Maximum density of induced 5-cycle is achieved by an iterated blow-up of 5-cycle

József Balogh* Ping Hu† Bernard Lidický‡ Florian Pfender§

April 30, 2015

Abstract

Let $C(n)$ denote the maximum number of induced copies of 5-cycles in graphs on n vertices. For n large enough, we show that $C(n) = a \cdot b \cdot c \cdot d \cdot e + C(a) + C(b) + C(c) + C(d) + C(e)$, where $a + b + c + d + e = n$ and a, b, c, d, e are as equal as possible.

Moreover, for n being a power of 5, we show that the unique graph on n vertices maximizing the number of induced 5-cycles is an iterated blow-up of a 5-cycle.

The proof uses flag algebra computations and stability methods.

1 Introduction

In 1975, Pippinger and Golumbic [20] conjectured that in graphs the maximum induced density of a k-cycle is $k!/(k^k - k)$ when $k \geq 5$. In this paper we solve their conjecture for $k = 5$. In addition, we also show that the extremal limit object is unique. The problem of maximizing the induced density of C_5 is also posted on `http://flagmatic.org` as one of the problems where the plain flag algebra method was applied but failed to provide an exact result. It was also mentioned by Razborov [25].

Problems of maximizing the number of induced copies of a fixed small graph H have attracted a lot of attention recently [8, 14, 29]. For a list of other results on this so called inducibility of small graphs of order up to 5, see the work of Even-Zohar and Linial [8].

*Department of Mathematics, University of Illinois, Urbana, IL 61801, USA and Bolyai Institute, University of Szeged, Szeged, Hungary E-mail: `jobal@math.uiuc.edu`. Research is partially supported by Simons Fellowship, NSF CAREER Grant DMS-0745185, Arnold O. Beckman Research Award (UIUC Campus Research Board 13039) and Marie Curie FP7-PEOPLE-2012-IIF 327763.

†Department of Mathematics, University of Illinois, Urbana, IL 61801, USA and University of Warwick, UK E-mail: `pinghu1@math.uiuc.edu`.

‡Department of Mathematics, Iowa State University, Ames, IA, E-mail: `lidicky@iastate.edu`.

§Department of Mathematical and Statistical Sciences, University of Colorado Denver, E-mail: `Florian.Pfender@ucdenver.edu`. Research is partially supported by a collaboration grant from the Simons Foundation.

Denote the $(k-1)$-times iterated blow-up of C_5 by $C_5^{k\times}$, see Figure 1. Let $\mathcal{G}_n$ be the set of all graphs on n vertices, and denote by $C(G)$ the number of induced copies of C_5 in a graph G. Define

$$C(n) = \max_{G\in\mathcal{G}_n} C(G).$$

We say a graph $G \in \mathcal{G}_n$ is *extremal* if $C(G) = C(n)$. Notice that, since C_5 is a self-complementary graph, G is extremal if and only if its complement is extremal. If n is a power of 5, we can exactly determine the unique extremal graph and thus $C(n)$.

Theorem 1. *For $k \geq 1$, the unique extremal graph in $\mathcal{G}_{5^k}$ is $C_5^{k\times}$.*

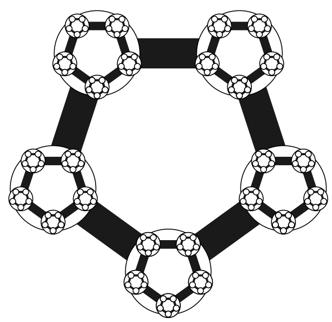


Figure 1: The graph $C_5^{k\times}$ maximizes the number of induced C_5s.

To prove Theorem 1, we first prove the following theorem. Note that this theorem is sufficient to determine the unique limit object (the graphon) maximizing the density of induced copies of C_5.

Theorem 2. *There exists n_0 such that for every $n \geq n_0$*

$$C(n) = a\cdot b\cdot c\cdot d\cdot e + C(a) + C(b) + C(c) + C(d) + C(e),$$

where $a+b+c+d+e = n$ and a,b,c,d,e are as equal as possible.

Moreover, if $G \in \mathcal{G}_n$ is an extremal graph, then $V(G)$ can be partitioned into five sets X_1, X_2, X_3, X_4, and X_5 of sizes a,b,c,d and e respectively, such that for $1 \leq i < j \leq 5$ and $x_i \in X_i$, $x_j \in X_j$, we have $x_ix_j \in E(G)$ if and only if $j - i \in \{1, 4\}$.

In the next section, we give a brief overview of our method, in Section 3 we prove Theorem 2, and in Section 4 we prove Theorem 1.

2 Method and Flag Algebras

Our method relies on the theory of flag algebras developed by Razborov [21]. Flag algebras can be used as a general tool to attack problems from extremal combinatorics. Flag algebras were used for a wide range of problems, for example the Caccetta-Häggkvist conjecture [15, 24], Turán-type problems in graphs [7, 11, 13, 19, 22, 26, 27], 3-graphs [9, 10] and hypercubes [1, 3], extremal problems in a colored environment [2, 4, 6], and also to problems in geometry [17] or extremal theory of permutations [5]. For more details on these applications, see a recent survey of Razborov [23].

A typical application of the so-called *plain flag algebra method* provides a bound on densities of substructures. To get a good bound, true inequalities and equalities involving the densities of substructures are combined with the help of semidefinite programming. This step is by now largely automated, there is even an open source application called Flagmatic [29], which gives easy to check certificates for the validity of this step. In some cases the bound is asymptotically sharp. Obtaining an exact result from the sharp bound usually consists of first bounding the densities of some small substructures by $o(1)$, which can be read off from the flag algebra computation. Forbidding these structures can yield a lot of information about the structures of the extremal structure. Finally, stability arguments are used to extract the precise extremal structure.

A similar approach can work in some cases where the bound on the desired density is not asymptotically sharp but merely very close to the extremal example. In this case, one may find bounds very close to 0 for a number of small substructures, and again these bounds may suffice for a stability argument.

Both of these 'lucky' cases happen most often when the extremal construction is 'clean', for example a simple blow-up of a small graph, replacing each vertex by a large independent set. Simple blow-ups of small graphs appear very often as extremal graphs, in fact there are large families of graphs whose extremal graphs for the inducibility are of this type, see Hatami, Hirst and Norin [12]. However, there are also many problems where the extremal construction is an iterated blow-up as shown by Pikhurko [18].

For our problem, the conjectured extremal graph has such an iterated structure, for which it is rare to obtain the precise density from plain flag algebra computations alone. One such rare example is the problem to determine the inducibility of small out-stars in oriented graphs [9] (note that the problem of inducibility of all out-stars was recently solved by Huang [16] using different techniques). Hladký, Král' and Norin announced that they found the inducibility of the oriented path of length 2, which also has an iterated extremal construction, via a flag algebra method. In [4] we determined the iterated extremal construction maximizing the number of rainbow triangles in 3-edge-colored complete graphs. Other than these three examples, we are not aware of any applications of flag algebras which completely determined an iterative structure.

For our question, a direct application of the plain method gives an upper bound on the limit value and shows that $\lim_{n\to\infty} C(n)/\binom{n}{5} < 0.03846157$, which is slightly more than the density of C_5 in the conjectured extremal construction, which is $\frac{1}{26} \approx 0.03846154$. This difference may appear very small, but the bounds on densities of subgraphs not appearing in the extremal structure are too weak to allow the standard methods to work.

Instead, we use flag algebras to find bounds on densities of other subgraphs, which appear with fairly high density in the extremal graph. This enables us to better control the slight lack of performance of the flag algebra bounds as these small errors have a weaker relative effect on larger densities. In the remainder of this section we will give a short description of this new method which provides a proof of Theorem 2, the most critical part of the proof of Theorem 1.

In studying the conjectured extremal example, the iterated blow-up $C_5^{k\times}$, one observes that the vast majority of induced C_5s contain a vertex in each of the five top-level sets. Starting with such a typical C_5 and picking an extra vertex, the adjacencies of this vertex to the C_5 determine conclusively to which top-level set the vertex belongs. Picking two extra vertices, the induced graph will be in one of two general classes: either the two additional vertices are in the same top-level set (we call this class $C31111$) or in different sets (we call this class $C22111$), see Figure 2.

With this observation in mind, we use flag algebra calculations to bound the densities of these two 7-vertex graph classes. We use the fact that we are studying the extremal example, and thus the induced density of C_5 can be bounded from below by $\frac{1}{26}$, the density in $C_5^{k\times}$ for $k \to \infty$. Using an averaging argument, we compute bounds on the number of graphs of these two classes a typical C_5 will lie in. We cannot expect very sharp bounds agreeing with the densities of a top-level C_5 in the iterated blow-up, as even in the iterated blow-up the lower level copies of C_5 affect the averaging. But this effect is small enough that these bounds enable us to go on.

Using a linear combination of the bounds on the numbers of graphs in $C31111$ and $C22111$ our now fixed typical base C_5 lies in, we can define five top-level sets and a left-over set, and bound the sizes of these sets. Further, we can even conclude that most edges and non-edges between the top-level sets follow the pattern of the base C_5, as otherwise the density of $C22111$ would be too small.

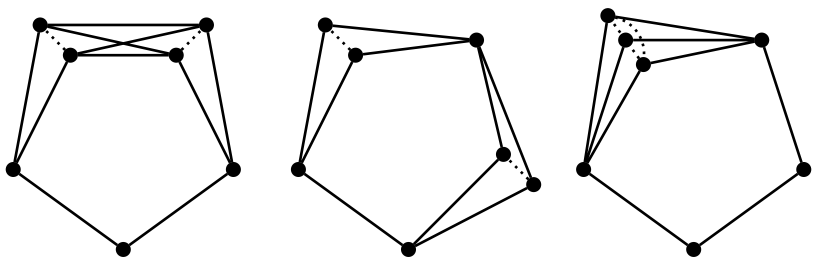
Using these bounds, we can use a fairly standard stability argument to show that in fact *all* edges and non-edges between the top-level sets follow the pattern of the base C_5 — if one of the pairs was out of pattern we could change it and increase the total number of C_5s.

In the next two steps, we show that the left-over set from above must be empty. First, we show that every vertex in the left-over set must look very different from the vertices in each of the top-level sets, again with a stability argument changing exactly one pair which is out of pattern. Then we show that this implies that this vertex lies in comparatively few C_5s to set up another standard stability argument: replacing this vertex by a copy of a vertex which is in at least an average number of C_5s would increase the total number of C_5s, a contradiction to the extremality. This last bound relies on the solution of a fairly well-behaved quadratic program, which can be relaxed to a program with only 5 variables. One could possibly solve this program with analytic means, but we doubt that this would give much added insight into the problem. Instead, we use a fairly simple brute-force discretization to approximate the solution in a rigorous way.

The final step of the proof of Theorem 2 is a convexity argument which shows that the top-level sets are balanced.

3 Proof of Theorem 2

In our proofs we consider densities of 7-vertex subgraphs. Guided by their prevalence in the conjectured extremal graph, the following two types of graphs will play an important role. We call a graph $C22111$ if it can be obtained from C_5 by duplicating two vertices. We call a graph $C31111$ if it can be obtained from C_5 by tripling one vertex. The edges between the original vertices and their copies are not specified, and there are two complementary types of $C22111$, depending on the adjacency of the two doubled vertices in C_5. Technically, $C22111$ and $C31111$ denote collections of several graphs. Examples of $C22111$ and $C31111$ are depicted in Figure 2. We slightly abuse notation by using $C22111$ and $C31111$ also to denote the densities of these graphs, i.e., the probability that randomly chosen 7 vertices induce the appropriate 7-vertex blow-up of C_5. Moreover, for a set of vertices Z we denote by $C22111(Z)$ and $C31111(Z)$ the densities of $C22111$ and $C31111$ containing Z, i.e., for a graph G on n vertices, $C22111(Z)$ ($C31111(Z)$) is the number of $C22111$ ($C31111$) containing Z divided by $\binom{n-|Z|}{7-|Z|}$.



$C22111$ $\qquad$ $C22111$ $\qquad$ $C31111$

Figure 2: Sketches of $C22111$ and $C31111$. The dotted edges may or may not be edges.

We start with the following statement.

Proposition 3. *There exists n_0 such that every extremal graph G on at least n_0 vertices satisfies:*

$$C_5 < 0.03846157;$$

$$4 \cdot C22111 - 11.94 \cdot C31111 \geq \frac{13498947603553891797877091863 91}{42000000000000000000000000000000} + o(1) > 0.003214. \tag{1}$$

Proof. This follows from a standard application of the plain flag algebra method. The first inequality was obtained by Flagmatic [29], which also provides the corresponding certificate. The computation by Flagmatic was done on 8 vertices. For the second inequality, we minimize the left side with the extra constraint that $C_5 \geq \frac{1}{26}$. We performed the computation on 7 vertices since the resulting bound was sufficient and rounding the solution is easier on 7 vertices than on 8. There are 6178 graphs to consider on 8 vertices while there are only 1044 on 7 vertices. It may be possible that we could use an upper bound on C_5 obtained on 7 vertices instead of 8 vertices. But since Flagmatic provides the result for 8 vertices, we used 8 vertices. For certificates, see `http://orion.math.iastate.edu/lidicky/pub/c5/`. $\square$

The expressions from Proposition 3 compare to the following limiting values in the iterated blow-up $C_5^{k\times}$, where $k \to \infty$:

$$C_5 = \frac{1}{26} \approx 0.03846154; \quad 4 \cdot C22111 - 11.94 \cdot C31111 = 4 \cdot \frac{5}{31} - 11.94 \cdot \frac{5}{93} \approx 0.0032258.$$

Notice that in the iterated blow-up of C_5, in the limit $4 \cdot C22111 - 12 \cdot C31111 = 0$. For our method to work, we need a lower bound greater than zero. On the other hand, computational experiments convinced us that the method works best if the bound is only slightly above zero, where a suitable factor is again determined by computations.

Let G be an extremal graph on n vertices, where n is sufficiently large to apply Proposition 3. Denote the set of all induced C_5s in G by $\mathcal{Z}$. We assume that $a \in \mathbb{R}$ and $Z = z_1z_2z_3z_4z_5$ is an induced C_5 maximizing $C22111(Z) - a \cdot C31111(Z)$. Then

$$(C22111(Z) - a \cdot C31111(Z)) \binom{n-5}{2} \geq \frac{1}{|\mathcal{Z}|} \sum_{Y \in \mathcal{Z}} (C22111(Y) - a \cdot C31111(Y)) \binom{n-5}{2} =$$

$$= \frac{(4 \cdot C22111 - 3a \cdot C31111)\binom{n}{7}}{C_5\binom{n}{5}} = \frac{\frac{4}{21}C22111 - \frac{a}{7}C31111}{C_5}\binom{n-5}{2}.$$

As mentioned above, computations indicate that we get the most useful bounds if $C22111(Z) - a \cdot C31111(Z)$ is close but not too close to 0. Using (1) and setting $a = 3.98$, we get

$$C22111(Z) - 3.98 \cdot C31111(Z) > 0.0039792. \tag{2}$$

For $1 \leq i \leq 5$, we define sets of vertices Z_i which look like z_i to the other vertices of Z. Formally,

$$Z_i := \{v \in V(G) : G[(Z \setminus z_i) \cup v] \cong C_5\} \text{ for } 1 \leq i \leq 5.$$

Note that $Z_i \cap Z_j = \emptyset$ for $i \neq j$. We call a pair v_iv_j *funky*, if v_iv_j is an edge but z_iz_j is not an edge or vice versa, where $v_i \in Z_i$, $v_j \in Z_j$, $1 \leq i < j \leq 5$. In other words, $G[Z \cup \{v_i, v_j\}] \ncong C22111$, i.e., every funky pair destroys a potential copy of $C22111(Z)$. Denote by E_f the set of funky pairs. With this notation, (2) implies that for large n we have

$$\sum_{1 \leq i < j \leq 5} |Z_i||Z_j| - |E_f| - 3.98 \sum_{i \in [5]} |Z_i|^2/2 > 0.003979\binom{n-5}{2}.$$

For any choice of sets $X_i \subseteq Z_i$, where $i \in [5]$, let $X_0 := V(G) \setminus \bigcup X_i$. Let f be the number of funky pairs not incident to vertices in X_0, divided by n^2 for normalization, and denote $x_i = \frac{1}{n}|X_i|$ for $i \in \{0, \ldots, 5\}$. Choose the X_i (possibly $X_i = Z_i$) such that the left hand side in

$$2 \sum_{1 \leq i < j \leq 5} x_ix_j - 2f - 3.98 \sum_{i \in [5]} x_i^2 > 0.003979 \tag{3}$$

is maximized. In order to simplify notation, we use $X_{i+5} = X_i$ and $x_{i+5} = x_i$ for all $i \geq 1$.

Claim 4. *The following inequalities are satisfied:*

$$0.19816 < x_i < 0.20184 \quad \textit{for} \quad i \in [5]; \tag{4}$$
$$x_0 < 0.00263; \tag{5}$$
$$f < 0.000011. \tag{6}$$

Proof. To obtain (4)–(6), we need to solve four quadratic programs. The objectives are to minimize x_1, maximize x_1, maximize x_0, and to maximize f, respectively. The constraints are (3) and $\sum_{i=0}^{5} x_i = 1$ in all four cases. By symmetry, bounds for x_1 apply also for x_2, x_3, x_4, and x_5.

Here we describe the process of obtaining the lower bound on x_1 in (4). We need to solve the following program (P):

$$(P)\begin{cases} \text{minimize} & x_1 \\ \text{subject to} & \sum_{i=0}^{5} x_i = 1, \\ & 2\sum_{1\le i<j\le 5} x_i x_j - 2f - 3.98\sum_{i\in[5]} x_i^2 > 0.003979, \\ & x_i \ge 0 \text{ for } i \in \{0,1,\ldots,5\}. \end{cases}$$

We claim that if (P) has a feasible solution S, then there exists a feasible solution S' of (P) where

$$S'(x_1) = S(x_1), \quad S'(f) = 0, \quad S'(x_0) = S(x_0),$$
$$S'(x_2) = S'(x_3) = S'(x_4) = S'(x_5) = \frac{1}{4}\big(1 - S(x_1) - S(x_0)\big).$$

Since x_2, x_3, x_4 and x_5 appear only in constraints, we only need to check whether (3) is satisfied. The left hand side of (3) can be rewritten as

$$\begin{aligned} & 2x_1 \sum_{2\le i<j\le 5} x_i + 2\sum_{2\le i<j\le 5} x_i x_j - 3.98 \sum_{1\le i<j\le 5} x_i^2 - 2f \\ = \ & 2x_1 \sum_{2\le i<j\le 5} x_i - \sum_{2\le i<j\le 5} (x_i - x_j)^2 - 0.98 \sum_{2\le i<j\le 5} x_i^2 - 3.98x_1^2 - 2f. \end{aligned}$$

Note that the term $\sum_{2\le i<j\le 5}(x_i - x_j)^2$ is minimized if $x_i = x_j$ for all $i, j \in \{2,3,4,5\}$. The term $x_2^2 + x_3^2 + x_4^2 + x_5^2$, subject to $x_2 + x_3 + x_4 + x_5$ being a constant, is also minimized if $x_i = x_j$ for all $i, j \in \{2,3,4,5\}$. Since $f \ge 0$, the term $2f$ is minimized when $f = 0$. Hence (3) is satisfied by S' and we can add the constraints $x_2 = x_3 = x_4 = x_5$ and $f = 0$ to bound x_1. The resulting program (P') is

$$(P')\begin{cases} \text{minimize} & x_1 \\ \text{subject to} & x_0 + x_1 + 4y = 1, \\ & 8x_1 y - 0.98 \cdot 4y^2 - 3.98x_1^2 \ge 0.003979, \\ & x_0, x_1, y \ge 0. \end{cases}$$

We solve (P') using Lagrange multipliers. We delegate the work to Sage [28] and we provide the Sage script at `http://orion.math.iastate.edu/lidicky/pub/c5/`. Finding an upper bound on x_1 is done by changing the objective to maximization.

Similarly, we can set $x_1 = x_2 = x_3 = x_4 = x_5 = 1/5$ to get an upper bound on f. We can set $f = 0$ and $x_1 = x_2 = x_3 = x_4 = x_5 = (1 - x_0)/5$ to get an upper bound on x_0. We omit the details. Sage scripts for solving the resulting programs are provided at `http://orion.math.iastate.edu/lidicky/pub/c5/`. □

For any vertex $v \in X_i, i \in [5]$ we use $d_f(v)$ to denote the number of funky pairs from v to $(X_1 \cup X_2 \cup X_3 \cup X_4 \cup X_5) \setminus X_i$ after normalizing by n. If we move v from X_1 to X_0, then the left hand side of (3) will decrease by

$$\frac{1}{n}\left(2(x_2 + x_3 + x_4 + x_5) - 2d_f(v) - 2 \cdot 3.98 \cdot x_1 + o(1)\right).$$

If this quantity was negative, then the left hand side of (3) could be increased by moving v to X_0, contradicting our choice of X_i. This together with (4) implies that

$$d_f(v) \le x_2 + x_3 + x_4 + x_5 - 3.98 \cdot x_1 + o(1) \le 1 - 4.98 \cdot x_1 + o(1) \le 0.0132. \tag{7}$$

Symmetric statements hold also for every vertex $v \in X_2 \cup X_3 \cup X_4 \cup X_5$.

Claim 5. *There are no funky pairs in $X_1 \cup X_2 \cup X_3 \cup X_4 \cup X_5$.*

Proof. Assume that there is a funky pair uv. By symmetry, we only need to consider two cases, either $u \in X_1, v \in X_2$ or $u \in X_1, v \in X_3$. In fact, it is sufficient to check the case where $u \in X_1$ and $v \in X_2$, so uv is not an edge. The other case then follows from considering the complement of G.

Let G' be a graph obtained from G by adding the edge uv, i.e., changing uv to be not funky. We compare the number of induced C_5s containing $\{u, v\}$ in G and in G'. In G', there are at least

$$\left[x_3x_4x_5 - (d_f(u) + d_f(v)) \max\{x_3x_4, x_3x_5, x_4x_5\} - f \cdot \max\{x_3, x_4, x_5\}\right] n^3$$

induced C_5s containing uv, since we can pick one vertex from each of X_3, X_4, X_5 to form an induced C_5 as long as none of the resulting nine pairs is funky.

Now we count the number of induced C_5s in G containing $\{u, v\}$. The number of such C_5s which contain vertices from X_0 is upper bounded by $x_0n^3/2$. Next we count the number of such C_5s avoiding X_0. Observe that there are no C_5s avoiding X_0 in which uv is the only funky pair.

The number of C_5s containing another funky pair $u'v'$ with $\{u, v\} \cap \{u', v'\} = \emptyset$ can be upper bounded by fn^3. We are left to count C_5s where the other funky pairs contain u or v. The number of C_5s containing at least two vertices other than u and v which are in funky pairs can be upper bounded by $(d_f(u)^2/2 + d_f(v)^2/2 + d_f(u)d_f(v))n^3$.

It remains to count only C_5s containing exactly one vertex w where uw and vw are the options for funky pairs. The number of choices of w is at most $(d_f(u)+d_f(v))n$. As $\{u, v, w\}$

is in an induced C_5, the set $\{u,v,w\}$ induces a path in either G or the complement of G. Let the middle vertex of that path be in X_i. If $G[\{u,v,w\}]$ is a path, then the remaining two vertices of a C_5 cannot be in $X_{i+1} \cup X_{i+4}$. If $G[\{u,v,w\}]$ is the complement of a path, then the remaining two vertices cannot be in $X_{i+2} \cup X_{i+3}$. Hence the remaining two vertices of a C_5 containing $\{u,v,w\}$ can be chosen from at most $3n \cdot \max\{x_i\}$ vertices. This gives an upper bound of $(d_f(u)+d_f(v))n\binom{3n\cdot\max\{x_i\}}{2}$ on the number of such C_5s.

Now we compare the number of induced C_5s containing uv in G and in G'. We use x_{max} and x_{min} to denote the upper and lower bound respectively from (4), use d_f to denote the upper bound on $d_f(u)$ and $d_f(v)$ from (7), and also use bounds from (5) and (6). The number of C_5s containing uv divided by n^3 is

$$\begin{aligned} \text{in } G: &\leq x_0/2 + f + 2d_f^2 + 9d_f x_{max}^2 \leq 0.0065;\\ \text{in } G': &\geq (x_{min} - 2d_f)x_{min}^2 - f x_{max} \geq 0.0067. \end{aligned}$$

This contradicts the extremality of G. $\square$

Next, we want to show that $X_0 = \emptyset$. For this, suppose that there exists an $x \in X_0$. We will add x to one of the X_i, $i \in [5]$ such that $d_f(x)$ is minimal. By symmetry, we may assume that x is added to X_1. Note that adding a single vertex to X_1 does not change any of the density bounds we used above by more than $o(1)$.

Claim 6. *For every $x \in X_0$, if x is added to X_1 then $d_f(x) \geq 0.0808$.*

Proof. Let xw be a funky pair, where $w \in X_2$. The case where $w \in X_3$ can be argued the same way by considering the complement of G. Let G' be obtained from G by adding the edge xw. Since G is extremal, we have $C(G') \leq C(G)$. The following analysis is similar to the proof of Claim 5, however, we can say a bit more since every funky pair contains x.

First we count induced C_5s containing xw in G. The number of induced C_5s containing xw and other vertices from X_0 is easily bounded from above by $x_0 n^3/2$.

Let F be an induced C_5 in G containing xw and avoiding $X_0 \setminus \{x\}$. Since all funky pairs contain x, $F - x$ is an induced path $p_0p_1p_2p_3$ without funky pairs. Either $p_j \in X_2$ for all $j \in \{0,1,2,3\}$ or there is an $i \in \{1,2,3,4,5\}$ such that $p_j \in X_{i+j}$ for all $j \in \{0,1,2,3\}$. The first case is depicted in Figure 3(a). Consider now the second case. If $i \in \{2,3,4\}$, then $xp_0p_1p_2p_3$ does not satisfy the definition of F. Hence $i \in \{1,5\}$ and the possible C_5s are depicted in Figure 3(b) and (c). In each of the three cases, F contains exactly two funky pairs, xw and xy. The location of y entirely determines the location of $F - x$. Hence the number of induced C_5s containing xw is at most $d_f(x)x_{max}^2 n^3$.

In G', there are at least $\big(x_3x_4x_5 - d_f(x)\cdot\max\{x_3x_4, x_3x_5, x_4x_5\}\big)n^3$ induced C_5s containing xw. We obtain

$$C(G)/n^3 \leq d_f(x)x_{max}^2 + x_0/2 \qquad \text{and} \qquad C(G')/n^3 \geq (x_{min} - d_f(x))x_{min}^2.$$

Since $C(G') \leq C(G)$, we have

$$(x_{min} - d_f(x))x_{min}^2 \leq d_f(x)x_{max}^2 + x_0/2,$$

which together with (4) and (5) gives $d_f(x) \geq 0.0808$. $\square$

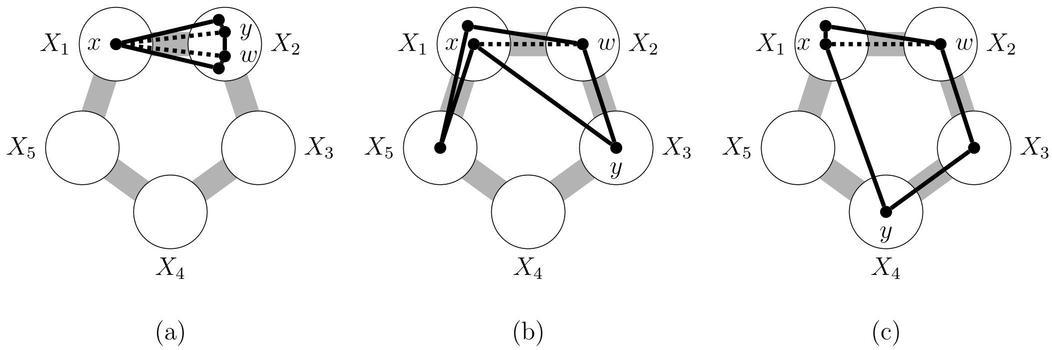


Figure 3: Possible C_5s with funky pair xw. They all have exactly one other funky pair xy. The dotted lines represent non-edges.

Claim 7. *Every vertex of the extremal graph G is in at least $(1/26+o(1))\binom{n}{4} \approx 0.001602564n^4$ induced C_5s.*

Proof. For every vertex $u \in V(G)$, denote by C_5^u the number of C_5s in G containing u. For any two vertices $u, v \in V(G)$, we show that $C_5^u - C_5^v < n^3$, which implies Claim 7. Denote by C_5^{uv} the number of C_5s in G containing both u and v. A trivial bound is $C_5^{uv} \leq \binom{n-2}{3}$.

Let G' be obtained from G by deleting v and duplicating u to u', i.e., for every vertex x we add the edge xu' iff xu is an edge. As G is extremal we have

$$0 \geq C(G') - C(G) \geq C_5^u - C_5^v - C_5^{uv} \geq C_5^u - C_5^v - \binom{n-2}{3}.$$

$\square$

Claim 8. *The set X_0 is empty.*

Proof. Assume that there is an $x \in X_0$. We count C_5^x, the number of induced C_5s containing x. Our goal is to show that C_5^x is smaller than the value in Claim 7, which is a contradiction. Let $a_i n$ be the number of neighbors of x in X_i and $b_i n$ be the number of non-neighbors of x in X_i for $i \in \{0, 1, 2, 3, 4, 5\}$.

The number of C_5s where the other four vertices are in $X_1 \cup X_2 \cup X_3 \cup X_4 \cup X_5$ is upper bounded by

$$\left(a_1 b_2 b_3 a_4 + a_2 b_3 b_4 a_5 + a_3 b_4 b_5 a_1 + a_4 b_5 b_1 a_2 + a_5 b_1 b_2 a_3 + \frac{1}{4}\sum_{i=1}^{5} a_i^2 b_i^2\right) n^4.$$

Moreover, we also need to include the C_5s containing vertices from X_0 in our bound, which we do very generously by increasing all variables by a_0 or b_0.

Since $x_i = a_i + b_i$, we can use (4) for every $i \in [5]$ as constraints. We also use Claim 6 to obtain constraints since it is possible to express $d_f(x)$ using a_is and b_is if x is added to X_j for all $i, j \in [5]$.

By combining the previous objective and constraints, we obtain the following program (P), whose objective gives an upper bound on the number of C_5s containing x divided by n^4.

$$(P)\begin{cases}\text{maximize} & \sum_{i=1}^5 (a_i+a_0)(b_{i+1}+b_0)(b_{i+2}+b_0)(a_{i+3}+a_0)+\frac{1}{4}\sum_{i=1}^5 a_i^2b_i^2\\ \text{subject to} & \sum_{i=0}^5 (a_i+b_i)=1,\\ & 0.19816 \le a_i+b_i \le 0.20184 \text{ for } i\in\{1,2,3,4,5\},\\ & a_0+b_0 \le 0.00263,\\ & b_2+b_5+a_3+a_4 \ge 0.0808,\\ & b_1+b_3+a_4+a_5 \ge 0.0808,\\ & b_2+b_4+a_1+a_5 \ge 0.0808,\\ & b_3+b_5+a_1+a_2 \ge 0.0808,\\ & b_4+b_1+a_2+a_3 \ge 0.0808,\\ & a_i,b_i \ge 0 \text{ for } i\in\{0,1,2,3,4,5\}.\end{cases}$$

Instead of solving (P) we solve a slight relaxation (P') with increased upper bounds on a_i+b_i, which allows us to drop a_0 and b_0. Since the objective function is maximizing, we can claim that a_i+b_i is always as large as possible, which decreases the number of the degrees of freedom.

$$(P')\begin{cases}\text{maximize} & f=\sum_{i=1}^5 a_ib_{i+1}b_{i+2}a_{i+3}+\frac{1}{4}\sum_{i=1}^5 a_i^2b_i^2\\ \text{subject to} & a_i+b_i=0.21 \text{ for } i\in\{1,2,3,4,5\},\\ & b_2+b_5+a_3+a_4 \ge 0.0808,\\ & b_1+b_3+a_4+a_5 \ge 0.0808,\\ & b_2+b_4+a_1+a_5 \ge 0.0808,\\ & b_3+b_5+a_1+a_2 \ge 0.0808,\\ & b_4+b_1+a_2+a_3 \ge 0.0808,\\ & a_i,b_i \ge 0 \text{ for } i\in\{1,2,3,4,5\}.\end{cases}$$

Note that the resulting program (P') has only 5 degrees of freedom. We find an upper bound on the solution of (P') by a brute force method. We discretize the space of possible solutions, and bound the gradient of the target function to control the behavior between the grid points.

For solving (P'), we fix a constant s which will correspond to the number of steps. For every a_i we check $s+1$ equally spaced values between 0 and 0.21 that include the boundaries. By this we have a grid of s^5 boxes where every feasible solution of (P'), and hence also of (P), is in one of the boxes.

Next we need to find the partial derivatives of f. Since f is symmetric, we only check the partial derivative with respect to a_1.

$$\frac{\partial f}{\partial a_1}=b_2b_3a_4+a_3b_4b_5+\frac{1}{2}a_1b_1^2.$$

We want to find an upper bound on $\frac{\partial f}{\partial a_1}$. Hence we assume $a_1 + b_1 = a_3 + b_3 = a_4 + b_4 = b_2 = b_5 = 0.21$ and we maximize

$$b_2 b_3 a_4 + a_3 b_4 b_5 = 0.21\left((0.21 - a_3)a_4 + a_3(0.21 - a_4)\right) = 0.21\left(0.21a_4 + 0.21a_3 - 2a_3a_4\right).$$

This is maximized if $a_3 = 0, a_4 = 0.21$ or $a_3 = 0.21, a_4 = 0$ and gives the value 0.21^3. Hence

$$\frac{1}{2}a_1 b_1^2 = \frac{4}{2}a_1 \cdot \frac{b_1}{2} \cdot \frac{b_1}{2} \le \frac{2(a_1 + b_1)^3}{3^3} = \frac{2 \cdot 0.21^3}{27}.$$

The resulting upper bound is

$$\frac{\partial f}{\partial a_1} \le 0.21^3 + \frac{2 \cdot 0.21^3}{27} < 0.001.$$

Hence in a box with side length t the value of f cannot be bigger than the value at a corner plus $5t/2 \cdot 0.001$. The factor $5t/2$ comes from the fact that the closest corner is in distance at most $t/2$ in each of the 5 coordinates.

If we set $s = 100$, we compute that the maximum over all grid points of (P'') is less than 0.00157. This can be checked by a computer program `mesh-opt.cpp` which computes the values at all grid points. With $t < 0.21/s = 0.0021$, we have $5t/2 \cdot 0.001 < 0.00001$. We conclude that x is in less than $0.00158n^4$ induced C_5s which contradicts Claim 7.

Let us note that if we had chosen $s = 200$, we could have concluded that x is less than $0.00147n^4$. □

We have just established the "outside" structure of G. Observe that in this outside structure, an induced C_5 can appear only if it either intersects each of the classes in exactly one vertex, or if it lies completely inside one of the classes. This implies that

$$C(n) = (x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5)n^5 + C(x_1 n) + C(x_2 n) + C(x_3 n) + C(x_4 n) + C(x_5 n).$$

By averaging over all subgraphs of G of order $n-1$, we can easily see that $C(n) \le \frac{n}{n-5}C(n-1)$ for all n, so

$$\ell := \lim_{n\to\infty} \frac{C(n)}{\binom{n}{5}}$$

exists. Therefore,

$$\ell + o(1) = 5! \cdot x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5 + \ell(x_1^5 + x_2^5 + x_3^5 + x_4^5 + x_5^5),$$

which implies that $x_i = \frac{1}{5} + o(1)$, and $\ell = \frac{1}{26}$, given the constraints on the x_i.

In order to prove Theorem 2, it remains to show that in fact $|X_i| - |X_j| \le 1$ for all $i, j \in \{1, \ldots, 5\}$.

Claim 9. *For n large enough, we have $|X_i| - |X_j| \le 1$ for all $i, j \in \{1, \ldots, 5\}$.*

Proof. By symmetry, assume for contradiction that $|X_1|-|X_2| \geq 2$. Let $v \in X_1$ where C_5^v is minimized over the vertices in X_1 and let $w \in X_2$ where C_5^w is maximized over the vertices in X_2. As G is extremal, $C_5^v + C_5^{vw} - C_5^w \geq 0$; otherwise, we can increase the number of C_5s by replacing v by a copy of w.

Let $y_i := |X_i| = x_i n$. By the monotonicity of $\frac{C(n)}{\binom{n}{5}}$, we have

$$\frac{1}{26} + o(1) \geq \frac{C(y_2)}{\binom{y_2}{5}} \geq \frac{C(y_1)}{\binom{y_1}{5}} \geq \frac{1}{26} - o(1).$$

Therefore, using $y_1 - y_2 \geq 2$, we have

$$\begin{aligned}
C_5^v + C_5^{vw} - C_5^w &\leq \frac{C(y_1)}{y_1} + y_2y_3y_4y_5 + y_3y_4y_5 - \frac{C(y_2)}{y_2} - y_1y_3y_4y_5 \\
&= \frac{y_2C(y_1) - y_1C(y_2)}{y_1y_2} + (y_2 - y_1 + 1)y_3y_4y_5 \\
&\leq \left(\frac{1}{26} + o(1)\right)\frac{1}{y_1y_2}\left(y_2\binom{y_1}{5} - y_1\binom{y_2}{5}\right) + (y_2 - y_1 + 1)y_3y_4y_5 \\
&\leq \left(\frac{1}{26 \cdot 5!} + o(1)\right)\left(y_1^4 - y_2^4\right) + (y_2 - y_1 + 1)y_3y_4y_5 \\
&= \left(\frac{1}{26 \cdot 5!} + o(1)\right)(y_1 - y_2)\left(y_1^3 + y_1^2y_2 + y_1y_2^2 + y_2^3\right) + (y_2 - y_1 + 1)y_3y_4y_5 \\
&= (y_1 - y_2)\left(\left(\frac{1}{26 \cdot 5!} + o(1)\right)\frac{4n^3}{125} - \frac{n^3}{125}\right) + \frac{(1 + o(1))n^3}{125} \\
&\leq \left(\frac{2}{26 \cdot 5!} + o(1)\right)\frac{4n^3}{125} - \frac{(1 + o(1))n^3}{125} < 0,
\end{aligned}$$

a contradiction. $\square$

With this claim, the proof of Theorem 2 is complete. $\square$

4 Proof of Theorem 1

Theorem 1 is a consequence of Theorem 2. The main proof idea is to take a minimal counterexample G and show that some blow-up of G contradicts Theorem 2.

Proof of Theorem 1. Theorem 1 is easily seen to be true for $k = 1$. Suppose for a contradiction that there is a graph G on $n = 5^k$ vertices with $C(G) \geq C(C_5^{k\times})$ that is not isomorphic to $C_5^{k\times}$, where $k \geq 2$ is minimal. Let n_0 be the n_0 from the statement of Theorem 2.

We say that a graph F of size $5m$ can be *5-partitioned*, if $V(F)$ can be partitioned into five sets X_1, X_2, X_3, X_4, X_5 with $|X_i| = m$ for all $i \in [5]$ and for every $1 \leq i < j \leq 5$, every $x_i \in X_i$ and $x_j \in X_j$ are adjacent if and only if $|i - j| \in \{1, 4\}$. Notice that this is the structure described by Theorem 2. Hence if $5m \geq n_0$, and F is extremal then F can be 5-partitioned.

If G can be 5-partitioned, then G is isomorphic to $C_5^{k\times}$ by the minimality of k, a contradiction. Therefore, G cannot be 5-partitioned.

Let H be an extremal graph on $5^\ell > n_0$ vertices. Blowing up every vertex of $C_5^{k\times}$ by a factor of 5^ℓ, and inserting H in every part, gives an extremal graph G_1 on $5^{k+\ell}$ vertices by ℓ applications of Theorem 2. On the other hand, the graph G_2 obtained by blowing up every vertex of G by a factor of 5^ℓ, and inserting H in every part, contains at least as many C_5s as G_1,

$$C(G_1) = 5^k \cdot C(H) + C(C_5^{k\times}) \cdot (5^\ell)^5, \qquad C(G_2) = 5^k \cdot C(H) + C(G) \cdot (5^\ell)^5,$$

so $C(G_1) \leq C(G_2)$. Hence G_2 must also be extremal. Therefore G_2 can be 5-partitioned into five sets X_1, X_2, X_3, X_4, X_5 with $|X_i| = 5^{k+\ell-1}$. In particular, two vertices in G_2 are in the same set X_i if and only if their adjacency pattern agrees on more than half of the remaining vertices. But this implies that for every copy H' of H inserted into the blow-up of G, all vertices of H' are in the same X_i, and thus the 5-partition of $V(G_2)$ gives a 5-partition of $V(G)$, a contradiction. $\square$

Acknowledgement

We would like to thank Jan Volec for fruitful discussions and two anonymous referees for a careful reading of the manuscript and several suggestions improving the write-up.

References

[1] R. Baber. Turán densities of hypercubes. Submitted, available as arXiv:1201.3587.

[2] R. Baber and J. Talbot. A solution to the 2/3 conjecture. *SIAM J. Discrete Math.*, 28(2):756–766, 2014.

[3] J. Balogh, P. Hu, B. Lidický, and H. Liu. Upper bounds on the size of 4- and 6-cycle-free subgraphs of the hypercube. *European Journal of Combinatorics*, 35:75–85, 2014.

[4] J. Balogh, P. Hu, B. Lidický, F. Pfender, J. Volec, and M. Young. Rainbow triangles in three-colored graphs. Submitted, avilable as arXiv:1408.5296.

[5] J. Balogh, P. Hu, B. Lidický, O. Pikhurko, B. Udvari, and J. Volec. Minimum number of monotone subsequences of length 4 in permutations. *To appear in Combinatorics, Probability and Computing.*

[6] J. Cummings, D. Kráľ, F. Pfender, K. Sperfeld, A. Treglown, and M. Young. Monochromatic triangles in three-coloured graphs. *J. Combin. Theory Ser. B*, 103(4):489–503, 2013.

[7] S. Das, H. Huang, J. Ma, H. Naves, and B. Sudakov. A problem of Erdős on the minimum number of k-cliques. *J. Combin. Theory Ser. B*, 103(3):344–373, 2013.

[8] C. Even-Zohar and N. Linial. A note on the inducibility of 4-vertex graphs. *Graphs and Combinatorics*, pages 1–14, 2014.

[9] V. Falgas-Ravry and E. R. Vaughan. Turán H-densities for 3-graphs. *Electron. J. Combin.*, 19:R40, 2012.

[10] R. Glebov, D. Kráľ, and J. Volec. A problem of Erdős and Sós on 3-graphs. Submitted, available as arXiv:1303.7372.

[11] A. Grzesik. On the maximum number of five-cycles in a triangle-free graph. *J. Combin. Theory Ser. B*, 102(5):1061–1066, 2012.

[12] H. Hatami, J. Hirst, and S. Norine. The inducibility of blow-up graphs. *Journal of Combinatorial Theory, Series B*, pages 196–212, 2014.

[13] H. Hatami, J. Hladký, D. Kráľ, S. Norine, and A. A. Razborov. On the number of pentagons in triangle-free graphs. *J. Combin. Theory Ser. A*, 120(3):722–732, 2013.

[14] J. Hirst. The inducibility of graphs on four vertices. *J. Graph Theory*, 75(3):231–243, 2014.

[15] J. Hladký, D. Kráľ, and S. Norine. Counting flags in triangle-free digraphs. In *European Conference on Combinatorics, Graph Theory and Applications (EuroComb 2009)*, volume 34 of *Electron. Notes Discrete Math.*, pages 621–625. Elsevier Sci. B. V., Amsterdam, 2009.

[16] H. Huang. On the maximum induced density of directed stars and related problems. *SIAM J. Discrete Math.*, 28(1):92–98, 2014.

[17] D. Kráľ, L. Mach, and J.-S. Sereni. A new lower bound based on Gromov's method of selecting heavily covered points. *Discrete Comput. Geom.*, 48(2):487–498, 2012.

[18] O. Pikhurko. On possible Turán densities. *Israel J. Math.*, 201(1):415–454, 2014.

[19] O. Pikhurko and E. R. Vaughan. Minimum number of k-cliques in graphs with bounded independence number. *To appear in Combin. Probab. Comput.*,.

[20] N. Pippenger and M. C. Golumbic. The inducibility of graphs. *J. Combin. Theory Ser. B*, 19(3):189–203, 1975.

[21] A. A. Razborov. Flag algebras. *Journal of Symbolic Logic*, 72(4):1239–1282, 2007.

[22] A. A. Razborov. On the minimal density of triangles in graphs. *Combin. Probab. Comput.*, 17(4):603–618, 2008.

[23] A. A. Razborov. Flag Algebras: An Interim Report. In *The Mathematics of Paul Erdős II*, pages 207–232. Springer, 2013.

[24] A. A. Razborov. On the Caccetta-Häggkvist conjecture with forbidden subgraphs. *J. Graph Theory*, 74(2):236–248, 2013.

[25] A. A. Razborov. Flag algebras: an interim report, 2014. Talk at IMA Annual Program `http://www.ima.umn.edu/videos/index.php?id=2756`.

[26] C. Reiher. The clique density theorem, 2012. Submitted, available as arXiv:1212.2454.

[27] K. Sperfeld. The inducibility of small oriented graphs, 2011. Manuscript arXiv:1111.4813.

[28] W. Stein et al. *Sage Mathematics Software (Version 6.3)*. The Sage Development Team, 2014. `http://www.sagemath.org`.

[29] E. R. Vaughan. Flagmatic 2.0, 2012. `http://flagmatic.org`.