averaging argument and by Lemma 12, there exists some edge e, such that

$$
\begin{aligned}
&\frac{1}{n^3}(6N_e(H_6) + 0.196n^2 N_e(H_3^T) + 0.366 \cdot nN_e(H_4^T)) \\
&\qquad \geq \frac{48\binom{n}{6}H_6 + 0.196n^2\binom{n}{4}H_3^T + 2 \cdot 0.366n\binom{n}{5}H_4^T}{0.25n^3\binom{n}{3}} \\
&\qquad = \frac{6 \cdot \frac{1}{15}H_6 + 0.196 \cdot \frac{1}{4}H_3^T + 0.366 \cdot \frac{1}{10}H_4^T}{0.25} + o(1) \\
&\qquad \geq \frac{0.0552798}{0.25} + o(1) > 0.221119 + o(1).
\end{aligned} \tag{3}
$$

Now fix such an edge e, say consisting of vertices v_1, v_2, v_3. Then we may create a partition as follows: If for a vertex u, the set $\{u\} \cup (\{v_1, v_2, v_3\} \setminus \{v_i\})$ induces an edge, place u in X_i. Otherwise, place u in T. Notice that u appears only in one of the four sets since G_n (and thus $\bar{G}_n$) is K_4^--free. Effectively what we are doing is placing u in X_i if it looks like v_i to the other vertices in e. Now, this partition determines upper bounds for all $N_e(G)$. In particular, one can see that

$$
\begin{aligned}
N_e(H_6) &= (X_1 - 1)(X_2 - 1)(X_3 - 1) - f_2\binom{n}{3} \\
N_e(H_3^T) &\leq tn \\
N_e(H_4^T) &\leq t(1-t)n^2
\end{aligned} \tag{4}
$$

The result then follows by combining (3) and (4).

□

Now that we see how Lemma 12 translates into a polynomial based on part sizes, we want to give a brief explanation why we chose the coefficients as we did. As we want to prove that G_n is very close to H_n, we are guided by the structure of H_n. In H_n, most edges span the three sets of the top level blow-up. Choosing such an edge, it is in close to $\frac{n^3}{27}$ copies of H_6. If we can somehow guarantee that there exists an edge in G_n which is in that many copies of H_6, we can recover these three sets via our definition of the X_i, with $x_1 = x_2 = x_3 = \frac{1}{3}$. But since we can not pin point such an edge, and our argument instead relies on an average over all edges, other edges in H_n reduce this average. To improve our bounds, we consider subgraphs of H_n which are frequent supergraphs of other edges in H_n.

The second most frequent type of edges in H_n lies completely in one set of the top-level, but spans the three sets of the next level. Using such an edge as our root yields $x_1 = x_2 = x_3 = 1/9$ with $t = \frac{2}{3}$. These edges are contained in many graphs isomorphic to H_3^T and H_4^T, while the first type of edges is contained in neither. That is why we choose to include these graphs to boost our bound.

Our goal is to construct an equation from these three graphs so that only the top level edges of H_n can beat the value we get from flag algebra, guaranteeing that an average edge is at the top level, and allows us to partition the vertices well into three parts. At the same time, we want that bound to be as high as possible, to give us the best possible partition. In other words, we want (2) to be feasible only close to $x_1 = x_2 = x_3 = \frac{1}{3}$. However, the larger our polynomial is at $x_1 = x_2 = x_3 = \frac{1}{9}$, the more this increases the right hand side of (1), and with this improves our bounds. Thus, we pick our coefficients for the $p(G)$ so that (2) is not feasible at $x_1 = x_2 = x_3 = 1/9$, but is very close to being feasible to get the maximum contribution from those second level edges. In practice, this does take a bit of guess work. You start with the densities you have in H_n, and choose your coefficients with a bit of room to account for the bounds from flag algebra not being perfect. You may have to run the plain flag algebra method multiple times to test different coefficients in order to find the best possible combination. In our case, our first try with a very large computation was close enough to being successful so that it was sufficient to run a second much smaller computation to slightly adjust the coefficients. One could further optimize the coefficients or add an additional $p(G)$ term to get more control over the resulting polynomial, or to take into account edges on the next level of the construction. But we will see that for our needs, the bounds we get work sufficiently well.

Claim 14. *For large enough n, for all partitions satisfying Lemma 13,*

$$\begin{aligned} x_3 &\geq 0.306 \\ x_1 &\leq 0.361 \\ t &\leq 0.0109 \\ f_2 &\leq 0.001104 \\ f_2 &\leq \tfrac{2}{9}(1-t)^3 + 0.196t + 0.366(1-t)t - 0.221119 \end{aligned} \tag{5}$$

Proof. Note that for any fixed t, the left hand side of (2) is maximized if $x_1 = x_2 = x_3 = \frac{1-t}{3}$. Thus, f_2 is bounded by (5). This polynomial attains its maximum at $t = 0$, implying that $f_2 \leq 0.001104$ and $|F_2| < 0.000184n^3$. Further, the polynomial is negative for $0.0109 < t \leq 1$, implying the bound on t.

A little bit of calculus shows that decreasing t to increase x_1 increases the left side of (2), so it suffices to set $f_2 = t = 0$ to bound x_1 and x_3. To bound x_3, we may further assume $x_1 = x_2$, leading to a one-variable expression. The same is true for x_1 using $x_2 = x_3$. Evaluating these expressions gives the claimed bounds. $\square$

The previous claim shows that G_n must be fairly close to the conjectured construction at the top level. Next, we want to establish improved bounds using Claim 11. In $\bar{G}_n$, every vertex has degree $\frac{1}{8}n^2 + o(n^2)$.

Let us first look at the normalized degree sum inside X_1 (in $\bar{G}_n$). For ease of notation, we neglect "little o" terms in most of the remainder. We have

$$\Sigma_1 := \frac{1}{n} \sum_{x \in X_1} \frac{d(x)}{n^2} = \frac{1}{8} x_1. \tag{6}$$

We will further partition $\Sigma_1 = \Sigma_I + \Sigma_R + \Sigma_F + \Sigma_T$, where Σ_I is the contribution from edges completely in X_1, Σ_R is the contribution from edges spanning all of X_1, X_2 and X_3, Σ_F is the contribution from edges in F_2, and Σ_T is the contribution from edges with at least one vertex in T.

We have $\Sigma_I \leq \frac{1}{8}x_1^3$ — note that the random process from G_n to $\bar{G}_n$ also affected the degrees inside X_1. We have exactly $\Sigma_R = x_1 x_2 x_3 - \frac{1}{6} f_2$.

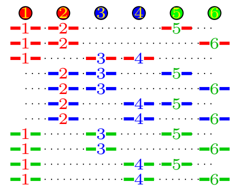


Figure 3: 3-graphs on 6 vertices having 4 normal edges (blue), 3 extra funky edges (red) where the first two are with coefficient two, and 4 funky non-edges (green).

Next, let us look at all edges of the types $u_1v_1v_2$, $u_1v_1v_3$, $v_1u_2v_2$, $v_1u_3v_3$ to bound Σ_F. The first two types each contribute 2 to Σ_F, the last two each contribute 1. To account for these edges, we will look at all subgraphs spanned by 6 vertices $\{u_1, v_1, u_2, v_2, u_3, v_3\}$. Note that if $u_1v_1v_2 \in E$, then at least one of $u_1v_2v_3$ and $v_1v_2v_3$ is in F_2 to avoid a K_4^-.

Claim 15 (⚙). *For all $\{K_4^-, C_5^-\}$-free 6-vertex 3-graphs on vertices $\{u_1, v_1, u_2, v_2, u_3, v_3\}$ the ratio of the count of edges $\sum_{t\in\{u,v\}} 2u_1v_1t_2 + 2u_1v_1t_3 + t_1v_2u_2 + t_1v_3u_3$ over the number of non-edges intersecting all three pairs u_1v_1, u_2v_2, u_3v_3.*

Proof. We generate all possible such $\{K_4^-, C_5^-\}$-free 6-vertex 3-graphs with a computer and find the ratio. Over all such 3-graphs, the largest ratio is $\frac{5}{4}$. Figure 3 depicts the graph achieving the ratio $\frac{5}{4}$. □

Our goal is to get a bound on Σ_F in terms of non-edges in F_2. For this, let us set up an auxiliary bipartite weighted multigraph M between the vertices $v_1 \in X_1$ and non-edges $w_1w_2w_3 \in F_2$. For every set $S = \{u_1, v_1, u_2, v_2, u_3, v_3\}$, we count the s_1 edges in $F_1 \cap S$ containing v_1, and the s_2 non-edges $w_1w_2w_3 \in F_2 \cap S$. Then, we add an edge of weight $\frac{s_1}{s_2}$ in M between v_1 and $w_1w_2w_3$. Notice that this is a multigraph since different S contain $\{v_1, w_1, w_2, w_3\}$.

For every set S containing v_1, the total edge weight of v_1 in M is s_1. So, counting over all sets S containing v_1, the total weight of the edges in M is at least $\frac{1}{2}x_2x_3^2n^3\Sigma_F$, as each F_1-edge is in at least $\frac{1}{2}x_2x_3^2n^3$ different sets S. On the other hand, every non-edge $w_1w_2w_3 \in F_2$ is in $x_1x_2x_3n^3$ sets S, so it has edge weight in M at most $\frac{5}{4}x_1x_2x_3n^3$ by the discussion before about 6-vertex graphs, so the total edge weight is bounded above by $\frac{5}{4}x_1x_2x_3n^3f_2\binom{n}{3}$. After normalizing the degree sum as before,

$$\Sigma_F \le \frac{1}{n^3}\frac{5}{2}f_2\binom{n}{3}\frac{x_1x_2x_3}{x_2x_3^2} = \frac{5x_1}{12x_3}f_2.$$

Finally, let us account for edges with vertices in T. Every vertex in T has degree $\frac{1}{8}n^2$, so there are at most $\frac{1}{8}tn^3$ edges containing a vertex in T and a vertex in X_1. Edges with one vertex in T and two vertices in X_1 count twice towards Σ_T. Since the link of every vertex is triangle free, there are at most $\frac{1}{4}tx_1^2n^3$ such edges.

So, in total, $\Sigma_T \le \frac{1}{8}t(1 + 2x_1^2)$. Taking everything together, we get

$$\frac{1}{8}x_1 = \Sigma_1 \le \frac{1}{8}x_1^3 + x_1x_2x_3 - \frac{1}{6}f_2 + \frac{5x_1}{12x_3}f_2 + \frac{1}{8}t(1 + 2x_1^2), \tag{7}$$

and consequently after multiplying by $24x_3$,

$$3x_1x_3 \le 3x_1^3x_3 + 24x_1x_2x_3^2 - 4x_3f_2 + 10x_1f_2 + 3x_3t + 6x_1^2x_3t. \tag{8}$$

This inequality is weakest for maximized f_2, so we may assume that (2) holds with equality, and we have

$$\begin{aligned} 0 \le &- 2.21119x_1 + 1.96x_1t + 0.884476x_3 + 2.216x_3t - 3x_1x_3 + 6x_1^2x_3t + 3x_1^3x_3 \\ &+ 60x_1^2x_2x_3 + 3.66x_1t - 3.66x_1t^2 - 1.464x_3t + 1.464x_3t^2. \end{aligned} \tag{9}$$

Combining (2) and (9), we find new bounds on the x_i via optimization. This time, the optimization is less straight forward. This is a polynomial optimization problem which could certainly be treated with dedicated methods we are not experts in. Instead, we formulate a flag algebra model and solve that. We encode the problem as an edgeless graph with four vertex colors, where the size of the parts are x_1, x_2, x_3, t. We can express all of our inequalities in this model, where monomials of degree d correspond to graphs on d vertices with suitable weights. We ask flag algebra to optimize the parameters however we like, giving us the following certified bounds with a moderately sized computation.

Claim 16 (⚙).

$$\begin{aligned} x_3 &\ge 0.31723 \\ x_2x_3 &\ge 0.10613 \\ x_1 &\le 0.33865 \\ x_1x_2 &\le 0.11378 \end{aligned}$$

Next, we will bound a term which will show up later in our argument.

Claim 17 (⚙). *In $\bar{G}_n$, $f_1 - f_2 + \frac{3}{4}(t^2 - t^3) \le 0.0003042$.*

Proof. We are setting up a flag algebra model on 3-graphs, in which vertices are partitioned into the 4 parts. This additional partition restricts further the size of subgraphs we can compute with, but this smaller computation gives bounds sufficient for our purpose. Again, the 3-graph has edge density 0.25 and is $\{C_5^-, K_4^-\}$-free, and all bounds in Claims 14 and 16 apply to this partition. We now find the claimed upper bound for $f_1 - f_2 + \frac{3}{4}(t^2 - t^3)$ with the plain flag algebra method. □

From now on, we assume the partition $X_1 \cup X_2 \cup X_3 \cup T$ is chosen so that the left side of (2) is maximized. All bounds we have established hold for this partition.

Next, we will look at the link $L(v)$ of a single vertex v, the 2-graph formed by the remaining pairs of vertices in edges containing v. The number of edges in $L(v)$ is $d(v) = \frac{1}{4}\binom{n}{2} + o(n^2)$. Further, this graph is triangle free as otherwise v would be the degree 3 vertex of a K_4^- in G_n.

Let $\{A, B, C\} = \{X_1, X_2, X_3\}$ such that the number of edges in $L(v)$ between A and B is maximized. This implies that $v \in C \cup T$. We will use normalized sizes $a = |A|/n$, $b = |B|/n$ and $c = |C|/n$.

We want to find an upper bound for the number of edges in $L(v)$ inside C, $d_C(v) = d_C\binom{n}{2}$. We know that the edge density in C (in $\bar{G} \cup v$) is at most $e(C) \le \frac{1}{4}$. This is still true if we blow up v into a copy of $\bar{G}_{\mu n}$ for any $\mu \geq 0$, as the random sparcification has, with high probability, affected $d_C(v)$ proportionally. Therefore, for all $\mu \geq 0$,

$$e(C)\binom{cn}{3} + \mu n d_C(v) + \frac{1}{4}\binom{\mu n}{3} \leq \frac{1}{4}\binom{(c+\mu)n}{3},$$

so after dividing by $\binom{n}{3}$,

$$3\mu d_C \leq \frac{1}{4}(c^3 + 3c^2\mu + 3c\mu^2) - e(C)c^3.$$

On the other hand, every vertex in T is in $\frac{1}{4}\binom{n}{2} + o(n^2)$ edges, so the total number of edges containing at least one vertex of T is at most $\frac{3}{4}t\binom{n}{3}$.

For any $x, y, z > 0$ with $x + y + z = 1$ we have $\frac{1}{4}(x^3 + y^3 + z^3) + 6xyz \leq \frac{1}{4}$, with the maximum attained only for $x = y = z = \frac{1}{3}$. As $a + b + c = 1 - t$, this scales to $\frac{1}{4}(a^3 + b^3 + c^3) + 6abc \leq \frac{1}{4}(1 - t)^3$. Now, using also that $e(A), e(B) \leq \frac{1}{4}$,

$$\begin{aligned}\tfrac{1}{4} &\leq e(C)c^3 + e(A)a^3 + e(B)b^3 + 6abc + f_1 - f_2 + \tfrac{3}{4}t \\ &\leq \tfrac{1}{4}(a^3 + b^3 + c^3) + 6abc - \tfrac{1}{4}c^3 + e(C)c^3 + f_1 - f_2 + \tfrac{3}{4}t \\ &\leq \tfrac{1}{4}(1 - t)^3 - \tfrac{1}{4}c^3 + e(C)c^3 + f_1 - f_2 + \tfrac{3}{4}t.\end{aligned}$$

Therefore,

$$e(C)c^3 \geq \frac{1}{4}c^3 - \frac{3}{4}t^2 + \frac{3}{4}t^3 - f_1 + f_2.$$

Together, this gives

$$3\mu d_C \leq \frac{3}{4}(c^2\mu + c\mu^2) + \frac{3}{4}(t^2 - t^3) + f_1 - f_2,$$

and

$$d_C \leq \frac{1}{4}(c^2 + c\mu) + \frac{1}{3\mu}(\tfrac{3}{4}(t^2 - t^3) + f_1 - f_2).$$

This bound is minimized if

$$\mu = 2\sqrt{\frac{\frac{3}{4}(t^2 - t^3) + f_1 - f_2}{3c}},$$

so

$$d_C - \frac{1}{4}c^2 \leq \sqrt{\frac{c(\frac{3}{4}(t^2 - t^3) + f_1 - f_2)}{3}},$$

and finally, for $d_C \geq \frac{1}{4}c^2$,

$$3(d_C - \tfrac{1}{4}c^2)^2 \leq c(\tfrac{3}{4}(t^2 - t^3) + f_1 - f_2). \tag{10}$$

We need to establish two more bounds. For easier readability we will denote the set a vertex belongs to in its index.

Assume that u_Av_A and v_Av_B are edges in $L(v)$. Then, for any vertex v_C, $u_Avv_Av_Bv_Cu_A$ is a C_5^- unless at least one of $u_Av_Bv_C$ and $v_Av_Bv_C$ is missing, and thus in F_2. We can count such sets of 4 vertices containing a P_3 in $L(v)$ and a non-edge in F_2 in two ways. We can start with the P_3 and add any vertex in C like we did above, giving a count of $p_3^{AAB}\binom{n}{3}cn$ (giving implicitely the definitions of the density p_3^{AAB} of such paths). Or, we can start with a non-edge in F_2, and add a vertex in A, giving a count of $f_2\binom{n}{3}an$. The second choice is an overcount, as not every vertex in A must yield a P_3, and as there can be more than one non-edge between these 4 vertices. We use $\textcircled{A}$, $\textcircled{B}$, $\textcircled{C}$, $\textcircled{T}$ to denote the single vertex graphs with the vertex in the respective set when writing the following inequalities in flag algebra notation. We get

$$\textcircled{A}\!-\!\textcircled{A}\!-\!\textcircled{B} * \textcircled{C} < f_2 * \textcircled{A}\,, \tag{11}$$
$$\textcircled{A}\!-\!\textcircled{B}\!-\!\textcircled{B} * \textcircled{C} < f_2 * \textcircled{B}\,, \tag{12}$$

where the second line follows from a symmetric argument.

For the second bound, assume that u_Av_C and v_Av_B are edges in $L(v)$. Then again, $vu_Av_Cv_Bv_Av$ is a C_5^- unless at least one of $u_Av_Bv_C$ and $v_Av_Bv_C$ is missing, and thus in F_2. Denoting by $e^{AB}\binom{n}{2}$ the number of edges from A to B (and at the same time $\textcircled{A}\!-\!\textcircled{B}$ is an edge from A to B in flag algebra notation), similar counts as above give us $e^{AB}\binom{n}{2}e^{AC}\binom{n}{2}$ and $f_2\binom{n}{3}|A|$ for the set of 4 vertices, which yields

$$3\,\textcircled{A}\!-\!\textcircled{B} * \textcircled{A}\!-\!\textcircled{C} < 2\,\textcircled{A} * f_2, \tag{13}$$
$$3\,\textcircled{A}\!-\!\textcircled{B} * \textcircled{B}\!-\!\textcircled{C} < 2\,\textcircled{B} * f_2. \tag{14}$$

Now we put all these bounds together, and formulate the following model for $L(v)$. As f_1 and f_2 do not exist inside the model, we replace them by their bounds $f_2 < (6\,\textcircled{A} * \textcircled{B} * \textcircled{C} + 0.196\,\textcircled{T} + 0.366(1 - \textcircled{T})\,\textcircled{T} - 0.221119)$ and $f_1 - f_2 + \frac{3}{4}(\textcircled{T}^2 - \textcircled{T}^3) < 0.0003064$.

(a) Four sets of vertices: A, B, C, T

(b) $\textcircled{A} + \textcircled{B} + \textcircled{C} + \textcircled{T} = \bigcirc = 1$

(c) $0.31748 < \textcircled{A}, \textcircled{B}, \textcircled{C} < 0.33833$

(d) $0.10622 < \textcircled{A} * \textcircled{B}, \textcircled{A} * \textcircled{C}, \textcircled{B} * \textcircled{C} < 0.11366$

(e) $0 \leq \textcircled{T} < 0.0109$

(f) triangle-free

(g) $\textcircled{A}\!-\!\textcircled{B} \geq \textcircled{A}\!-\!\textcircled{C}$, $\textcircled{A}\!-\!\textcircled{B} \geq \textcircled{B}\!-\!\textcircled{C}$

(h) $\bigcirc\!-\!\bigcirc \geq \frac{1}{4}$

(i) $6\,\textcircled{A} * \textcircled{B} * \textcircled{C} + 0.196\,\textcircled{T} + 0.366(1 - \textcircled{T})\,\textcircled{T} - 0.221119 \geq 0$

(j) $\textcircled{A}\!-\!\textcircled{A}\!-\!\textcircled{B} * \textcircled{C} \leq \textcircled{A} * (6\,\textcircled{A} * \textcircled{B} * \textcircled{C} + 0.196\,\textcircled{T} + 0.366(1 - \textcircled{T})\,\textcircled{T} - 0.221119)$
$\textcircled{A}\!-\!\textcircled{B}\!-\!\textcircled{B} * \textcircled{C} \leq \textcircled{B} * (6\,\textcircled{A} * \textcircled{B} * \textcircled{C} + 0.196\,\textcircled{T} + 0.366(1 - \textcircled{T})\,\textcircled{T} - 0.221119)$

(k) $3\,\textcircled{A}\!-\!\textcircled{B} * \textcircled{A}\!-\!\textcircled{C} \leq 2\,\textcircled{A} * (6\,\textcircled{A} * \textcircled{B} * \textcircled{C} + 0.196\,\textcircled{T} + 0.366(1 - \textcircled{T})\,\textcircled{T} - 0.221119)$
$3\,\textcircled{A}\!-\!\textcircled{B} * \textcircled{B}\!-\!\textcircled{C} \leq 2\,\textcircled{B} * (6\,\textcircled{A} * \textcircled{B} * \textcircled{C} + 0.196\,\textcircled{T} + 0.366(1 - \textcircled{T})\,\textcircled{T} - 0.221119)$

(l) $3(\textcircled{C}\!-\!\textcircled{C} - \frac{1}{4}\textcircled{C}^2)^2 \leq \textcircled{C} * 0.0003064$

Claim 18 (⚙). *In $L(v)$, $\textcircled{A}\!-\!\textcircled{B} > 0.1942$.*

Proof. In our model for $L(v)$, we find a lower bound for $\textcircled{A}\!-\!\textcircled{B}$ with a moderate flag algebra computation. □

Claim 19. *$T = \emptyset$.*

Proof. Assume that $v \in T$. As a member of T, it contributes (normalized) $0.196 + 0.366(1 - t) \leq 0.562$ to the left side of (2). If we move it to C, it contributes $6 \cdot \frac{1}{2}e^{AB} > 0.58$. As the partition maximizes the left side, this is a contradiction. As v was chosen arbitrarily, the claim follows. □

At this point, we could recompute and improve a few of our bounds using the fact that $T = \emptyset$. It turns out that this is not needed, so we skip this step.

Claim 20 (⚙). *Every vertex is in at most $0.01452\binom{n}{2}$ non-edges in F_2.*

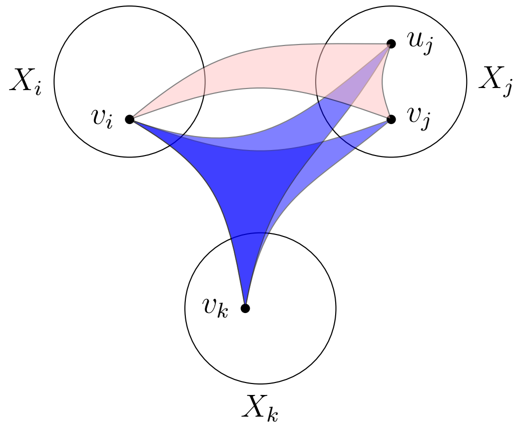
Proof. For an arbitrary vertex v, we use the model for $L(v)$ with $T = \emptyset$. Remember that $v \in C \cup T$, so in fact $v \in C$ here. Every non-edge in $L(v)$ between A and B corresponds to a non-edge in F_2 containing v. Using a moderate flag algebra computation, we find an upper bound for $2\,\textcircled{A} * \textcircled{B} - \textcircled{A}\!-\!\textcircled{B}$, the density of such non-edges in $L(v)$. □

While some of our arguments relied on being in $\bar{G}_n$, note that all resulting bounds apply to G_n as well. In particular the only difference is that F_1 can only be larger in G_n, and F_2 can only be smaller. The only stated bound that may not apply to G_n is the bound in Claim 17. But this bound was only used as an intermediate step to find further bounds for the partition, which then apply to G_n as well as the partition is the same. The remaining argument starting here is in G_n.

Claim 21. *In the partition $X_1 \cup X_2 \cup X_3$ of G_n, $F_1 = F_2 = \emptyset$.*

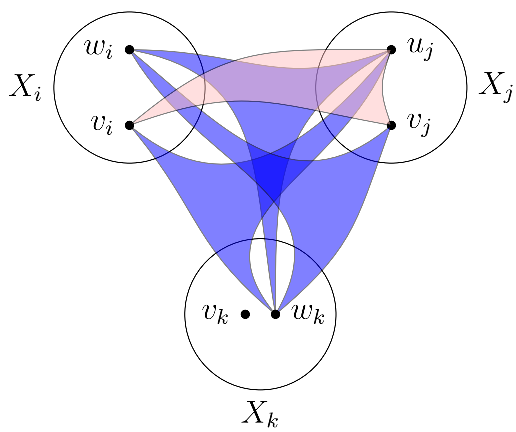
Proof. For ease of notation, indices of vertices indicate the index of the set they belong to. If $|F_2| > |F_1|$, we can delete all edges in F_1 and add all edges in F_2 to create a graph with more edges, contradicting the extremality of G_n. So we may assume that $f_1 \geq f_2$ and $F_1 \neq \emptyset$.

For every v_k, either $v_iv_jv_k$ or $v_iu_jv_k$ is a funky non-edge to avoid a K_4^-, so each funky edge $v_iu_jv_j$ intersects at least $x_kn \geq x_3n$ funky non-edges in two vertices. Thus, there are at least $f_1\binom{n}{3}x_3n$ such pairs of intersecting edges and non-edges. Counting the pairs the other way and averaging, we see that there exists a funky non-edge intersecting at least $x_3n > 0.31824n$ funky edges in two vertices.



Taking symmetry of i, j, k into account, we may assume that $v_iv_jv_k$ is such a funky non-edge with more than $\frac{0.31824n}{6} = 0.05304n$ vertices u_j such that $v_iu_jv_j$ is a funky edge. Let $Z_j = \{u_j : v_iu_jv_j \in E\}$, $z_jn = |Z_j|$.

For any w_i, w_k, and $u_j \in Z_j$, notice that $v_jv_iu_jw_kw_iv_j$ is a C_5^- unless one of $v_iu_jw_k$, $w_iu_jw_k$, $w_iv_jw_k$ is a funky non-edge.



In total, by Claim 16, we have $x_iz_jx_kn^3 \geq 0.10648z_jn^3$ choices for w_i, u_j, w_k.

By Claim 20, we can not have many non-edges $w_iv_jw_k$, as they are all incident to v_j. To be precise, we can have fewer than $0.0146\binom{n}{2} < 0.0073n^2$ such non-edges. Thus, out of the more than $0.10648z_jn^3$ choices of w_i, u_j, w_k, at least $(0.10648 - 0.0073)z_jn^3 > 0.0052n^3$ choices must yield a non-edge of the types $v_iu_jw_k$ or $w_iu_jw_k$.

We can have at most $f_2\binom{n}{3} < 0.00018n^3$ non-edges $w_iu_jw_k$. This still leaves more than $0.005n^3$ choices which force non-edges $v_iu_jw_k$. But since again by Claim 20, v_i is in at most $0.0073n^2$ such non-edges, this accounts for at most $0.0073x_in^3 < 0.0025n^3$ choices, a contradiction proving the claim. $\square$

Claim 22. *The edge density d in G_n is $\frac{1}{4} + o(1)$. Further, $x_i = \frac{1}{3} + o(1)$ for all i.*

Proof. As $F_2 = \emptyset$, the subgraphs spanned by the X_i are extremal themselves, and thus have $(d + o(1))\binom{|X_i|}{3}$ edges. Therefore,

$$d = 6x_1x_2x_3 + (d + o(1))(x_1^3 + x_2^3 + x_3^3).$$

For fixed $d < 0.2502$ and $0.3 < x_3 \leq x_2 \leq x_1 < 0.35$, the right side is maximized by $x_1 = x_2 = x_3 = \frac{1}{3}$. Solving for d gives $d = \frac{1}{4} + o(1)$. $\square$

To finish the proof of Lemma 9, the only part remaining to show is that the partition is balanced.

Claim 23. *For large enough n, $|X_1| - |X_3| \leq 1$.*

Proof. For the sake of contradiction, assume that $|X_1| - 1 \geq |X_3| + 1$. Remove one vertex of minimum degree from X_1 and duplicate a vertex of maximum degree from X_3. This way we are deleting at most $(\frac{1}{4} + o(1))\binom{|X_1|}{2}$ edges from X_1, and add at least $(\frac{1}{4} + o(1))\binom{|X_3|}{2}$ edges to X_3, for a net loss of at most

$$\begin{aligned}(\tfrac{1}{4} + o(1))\left(\binom{|X_1|}{2} - \binom{|X_3|}{2}\right) &= (\tfrac{1}{8} + o(1))(|X_1| + |X_3|)(|X_1| - |X_3|) \\ &= (\tfrac{1}{12} + o(1))(|X_1| - |X_3|)n\end{aligned}$$

edges inside these two sets. On the other hand, we are gaining

$$\begin{aligned}(|X_1| - 1)|X_2|(|X_3| + 1) - |X_1||X_2||X_3| &= (|X_1| - |X_3| - 1)|X_2| \\ &= (\tfrac{1}{3} + o(1)(|X_1| - |X_3| - 1)n\end{aligned}$$

edges spanning all three sets. As $|X_1| - |X_3| \geq 2$, we have $2(|X_1| - |X_3| - 1) \geq |X_1| - |X_3|$, so we gained a total of at least $(\frac{1}{12} + o(1))n$ edges, contradicting the maximality of G_n. $\square$

This establishes Lemma 9. $\square$

4 Hypergraph limits and Turán density

The Hypergraph Removal Lemma can be deduced from the celebrated Hypergraph Regularity Lemma (independently proved by Rödl, Nagle, Skokan, Schacht and Kohayakawa [23], and Gowers [14]). While merely stating the regularity lemma is beyond the scope of this paper, the removal lemma is easy to state and use.

Theorem 24 (Hypergraph Removal Lemma ([14], [23]))**.** *Let $r \geq 2$, and let H be an r-uniform hypergraph on k vertices. Let G be an r-uniform hypergraph on n vertices containing $o(n^k)$ (not necessarily induced) copies of H. Then one may remove $o(n^r)$ edges from G so that the resulting hypergraph is H-free.*

With this, Theorem 4 easily follows from Lemma 9.

Proof of Theorem 4. By Proposition 8, we can use the Hypergraph Removal Lemma to remove $o(n^3)$ edges from F_n to arrive at a $\{C_5^-, K_4^-\}$-free graph F_n'. Writing $e(H)$ for the edge density of a hypergraph H, we have

$$e(G_n) + o(1) = e(H_n) \leq e(F_n) = e(F_n') + o(1) \leq e(G_n) + o(1) = \frac{1}{4} + o(1),$$

implying $e(F_n) = \frac{1}{4} + o(1)$ and thus Theorem 4. $\square$

For a 3-graph H, we write $H[t] = H(E_t, E_t, \dots, E_t)$ for the balanced blow-up of H by $|H|$ copies of the empty 3-graph on t vertices. Proposition 8 is a special instance of a more general Theorem (see section 2 in [16]).

Theorem 25. *For every 3-graph H and positive integer t, if a 3-graph G on n vertices has no subgraph isomorphic to $H[t]$, then $p(H, G) = o(1)$.*

For a vertex v, let us denote copies of v by $v^1, v^2, \dots$. Proposition 8 then follows from Theorem 25 since C_5^- is in $K_4^-[2]$ as can be seen by the sequence 13243^11. In [4], it is shown that

Proposition 26. *Every C_ℓ^- with $\ell \geq 7$ and ℓ not divisible by 3 is contained in a blow-up of C_5^-.*

Proof. We paraphrase the argument, and first note that 13243^154^11 gives a C_7^- in $C_5^-[2]$. Next, observe that C_{k+3}^- is contained in $C_k^-[2]$ by considering the string $1231^12^13^14\dots k1$. The Claim now follows by induction starting in C_5^- and C_7^-. □

Proof of Theorem 5. To prove the theorem, note that Theorem 25 and Proposition 26 imply that for every $\ell \geq 7$ not divisible by 3, every C_ℓ^--free graph has $C_5^- = o(1)$. By the Hypergraph Removal Lemma, we can delete $o(n^3)$ edges to destroy all copies of C_5^-, and then Theorem 4 implies that $ex\left(C_\ell^-, n\right) \leq (\frac{1}{4} + o(1)\binom{n}{3}$. As H_n does not contain any copies of C_ℓ^-, the theorem follows. □

"Graphons, short for graph functions, are the limiting objects for sequences of large, finite graphs with respect to the so-called cut metric. They were introduced and developed by Borgs, Chayes, Lovász, Sós, Szegedy, and Vesztergombi in [5] and [19]." A short explainer by Glasscock, from which we cited the previous sentence, can be found in [13], and an extensive treatment by Lovász appeared in [18].

Later in [7], Elek and Szegedy developed the corresponding theory of hypergraphons, which are limit objects of hypergraphs. Similarly and closely related to the Hypergraph Regularity Lemma, stating the definition of a hypergraphon is beyond the scope of this paper, an introduction is contained in [18]. For our purposes, it is enough to consider the following theorem.

Theorem 27 ([7])**.** *Let (P_n) and (Q_n) be two sequences of r-uniform hypergraphs on n vertices. Then (P_n) and (Q_n) converge to the same unique hypergraphon, if and only if for every r-uniform hypergraph H, $\lim_{n\to\infty} p(H, P_n) = \lim_{n\to\infty} p(H, Q_n)$.*

It follows easily from an inductive application of Theorem 9 that (H_n) and (G_n) converge to the same hypergraphon W. In fact, we can label their vertices in a way that their edit distance is $o(n^3)$, a property stronger than their convergence to the same limit.

We say that a hypergraphon R is finitely forcible if there exists a finite set $\{P_1, P_2, \dots, P_k\}$ of r-uniform hypergraphs and real numbers $p_1, \dots, p_k \in [0, 1]$ so that for all sequences (Q_n) of r-uniform hypergraphs on n vertices, $\forall i : \lim p(P_i, Q_n) = p_i$ implies that R is the limit of (Q_n). In our next statement, we show that W is finitely forcible with only two subhypergraph densities.

Theorem 28. *Let (Q_n) be a sequence of 3-uniform hypergraphs on n vertices with $\lim p(C_5^-, Q_n) = 0$ and $\lim p(K_3, Q_n) = \frac{1}{4}$. Then (Q_n) converges to W, the limit of (H_n).*

Proof. Using the Hypergraph Removal Lemma, Theorem 25, Proposition 26, and the Hypergraph Removal Lemma a second time, we can delete $o(n^3)$ edges from Q_n to construct $\{C_5^-, K_4^-\}$-free hypergraphs Q'_n with $p(H, Q_n) = p(H, Q'_n) + o(1)$ for every 3-uniform H. Theorem 4 implies that $\Delta(Q'_n) = (\frac{1}{4} + o(1))\binom{n}{2}$. Let $\varepsilon = \varepsilon(n) > 0$ with $\lim_{n\to\infty} \varepsilon(n) = 0$ to be chosen later, and define

$$S_n = \left\{ v \in V(Q_n) : d(v) < (e(Q'_n) - \varepsilon)\binom{n}{2} \right\}$$

be the set of small degree vertices. If $\rho\binom{n}{2} = \Delta(Q'_n) - e(Q'_n)\binom{n}{2}$, then $|S_n|\varepsilon \leq n\rho < n\rho + 1$, so choosing $\varepsilon = \sqrt{\rho + \frac{1}{n}} = o(1)$ gives $|S_n| < \varepsilon n = o(n)$. Let $Q''_n = Q'_n - S_n$, then $\Delta(Q''_n) - \delta(Q''_n) = o(n^2)$.

With this and a bit of care handling the $o(1)$ terms, we can apply all steps of the proof of Theorem 9 until before Claim 21 to Q''_n instead of $\bar{G}_n$. Claim 21 is the first time we use extremality of G_n in the proof.

With essentially the same proof as in Claim 21, we can instead prove that $f_1 = o(1)$ and $f_2 = o(1)$. The same then still holds if we add S_n to X_1. This suffices to show that $x_i = \frac{1}{3} + o(1)$, and we have therefore, after adding back the $o(n^3)$ edges we removed in the very beginning, the following claim.

Claim 29. *The vertices of Q_n can be partitioned $V(Q_n) = X_1 \cup X_2 \cup X_3$ with $x_i = \frac{1}{3} + o(1)$ such that $f_1 = o(1)$ and $f_2 = o(1)$, and the induced hypergraphs on the X_i each have edge density $e(X_i) = \frac{1}{4} + o(1)$.*

Applying this claim inductively to the X_i proves the theorem. □

5 Proofs of Theorems 6 and 7

Theorem 6. *For $\ell \geq 5$ not divisible by 3, let $F_{\ell,n}$ be a C_ℓ^--free 3-graph on n vertices with a maximum number of edges. For n sufficiently large, there is a partition $V(F_{\ell,n}) = X_1 \cup X_2 \cup X_3$ such that $v_1v_2v_3 \in E(F_{\ell,n})$ for all $v_1 \in X_1, v_2 \in X_2, v_3 \in X_3$, and $|X_1|, |X_2|, |X_3| \in \{\lfloor \frac{n}{3} \rfloor, \lceil \frac{n}{3} \rceil\}$.*

Proof of Theorem 6. By Theorem 28, (F_n) converges to H_n. It remains to show that this convergence is even stronger, and $F_1 = F_2 = \emptyset$ if we more carefully partition S_n from the proof of Theorem 28 across the X_i. Once we prove this, Claim 23 and thus the full theorem follows.

We define F'_n, F''_n, S_n as in the proof of Theorem 28 and proceed in the proof until Claim 21. Add back the missing edges of F_n, and partition S_n across the X_i to maximize $6x_1x_2x_3 - f_2$. This only affected $o(n^3)$ edges, so we still have $f_1 = o(1)$ and $f_2 = o(1)$. If $f_2 > f_1$, we can increase the number of edges without creating a C_ℓ^- if we replace all edges in F_1 by all edges in F_2, a contradiction showing that $f_2 \leq f_1$.

In the following, we describe the argument for $\ell = 5$, but it is straight forward to adapt it to general ℓ.

Let us now consider the link of a vertex $L(v)$ again, for $\{A, B, C\} = \{X_1, X_2, X_3\}$. We get a new model with the new bounds we know for any vertex v, by symmetry we may assume $v \in C$. As flag algebra computes in the limit, we omit $o(1)$ terms.

(a) Three sets of vertices: A, B, C

(b) $\cancel{A + B + C = 1}$

(c) A, B, C $= \frac{1}{3}$

(d) $\cancel{0.10649 < A * B, A * C, B * C < 0.11354}$

(e) $\cancel{0 \leq T < 0.01022}$

(f) ~~triangle-free~~

(g) $A{-}B \geq A{-}C, B{-}C$

(h) $A{-}A + B{-}B + C{-}C + A{-}B + A{-}C + B{-}C = \bigcirc{-}\bigcirc = \frac{1}{4}$

(i) $\cancel{6 A * B * C + 0.196 T + 0.366(1 - T) T - 0.22118 \geq 0}$

(j) $A{-}A{-}B + A{-}A{-}B = 0$
$A{-}B{-}B + A{-}B{-}B = 0$

(k) $A{-}C = 0$
$B{-}C = 0$

(l) $C{-}C \leq \frac{1}{4} C^2 = \frac{1}{36}$
$A{-}A, B{-}B \leq \frac{1}{36}$

Let us give some explanations. We do not need (d) anymore. The link of v may contain triangles. Denoting the corresponding triangle densities by k_3^{AAB} and k_3^{ABB}, (j) follows from the same argument as before. Similarly, (k) follows together with (g). Note that neither of these arguments used that G_n is K_4^--free. Finally, (l) follows for e^{CC} with the same argument, but there are more straight forward ways to prove it. Symmetrical arguments apply to e^{AA} and e^{BB}.

From (j) we see that all but $o(n)$ vertices in A have either $o(n)$ neighbors in A or $o(n)$ neighbors in B. A symmetric statement is true for vertices in B. Denoting the number of vertices in A with more than $o(n)$ neighbors in A by $s_A n$, and the number of vertices in B with more than $o(n)$ neighbors in B by $s_B n$, we have $0 \leq s_A, s_B \leq \frac{1}{3}$, and

$$e^{AA} + e^{AB} + e^{BB} \leq s_A^2 + 2(\tfrac{1}{3} - s_A)(\tfrac{1}{3} - s_B) + s_B^2 + o(1).$$

This counts all the e^{AA}- and e^{BB}-edges in two cliques of sizes $s_A n$ and $s_B n$, and the e^{AB}-edges in a complete bipartite graph not overlapping the cliques. But then, (h) and (l) imply that $e^{AA} + e^{AB} + e^{BB} = \frac{2}{9} + o(1)$, which is only possible for $s_A = s_B = o(1)$ (noting that $e^{AA} = \frac{1}{9} + o(1)$ is ruled out by (j)), and $A{-}B = 2 A * B + o(1)$. We conclude that

Claim 30. *The number of non-edges in F_2 containing v is $o(n^2)$.*

The remainder of the proof is again similar to Lemma 9, but easier. Assume for the sake of contradiction that $F_1 \neq \emptyset$, and $u_A v_A v_B \in F_1$. Then for any choice of w_B, w_C, $u_A v_B v_A w_C w_B u_A$ is a C_5^- unless at least one of $u_A w_B w_C, v_A w_B w_C, v_A v_B w_C$ is missing. By Claim 30, $u_A w_B w_C$ and $v_A w_B w_C$ can miss only $o(n^2)$ times, so for the remaining $\frac{n^2}{9} + o(n^2)$ choices, $v_A v_B w_C \in F_2$ (and in fact $u_A v_B w_C \in F_2$ as well considering $v_A v_B u_A w_C w_B v_A$). In other words, $u_A v_A v_B$ intersects at least $\frac{2}{3}n + o(n)$ non-edges in F_2 in two vertices. Counting this structure for all edges in F_1, and noting that $f_1 \geq f_2$, this implies that there exists some non-edge $v_A v_B v_C \in F_2$ which intersects at least $\frac{2}{3}n + o(n)$ edges in F_1.

In particular, we may assume by symmetry that $v_A v_B$ is in at least $\frac{1}{9}n + o(n)$ edges of type $v_A v_B u_B$. For any pair w_A, w_C, and any u_B with $v_A v_B u_B \in F_1$, at least one of $w_A u_B w_C$, $v_A u_B w_C$, $w_A v_B w_C$ must be missing. But there are $o(n^3)$ non-edges of type $w_A u_B w_C$ in F_2. Further, by Claim 30, there are $o(n^2)$ non-edges of types $v_A u_B w_c$ and $w_A v_B w_c$, each accounting for at most $\frac{1}{3}n + o(n)$ choices of the at least $\frac{1}{81}n^3 + o(n^3)$ choices of w_A, u_B, w_C. This contradiction shows that $F_1 = F_2 = \emptyset$.

The only remaining Claim 23 from Lemma 9 follows with the same proof, establishing Theorem 6. □

Theorem 7.
$$\left| ex\left(C_\ell^-, n\right) - \frac{n^3}{24} \right| < \frac{1}{6} n \log_3 n + O(n).$$

Proof of Theorem 7. As described in Section 1, for some large enough M, for every k and $3^{k+1}M > n \geq 3^k M$, F_n agrees on the first k levels of H_n from the outside in. This shows that F_n and H_n are isomorphic up to changing $3^k \binom{M}{3} = O(n)$ edges.

Thus, it suffices to show that

$$\left| ||H_n|| - \frac{n^3}{24} \right| < \frac{1}{6} n \log_3 n + O(n).$$

For this, we use induction on n and consider three cases $n = 3k$, $n = 3k+1$, and $n = 3k+2$. Note that the start of the induction is trivial due to the $O(n)$ term.

Then we have for $n = 3k$,

$$\begin{aligned} \left| ||H_{3k}|| - \frac{(3k)^3}{24} \right| &= \left| k^3 + 3||H_k|| - \frac{(3k)^3}{24} \right| \\ &\leq \left| k^3 + 3\frac{k^3}{24} - \frac{(3k)^3}{24} \right| + \frac{3}{6} k \log_3 k + O(3k) \\ &< \frac{1}{6} n \log_3 n + O(n). \end{aligned}$$

Similarly for $n = 3k+1$,

$$\begin{aligned}
\left| ||H_{3k+1}|| - \frac{(3k+1)^3}{24} \right| &= \left| k^3 + k^2 + 2||H_k|| + ||H_{k+1}|| - \frac{(3k+1)^3}{24} \right| \\
&\leq \left| k^3 + k^2 + 2\frac{k^3}{24} + \frac{(k+1)^3}{24} - \frac{(3k+1)^3}{24} \right| \\
&\quad + \frac{1}{6}(3k+1)\log_3(k+1) + O(3k+1) \\
&< \frac{6k}{24} + \frac{1}{6}n(\log_3 n - \tfrac{1}{2}) + O(n) \\
&< \frac{1}{6}n\log_3 n + O(n),
\end{aligned}$$

and for $n = 3k+2$,

$$\begin{aligned}
\left| ||H_{3k+2}|| - \frac{(3k+2)^3}{24} \right| &= \left| k^3 + 2k^2 + k + ||H_k|| + 2||H_{k+1}|| - \frac{(3k+2)^3}{24} \right| \\
&\leq \left| k^3 + 2k^2 + k + \frac{k^3}{24} + 2\frac{(k+1)^3}{24} - \frac{(3k+2)^3}{24} \right| \\
&\quad + \frac{1}{6}(3k+2)\log_3(k+1) + O(3k+2) \\
&< \frac{6k+6}{24} + \frac{1}{6}n(\log_3 n - \tfrac{2}{3}) + O(n) \\
&< \frac{1}{6}n\log_3 n + O(n).
\end{aligned}$$

□

Acknowledgments

This work used the computing resources at the Center for Computational Mathematics, University of Colorado Denver, including the Alderaan cluster, supported by the National Science Foundation award OAC-2019089.

After publishing a preprint of this manuscript, we learned that Bodnár, León, Liu, and Pikhurko are also working on this problem and are close to a final independent manuscript.

References

[1] J. Balogh, F. C. Clemen, and B. Lidický. Hypergraph Turán problems in ℓ_2-norm. In Surveys in combinatorics 2022, volume 481 of London Math. Soc. Lecture Note Ser., pages 21–63. Cambridge Univ. Press, Cambridge, 2022.

[2] J. Balogh, P. Hu, B. Lidický, and F. Pfender. Maximum density of induced 5-cycle is achieved by an iterated blow-up of 5-cycle. European J. Combin., 52:47–58, 2016.

[3] J. Balogh, P. Hu, B. Lidický, F. Pfender, J. Volec, and M. Young. Rainbow triangles in three-colored graphs. J. Combin. Theory Ser. B, 126:83–113, 2017.

[4] J. Balogh and H. Luo. Turán density of long tight cycle minus one hyperedge. Combinatorica, 44(5):949–976, Apr. 2024.

[5] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. Adv. Math., 219(6):1801–1851, 2008.

[6] D. De Caen and Z. Füredi. The maximum size of 3-uniform hypergraphs not containing a Fano plane. J. Combin. Theory Ser. B, 78(2):274–276, 2000.

[7] G. Elek and B. Szegedy. A measure-theoretic approach to the theory of dense hypergraphs. Adv. Math., 231(3-4):1731–1772, 2012.

[8] P. Erdős. On extremal problems of graphs and generalized graphs. Israel J. Math., 2:183–190, 1964.

[9] P. Erdős and M. Simonovits. A limit theorem in graph theory. Studia Sci. Math. Hungar., 1:51–57, 1966.

[10] P. Erdős and A. H. Stone. On the structure of linear graphs. Bull. Amer. Math. Soc., 52:1087–1091, 1946.

[11] P. Frankl and Z. Füredi. A new generalization of the Erdos-Ko-Rado theorem. Combinatorica, 3(3-4):341–349, 1983.

[12] Z. Füredi, O. Pikhurko, and M. Simonovits. The Turán density of the hypergraph $\{abc, ade, bde, cde\}$. Electron. J. Combin., 10:Research Paper 18, 7, 2003.

[13] D. Glasscock. What is ... a graphon? Notices Amer. Math. Soc., 62(1):46–48, 2015.

[14] W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. Ann. of Math. (2), 166(3):897–946, 2007.

[15] N. Kamčev, S. Letzter, and A. Pokrovskiy. The Turán density of tight cycles in three-uniform hypergraphs. Int. Math. Res. Not. IMRN, (6):4804–4841, 2024.

[16] P. Keevash. Hypergraph Turán problems. Surveys in Combinatorics, 06 2011.

[17] B. Lidický, C. Mattes, and F. Pfender. C_5 is almost a fractalizer. J. Graph Theory, 104(1):220–244, 2023.

[18] L. Lovász. Large networks and graph limits, volume 60 of American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, RI, 2012.

[19] L. Lovász and B. Szegedy. Limits of dense graph sequences. J. Combin. Theory Ser. B, 96(6):933–957, 2006.

[20] D. Mubayi. On hypergraphs with every four points spanning at most two triples. Electron. J. Combin., 10:Note 10, 4, 2003.

[21] D. Mubayi, O. Pikhurko, and B. Sudakov. AimPL: Hypergraph Turán problem, 2011.

[22] A. A. Razborov. Flag algebras. J. Symbolic Logic, 72(4):1239–1282, 2007.

[23] V. Rödl, B. Nagle, J. Skokan, M. Schacht, and Y. Kohayakawa. The hypergraph regularity method and its applications. Proc. Natl. Acad. Sci. USA, 102(23):8109–8113, 2005.

[24] P. Turán. Eine Extremalaufgabe aus der Graphentheorie. Mat. Fiz. Lapok, 48:436–452, 1941.