Minimizing the number of 5-cycles in graphs with given edge-density

Patrick Bennett *   Andrzej Dudek †   Bernard Lidický ‡

March 2, 2018

Abstract

Motivated by the work of Razborov about the minimal density of triangles in graphs we study the minimal density of cycles C_5. We show that every graph of order n and size $\left(1-\frac{1}{k}\right)\binom{n}{2}$, where $k \geq 3$ is an integer, contains at least

$$\left(\frac{1}{10}-\frac{1}{2k}+\frac{1}{k^2}-\frac{1}{k^3}+\frac{2}{5k^4}\right)n^5+o(n^5)$$

copies of C_5. This bound is optimal, since a matching upper bound is given by the balanced complete k-partite graph. The proof is based on the flag algebras framework. We also provide a stability result for $2 \leq k \leq 73$.

1 Introduction

It is believed that *extremal graph theory* was started by Turán [24] when he proved that any graph on n vertices with more than $\frac{r-2}{2(r-1)}n^2$ edges must contain an r-clique (i.e., a copy of K_r). The case $r = 3$ was earlier proved by Mantel [14]. The general Turán problem is to determine the minimum number $\mathrm{ex}(n, H)$ of edges in an n vertex graph that guarantees a copy of a graph H, and has been very widely studied. The Erdős and Stone theorem [6] was a major breakthrough which asymptotically determined the value of $\mathrm{ex}(n, H)$ for all nonbipartite H. For such H we have

$$\mathrm{ex}(n, H) = \frac{\chi(G)-2}{2(\chi(G)-1)}n^2 + o(n^2).$$

*Department of Mathematics, Western Michigan University, Kalamazoo, MI, USA. E-mail: patrick.bennett@wmich.edu. Supported in part by Simons Foundation Grant #426894.

†Department of Mathematics, Western Michigan University, Kalamazoo, MI, USA. E-mail: andrzej.dudek@wmich.edu. Supported in part by Simons Foundation Grant #522400.

‡Department of Mathematics, Iowa State University. Ames, IA, USA. E-mail: lidicky@iastate.edu. Supported in part by NSF grant DMS-1600390.

The natural quantitative question that arises is how many copies of H must be contained in a graph G on n vertices and $m > ex(n, H)$ edges. This question has also been well studied. Obviously the number of edges m can be expressed as a density parameter p such that $m = p\binom{n}{2}$. Therefore, we will use the following notation. Let G be a (large) graph of order n and H a small one. Define $\nu_H(G)$ to be the number of copies (not necessary induced) of H in G and the corresponding density as

$$d_H(G) = \frac{\nu_H(G)}{|V(G)|^{|V(H)|}}.$$

Furthermore, for a given number $p \in [0, 1]$ let

$$d_H(p) = \lim_{n\to\infty} \min_G d_H(G),$$

where the minimum is taken over all graphs G of order n and size $p\binom{n}{2}$, assuming the limit exists.

When $H = K_3$ (that means it is a triangle) Moon and Moser [15] and also independently Nordhaus and Stewart [17] determined $d_{K_3}(p)$ for any $p = 1 - \frac{1}{k}$, where k is a positive integer. We call such $p = 1 - \frac{1}{k}$ a *Turán density*. Some other partial results for $H = K_r$ were established by Lovász and Simonovits [13]. However, for arbitrary p these problems remained open for over 50 years.

In 2007 Razborov in his seminal paper [20] introduced the so-called *flag algebras* and determined $d_{K_3}(p)$ for any p [21]. Subsequently, Pikhurko and Razborov [18] characterized the nearly extremal graphs. Very recently, Liu, Pikhurko and Staden [12] found the precise minimum number of triangles among graphs with a given number of edges. Nikiforov [16] found $d_{K_4}(p)$ for all p, and then Reiher [22] found $d_{K_r}(p)$ for all r and p.

In this paper we address the minimum density of the 5-cycle, C_5, in a graph with given edge density. We chose to investigate C_5 instead of C_4 since it is known due to Sidorenko [23] that for any fixed constant edge density p, the minimum C_4-density is achieved asymptotically by the random graph $G_{n,p}$. It is worth mentioning some other research related to 5-cycles. Specifically, Grzesik [8] and independently Hatami, Hladký, Král', Norine and Razborov [9] proved that the maximum density of 5-cycles in a triangle-free graph that is large or its number of vertices is a power of 5 is achieved by the balanced blow-up of a 5-cycle. The extension to graphs of all sizes, with one exception on 8 vertices, was done by Lidický and Pfender [11]. This settled in the affirmative a conjecture of Erdős [5]. On the other hand, Balogh, Hu, and Lidický, and Pfender [2] studied the problem of maximizing induced 5-cycles, and proved that this is also achieved by the balanced iterated blow-up of a 5-cycle. This confirmed a special case of a conjecture of Pippinger and Golumbic [19].

Here we present the main result of this paper.

Theorem 1. *Let $k \geq 3$ be an integer and $p = 1 - \frac{1}{k}$. Then,*

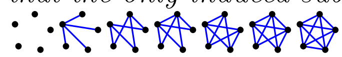
$$d_{C_5}(p) = \frac{1}{10} - \frac{1}{2k} + \frac{1}{k^2} - \frac{1}{k^3} + \frac{2}{5k^4}.$$

Observe that this bound is consistent with the case $k = 2$ for which $d_{C_5}(\frac{1}{2}) = 0$. (A complete balanced bipartite graph has the right density and no copy of C_5.) Although the proof of Theorem 1 is based on the flag algebras framework, it does not require using of any SDP solver (see Section 2).

We also show the following stability-type result.

Theorem 2. *Let G be a graph on n vertices for large n, such that G has edge density $p = 1 - \frac{1}{k}$ for $k \geq 2$ and*

$$d_{C_5}(G) \leq d_{C_5}(p) + \epsilon$$

for some positive but sufficiently small ϵ. Assume further that the only induced subgraphs on five vertices with density more than ϵ are the graphs in:  *. Then G has edit distance at most δn^2 from the Turán graph T_n^k, for some function $\delta = \delta(\epsilon) \to 0$ as $\epsilon \to 0$.*

The proof of this theorem is technical but elementary (see Section 3). We are also able to show that for each $k \in \{2, \dots, 73\}$ in Theorem 2 the assumption about non-zero induced subgraph densities hold. (Here actually we use an SDP solver.) Thus, for any $k \in \{2, \dots, 73\}$ the graphs with density $p = 1 - \frac{1}{k}$ that minimize the number of copies of C_5 are "close" to the Turán graph.

We also discuss extremal constructions and provide a general upper bound on $d_{C_5}(p)$ for any p (see Section 4).

2 Proof of the main theorem

2.1 Upper bound

Let T_k^n be a complete balanced k-partite graph on n vertices. By considering the sequence of graphs T_k^n, we get

$$d_{C_5}(T_k^n) = \frac{\left[\frac{1}{10}(k)_5 + \frac{1}{2}(k)_4 + \frac{1}{2}(k)_3\right]\left(\frac{n}{k}\right)^5}{n^5} + o(1),$$

where $(k)_\ell = k(k-1)\cdots(k-\ell+1)$ is the *falling factorial*. To justify the numerator, we count the number of C_5 copies with vertices in parts V_1, V_2, V_3, V_4, V_5 of the partition. These parts may not all be distinct: for example we may have $V_1 = V_3$. However T_k^n has no edges within these parts and so we know $V_i \neq V_{i+1}$. We count copies of C_5 by grouping them according to how many distinct parts there are among $V_1, \dots, V_5$. Now there are asymptotically $\frac{1}{10}(k)_5\left(\frac{n}{k}\right)^5$ copies that hit 5 different parts (label 5 distinct parts, choose one vertex in each part, and divide by 10 for overcounting). Now asymptotically there are $\frac{1}{2}(k)_4\left(\frac{n}{k}\right)^5$ hitting 4 parts, and $\frac{1}{2}(k)_3\left(\frac{n}{k}\right)^5$ hitting 3 parts.

Simplifying, we get that

$$d_{C_5}(T_k^n) = \frac{1}{10} - \frac{1}{2k} + \frac{1}{k^2} - \frac{1}{k^3} + \frac{2}{5k^4} + o(1),$$

which implies the upper bound in Theorem 1.

2.2 Lower bound

2.2.1 Preliminaries

The proof of the lower bound in Theorem 1 relies on the celebrated flag algebra method introduced by Razborov [20]. Here we briefly discuss the main idea behind this approach.

Let $(G_n)_{n\in\mathbb{N}}$ be a sequence of graphs, such that order of G_n increases. Such a sequence is called *convergent* if for every fixed graph H, the density of H in G_n converges, i.e., for every H there exists some number $\phi(H)$, such that

$$\lim_{n\to\infty} p(H, G_n) = \phi(H),$$

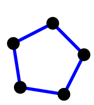
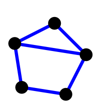
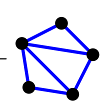
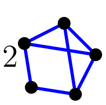
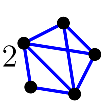
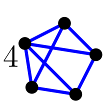
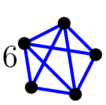
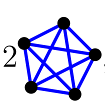
where $p(H, G)$ is the probability that $|H|$ vertices chosen uniformly at random from $V(G)$ induce a copy of H (simply containing a copy of H is not enough, it must be induced here). Notice that any sequence of graphs that increases in size has a convergent subsequence G_{n_i}. Thus, without loss of generality we assume G_n is convergent. Note that ϕ cannot be an arbitrary function since it must satisfy many obvious identities such as $\phi(edge) + \phi(nonedge) = 1$.

Interestingly, these ϕ exactly correspond to homomorphisms that we now describe. Denote by $\mathcal{F}$ the set of all graphs and by $\mathcal{F}_\ell$ the set of graphs of size ℓ. Let $\mathbb{R}\mathcal{F}$ be the set of all finite formal linear combinations of graphs in $\mathcal{F}$ with real coefficients. It comes with the natural operations of addition and multiplication by a real number. Let $\mathcal{K}$ be a linear subspace generated by all linear combinations

$$F - \sum_{H\in\mathcal{F}_\ell} p(F, H) \cdot H, \tag{1}$$

where $\ell > |V(F)|$. Notice that ϕ evaluated at any element of $\mathcal{K}$ gives 0. Finally, let $\mathcal{A}$ be $\mathbb{R}\mathcal{F}$ factorized by $\mathcal{K}$. It is possible to define multiplication on $\mathcal{A}$, which we do in Section 2.2.3. It can be proved that $\mathcal{A}$ is indeed an algebra. Now limits of convergent graph sequences correspond to homomorphism ϕ from $\mathcal{A}$ to $\mathbb{R}$ such that $\phi(F) \geq 0$ for all $F \in \mathcal{F}$. Denote the set of all such homomorphisms by $Hom^+(\mathcal{A}, \mathbb{R})$.

Let OPT be the following linear combination, which counts the C_5 copies using induced subgraphs :

$OPT =$  $+$  $+$  $+ 2$  $+ 2$  $+ 4$  $+ 6$  $+ 12$ ,

where the coefficient of each graph is the number of copies of C_5 it contains. Thus,

$$\phi(OPT) = 120 \lim_{n\to\infty} d_{C_5}(G_n).$$

The factor 120 comes from the fact that $p(C_5, G_n)$ is the probability that 5 vertices chosen uniformly at random from $V(G_n)$ induce a copy of C_5. So we have $\binom{n}{5} \approx \frac{n^5}{120}$ choices. Notice that OPT is written as a linear combination of all 34 graphs on 5-vertices, but 26 of them have coefficient 0. For short, we will write it as

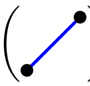
$$OPT = \sum_{F \in \mathcal{F}_5} c_F^{OPT} F, \tag{2}$$

where nonzero entries of c_F^{OPT} are above.

Our goal is to find a good lower bound on

$$\min_{\phi \in Hom^+(\mathcal{A},\mathbb{R})} \phi(OPT).$$

If we know that the density of edges is at least p, we say that we consider only ϕ that satisfies this additional constraint:

ϕ () $\geq p$ (3)

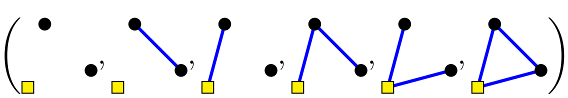
A particular instance of (1) is

$$\phi(K_2) = \phi\left(\sum_{F \in \mathcal{F}_5} p(K_2, F) \cdot F\right).$$

Our next goal is to find a suitable $A \in \mathcal{A}$, such that $\phi(A) \geq 0$ for all $\phi \in Hom^+(\mathcal{A},\mathbb{R})$ and use it in calculations. In particular, we will use it as $\phi(OPT) \geq \phi(OPT) - \phi(A) = \phi(OPT - A) \geq c$, where c is some resulting number. Recall that A can be represented as a linear combinations of graphs. Moreover, A may contain both positive and negative coefficients and these coefficients may combine with coefficients in OPT and make the resulting lower bound c more obvious.

It is possible to find such A by considering graphs with a few labeled vertices. In our case, we only need one labeled vertex. Similarly to defining the algebra $\mathcal{A}$ and limits of convergent graph sequences, one can define limits of graph sequences, where every graph has exactly one labeled vertex. This gives an algebra $\mathcal{A}^1$ and homomorphisms $Hom^+(\mathcal{A}^1,\mathbb{R})$. In the following, we depict the labeled vertex by a square.

Let X be the following vector

$X =$  ^{T}.

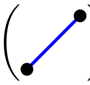
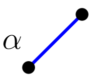
Notice that X is the vector of all graphs on 3 vertices with exactly one labeled vertex (the yellow square). For isomorphism, the labeled vertex must be preserved but the remaining vertices may be swapped. If M is a positive semidefinite matrix in $\mathbb{R}^{6\times 6}$, then for every $\phi^1 \in Hom^+(\mathcal{A}^1,\mathbb{R})$ holds

$$0 \leq \phi^1\left(X^T M X\right).$$

This can be seen since ϕ^1 is a homomorphism, so equivalent would be $0 \leq \phi^1\left(X^T\right) M \phi^1(X)$, where by $\phi^1(X)$ we mean application of ϕ^1 to each coordinate of X. In this case, there exists an unlabeling (i.e. averaging) linear operator, $[\![\cdot]\!]_1$, such that we get rid of the labeled vertex and get a linear combination of unlabeled graphs, such that for all $\phi \in Hom^+(\mathcal{A}, \mathbb{R})$

$$0 \leq \phi\left([\![X^T M X]\!]_1\right) \qquad\qquad [\![X^T M X]\!]_1 = \sum_{F \in \mathcal{F}_5} c_F^M \cdot F. \tag{4}$$

In Section 2.2.3 we will explain precisely how to calculate coefficients c_F^M. Next we take the sum of equations (2), (3), and (4), where α is any nonnegative constant and

$$\phi(OPT) \geq \phi(OPT) + \alpha \left(p - \phi\left(\right.\right.$$

$$\left.\left.\right)\right) - \phi\left([\![X^T M X]\!]_1\right)$$
$$= \phi\left(OPT + \alpha p - \alpha\right.$$

$$\left. - [\![X^T M X]\!]_1\right)$$
$$= \phi\left(\sum_{F \in \mathcal{F}_5} \left(c_F^{OPT} + \alpha p - \alpha \cdot p(K_2, F) - c_F^M\right) \cdot F\right)$$

(In Appendix A we provide c_F^{OPT} and $p(K_2, F)$ for each $F \in \mathcal{F}_5$.) For simplicity, we use for every F

$$c_F = \left(c_F^{OPT} + \alpha p - \alpha \cdot p(K_2, F) - c_F^M\right).$$

With this notation

$$\phi(OPT) \geq \phi\left(\sum_{F \in \mathcal{F}_5} c_F \cdot F\right) \geq \min_{F \in \mathcal{F}_5} c_F \cdot \phi\left(\sum_{F \in \mathcal{F}_5} F\right) = \min_{F \in \mathcal{F}_5} c_F, \tag{5}$$

where c_F is a number that depends on the choice of M and α. We can optimize M and α to maximize $\min_{F \in \mathcal{F}_5} c_F$. This will give a lower bound on $120 d_{C_5}(p)$.

2.2.2 Finding the optimum

We will slightly modify this general setup. Let $k \geq 3$ be fixed such that $p = 1 - 1/k$. Then, $\phi(K_2) \geq p$ and so, in particular, we have

$$0 \geq 10(k-1) - 10k\phi(K_2).$$

Thus,

$$\begin{aligned}
\phi(OPT) &\geq \phi(OPT) + \alpha\left(10(k-1) - 10k\phi(K_2)\right) - \phi\left([\![X^T M X]\!]_1\right) \\
&= \phi\left(OPT + 10(k-1)\alpha - 10k\alpha \cdot K_2 - [\![X^T M X]\!]_1\right) \\
&= \phi\left(\sum_{F \in \mathcal{F}_5} \left(c_F^{OPT} + 10(k-1)\alpha - 10k\alpha \cdot p(K_2, F) - c_F^M\right) F\right)
\end{aligned}$$

and so now

$$c_F = c_F^{OPT} + 10(k-1)\alpha - 10k\alpha \cdot p(K_2, F) - c_F^M.$$

Now we show that for some certain M and α, $c_F \geq 12 - 60/k + 120/k^2 - 120/k^3 + 48/k^4$ for any $F \in \mathcal{F}_5$ yielding

$$\phi(OPT) \geq 12 - \frac{60}{k} + \frac{120}{k^2} - \frac{120}{k^3} + \frac{48}{k^4}.$$

Let

$$\alpha = \frac{1}{5k^4}\left(30k^3 - 120k^2 + 180k - 96\right).$$

It is easy to check that $\alpha > 0$ for any $k \geq 3$. In order to define matrix M we define first two matrices A and B as follows:

$$A = \begin{pmatrix} 32k^2 - 96k + 96 & 0 & 4k^2 - 16k \\ 0 & 10k^4 - 30k^3 - 8k^2 + 96k - 96 & -10k^4 + 35k^3 - 4k^2 - 80k + 96 \\ 4k^2 - 16k & -10k^4 + 35k^3 - 4k^2 - 80k + 96 & 10k^4 - 40k^3 + 24k^2 + 64k - 96 \end{pmatrix}$$

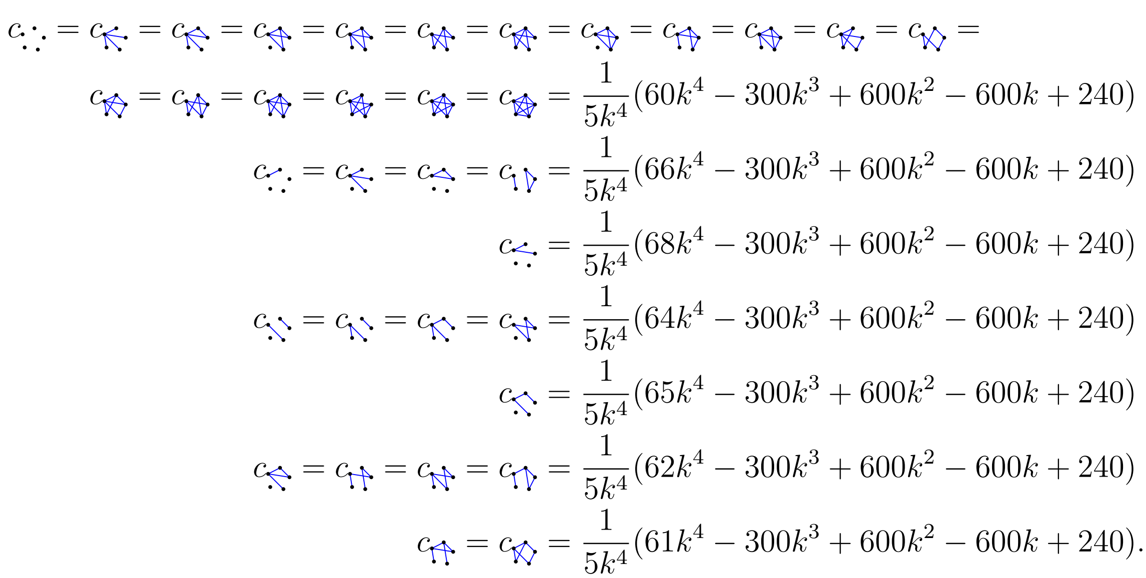
and

$$B = \begin{pmatrix} k-1 & 1 & k-2 & 0 & k-3 & -1 \\ 0 & 2 & k-2 & 0 & 2k-4 & -2 \\ 0 & 0 & k-1 & -1 & 2k-2 & -2 \end{pmatrix}.$$

It is easy to verify (by checking principal minors) that A is positive definite for any $k \geq 3$. Therefore, matrix

$$M = \frac{3}{2k^4} B^T A B$$

is positive semidefinite. In Section 2.2.4 we briefly describe how we determined matrices A and B. With this choice of M and α one can verify using for example Maple (see Appendix B) that coefficients c_F satisfy:



$$c = c = c = c = c = c = c = c = c = c = c = c =$$
$$c = c = c = c = c = c = \frac{1}{5k^4}(60k^4 - 300k^3 + 600k^2 - 600k + 240)$$
$$c = c = c = c = \frac{1}{5k^4}(66k^4 - 300k^3 + 600k^2 - 600k + 240)$$
$$c = \frac{1}{5k^4}(68k^4 - 300k^3 + 600k^2 - 600k + 240)$$
$$c = c = c = c = \frac{1}{5k^4}(64k^4 - 300k^3 + 600k^2 - 600k + 240)$$
$$c = \frac{1}{5k^4}(65k^4 - 300k^3 + 600k^2 - 600k + 240)$$
$$c = c = c = c = \frac{1}{5k^4}(62k^4 - 300k^3 + 600k^2 - 600k + 240)$$
$$c = c = \frac{1}{5k^4}(61k^4 - 300k^3 + 600k^2 - 600k + 240).$$

Since the entries only ever disagree in the k^4 coefficient we get that

$$\begin{aligned}\phi(OPT) &\geq \min_{F \in \mathcal{F}_5} c_F \\ &= \frac{1}{5k^4}(60k^4 - 300k^3 + 600k^2 - 600k + 240) = 12 - \frac{60}{k} + \frac{120}{k^2} - \frac{120}{k^3} + \frac{48}{k^4}.\end{aligned}$$

2.2.3 Products of graphs and determining c_F^M coefficients

First, we define product of unlabeled graphs. For a graph G, denote $|V(G)|$ by $|G|$. Let F_1, F_2, F in $\mathcal{F}$ such that $|F_1| + |F_2| \leq |F|$. Choose uniformly at random two disjoint subsets X_1 and X_2 of $V(F)$ of sizes $|F_1|$ and $|F_2|$, respectively. Denote by $p(F_1, F_2; F)$ the probability that $F[X_1]$ is isomorphic to F_1 and $F[X_2]$ is isomorphic to F_2. Finally, the product of F_1 and F_2 is defined as

$$F_1 \times F_2 = \sum_{F \in \mathcal{F}_{|F_1|+|F_2|}} p(F_1, F_2; F) \cdot F.$$

The product can be extended to linear combinations of graphs and gives a multiplication operation in $\mathcal{A}$.

The product in $\mathcal{A}^1$ is defined along the same lines as in $\mathcal{A}$ but the intersection of X_1 and X_2 is exactly the labeled vertex. A more precise definition follows. Let F_1, F_2, F in $\mathcal{F}^1$ such that $|F_1| + |F_2| \leq |F| - 1$. Choose uniformly at random subsets X_1 and X_2 of $V(F)$ of sizes $|F_1|$ and $|F_2|$, respectively whose intersection is exactly the one labeled vertex. Denote by $p(F_1, F_2; F)$ the probability that $F[X_1]$ is isomorphic to F_1 and $F[X_2]$ is isomorphic to F_2, where isomorphism preserves the labeled vertex. Finally, the product of F_1 and F_2 is defined as

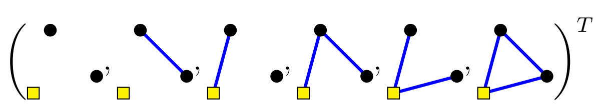
$$F_1 \times F_2 = \sum_{F \in \mathcal{F}_{|F_1|+|F_2|-1}} p(F_1, F_2; F) \cdot F.$$

Next we define the unlabeling operator $[\![\cdot]\!]_1 : \mathcal{F}^1 \to \mathbb{R}\mathcal{F}$. We extend $[\![\cdot]\!]_1$ to a linear function $\mathbb{R}\mathcal{F}^1 \to \mathbb{R}\mathcal{F}$ which we also call $[\![\cdot]\!]_1$. Let $F \in \mathcal{F}^1$. Denote by $G \in \mathcal{F}$ the graph obtained from F by unlabeling the labeled vertex. Let v be a vertex in G chosen uniformly at random. Let q be the probability, that G with labeled v is isomorphic to F. Then

$$[\![F]\!]_1 = q \cdot G.$$

Recall that X is the vector of all 3 vertex types.

$$X = (X_1, X_2, X_3, X_4, X_5, X_6)^T =$$



In Appendix A we list all coefficients for products in $\mathcal{F}_3^1$, after unlabeling and multiplying by a scaling factor of 30 to clear denominators. Then we obtain that

$$[\![X^T M X]\!]_1 = \sum_{i=1}^{6} \sum_{j=1}^{6} M_{i,j} [\![X_i \times X_j]\!]_1 = \sum_{F \in \mathcal{F}_5} c_F^M \cdot F,$$

since each $[\![X_i \times X_j]\!]_1$ is a linear combination of graphs in $\mathcal{F}_5$.

2.2.4 Guessing matrices A and B

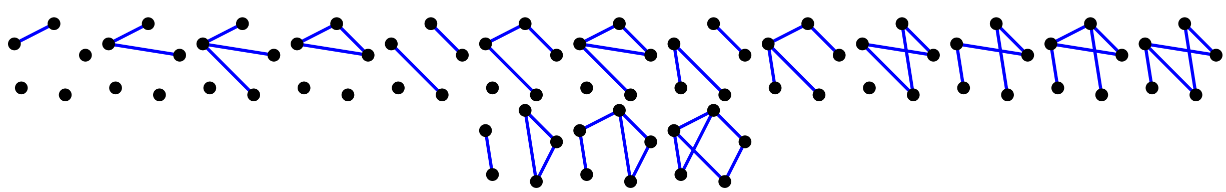
In this paragraph we describe how we obtained matrices A and B. First, we used semidefinite programming to find matrices say M for several small odd values of k. Notice that if (5) is applied to the extremal construction, then the left-hand side is equal to the right-hand side. That means that all inequalities used are actually equalities. In particular, $\phi\left(\llbracket X^T M X \rrbracket_1\right) = 0$. Since M is a positive semidefinite matrix, X evaluated on the extremal example must give an eigenvector of M corresponding to the eigenvalue 0. The matrix B was obtained by projecting onto the space orthogonal to three zero eigenvectors of M. As noted before, we had one zero eigenvector to start with. By looking at all eigenvectors of M, we managed to guess another zero eigenvector. We tried projection with the two zero eigenvectors and found the third one in the projection. After having obtained matrices B, we observed that a suitable A exists even if we set the coordinate $[1, 2]$ and $[2, 1]$ to 0. With proper scaling of the objective function, we were getting nice matrices from the CSDP [3] solver with all entries integers. By using the solutions for several values of k, we calculated a polynomial function of k fitting each entry in matrix A. Finally we observed that the same matrices A and B also work for even values of k.

3 Stability

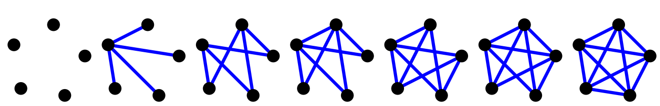
We believe that the complete k-partite graph is the one that minimize the number of C_5's for $p = 1 - \frac{1}{k}$. In general we were unable to prove it but we observed the following. Notice that if we have an extremal construction, then from the very first part of (5),

$$12 - \frac{60}{k} + \frac{120}{k^2} - \frac{120}{k^3} + \frac{48}{k^4} \geq \phi(OPT) \geq \min_{F \in \mathcal{F}_5} c_F = 12 - \frac{60}{k} + \frac{120}{k^2} - \frac{120}{k^3} + \frac{48}{k^4}$$

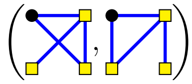
we observe that if $c_F > \min_{H \in \mathcal{F}_5} c_H$, then $\phi(F) = 0$. Otherwise, we get a contradiction. In this way, we obtain a list of forbidden graphs for extremal constructions in the sense that their density must be zero in the limit. Thus, the following graphs have zero density in the limit:



As a matter of fact for $2 \leq k \leq 73$ one can show by using flags with more labeled vertices that the only possible graphs with nonzero density must belong to the following list $\mathcal{L}$:



We perform a calculation analogous to the previous calculation. The main difference is that we include $[\![X_2^T M_2 X_2]\!]$, where M_2 is a positive semidefinite matrix in $\mathbb{R}^2$ and

$X_2 =$  .

For each $k \in \{2, \ldots, 73\}$, we were able to construct particular M and M_2, such that only graphs in $\mathcal{L}$ may have nonzero density. But unlike in the previous case, we were not able to construct M and M_2 as functions of k. Certificates for the flag algebra calculations are available at [10]. For convenience, we restate here the statement of Theorem 2.

Theorem 2. *Let G be a graph on n vertices for large n, such that G has edge density $p = 1 - \frac{1}{k}$ for $k \geq 2$ and*

$$d_{C_5}(G) \leq d_{C_5}(p) + \epsilon$$

for some positive but sufficiently small ϵ. Assume further that the only induced subgraphs on five vertices with density more than ϵ are the graphs in list $\mathcal{L}$ (we know that this assumption holds for $2 \leq k \leq 73$). Then G has edit distance at most δn^2 from the Turán graph T_n^k, for some function $\delta = \delta(\epsilon) \to 0$ as $\epsilon \to 0$.

For the proof of Theorem 2 we will use the following lemma:

Lemma 3. *Suppose a graph J on n vertices has a subgraph X such that*

(i) X has x vertices where $\epsilon' n \leq x \leq (1-\epsilon')n$ and edge density $q \leq \frac{1}{2}$

(ii) X is complete to $V(J) \setminus X$

(iii) X contains at least $\frac{1}{2}x^4 q^3 + \epsilon' x^4$ copies of P_4.

Then there exists a graph J' on n vertices with asymptotically the same edge density as J and

$$d_{C_5}(J') \leq d_{C_5}(J) - \frac{1}{2}(\epsilon')^6.$$

Proof of Lemma. Note first that conditions (i) and (ii) imply that J is dense since it has at least $\epsilon'(1-\epsilon')n^2 edges$. We make J' by replacing X with a X', which is a random balanced bipartite graph with edge probability $2q$. We will not change the rest of the graph, so $J' - X' = J - X$. W.h.p. X' has edge density asymptotically q and so J' has asymptotically the same edge density as J. We will argue that J' has much fewer copies of C_5 than J has, by considering several possible types of C_5 copies.

We will compare the copies according to how they intersect X (for counting copies of C_5 in the graph J) or X' (in J'). Specifically, since X is complete to the rest of J we have

$$\nu_{C_5}(J) = \sum_H m_H \nu_H(X) \cdot \nu_{C_5 - H}(J - X)$$

where the sum is over all induced subgraphs $H \subseteq C_5$, and the coefficient m_H is the number of C_5 copies contained in the graph formed by taking a copy of H and a copy of $C_5 - H$ with every possible edge in between. Recall that $\nu_H(G)$ counts the number of (not necessarily induced) copies of H in G. Similarly, we have

$$\nu_{C_5}(J') = \sum_H m_H \nu_H(X') \cdot \nu_{C_5-H}(J' - X') = \sum_H m_H \nu_H(X') \cdot \nu_{C_5-H}(J - X),$$

since $J' - X' = J - X$. So we will compare $\nu_H(X)$ with $\nu_H(X')$ for each H. Specifically we will show that $\nu_H(X') \leq (1 + o(1))\nu_H(X)$ for each H, and that this inequality holds with some room for $H = P_4$.

Some easy cases: when H has no vertices, $\nu_H(X) = \nu_H(X') = 1$. When H is a single vertex, $\nu_H(X) = \nu_H(X') = x$. When H is just an edge, $\nu_H(X) = (1 + o(1))\nu_H(X') = (1 + o(1))\binom{x}{2}q$. When H has 2 vertices and no edge we have $\nu_H(X') = \nu_H(X) = \binom{x}{2}$. When H is the graph on 3 vertices consisting of an edge and an isolated vertex, we have $\nu_H(X') = (1 + o(1))\nu_H(X) = (1 + o(1))x\binom{x}{2}q$.

When $H = P_3$ (the path of length 2) we have

$$\nu_{P_3}(X') = 2\binom{\frac{x}{2}}{2}\frac{x}{2}(2q)^2 = (1 + o(1))\frac{1}{2}x^3q^2$$

which we compare to

$$\nu_{P_3}(X) = \sum_{v \in X} \binom{|N(v) \cap X|}{2} \geq x \cdot \binom{\frac{2q\binom{x}{2}}{x}}{2} = (1 + o(1))\frac{1}{2}x^3q^2.$$

Finally we consider the case $H = P_4$. We have

$$\nu_{P_4}(X') = 2\binom{\frac{x}{2}}{2} \cdot 2\binom{\frac{x}{2}}{2}(2q)^3 = (1 + o(1))\frac{1}{2}x^4q^3$$

which we compare to

$$\nu_{P_4}(X) = \frac{1}{2}x^4q^3 + \epsilon' x^4.$$

Taking all possible H into account, we see that

$$\begin{aligned}
\nu_{C_5}(J) - \nu_{C_5}(J') &= \sum_H [\nu_H(X) - \nu_H(X')] \cdot \nu_{C_5-H}(J - X) \\
&\geq [\nu_{P_4}(X) - \nu_{P_4}(X')] \cdot \nu_{C_5-P_4}(J - X) \\
&\geq (1 + o(1))\epsilon' x^4 \cdot (n - x) \\
&> \frac{1}{2}(\epsilon')^6 n^5
\end{aligned}$$

and so

$$d_{C_5}(J') \leq d_{C_5}(J) - \frac{1}{2}(\epsilon')^6.$$

□

Proof of Theorem 2. By the induced graph removal lemma (see, e.g., [1, 4]) we can eliminate all induced subgraphs of G that are in $\mathcal{L}^c$ by adding or removing at most αn^2 edges, for some $\alpha = \alpha(\epsilon) \to 0$ as $\epsilon \to 0$. Call this new graph G', which has edge density p', where $p - 2\alpha \le p' \le p + 2\alpha$. Notice that G' has no triple inducing exactly one edge, since we have removed all 5-vertex subgraphs that contain any such triples. Now it is easy to see that G' is a complete k'-partite graph for some k'. Say the parts of G' are $X_1, \ldots, X_{k'}$. Also, note that since adding (or removing) one edge to G creates (or destroys) at most n^3 copies of C_5, we have

$$d_{C_5}(G) = d_{C_5}(G') + O(\alpha),$$

and

$$d_{C_5}(p) = d_{C_5}(p') + O(\alpha)$$

(recall that we use big-O notation to replace quantities that are bounded in absolute value, and the quantity being replaced may be negative). Now

$$d_{C_5}(G') \le d_{C_5}(G) + O(\alpha) \le d_{C_5}(p) + \epsilon + O(\alpha) \le d_{C_5}(p') + O(\epsilon + \alpha) \tag{6}$$

and so G' has nearly the minimum C_5-density among graphs with edge density p'.

In the following, we will need a parameter $\beta = \beta(\epsilon) = (\epsilon + \alpha(\epsilon))^{1/100}$.

Claim 4. *We are done unless we have the following. For any $i \neq j$, $|X_i| + |X_j| \le (1 - \beta)n$.*

Proof. WLOG, suppose for contradiction that $|X_1| + |X_2| \ge (1-\beta)n$, so the number of edges in G' is at most

$$\binom{n}{2} - \binom{|X_1|}{2} - \binom{|X_2|}{2} \le \binom{n}{2} - 2\binom{\frac{(1-\beta)n}{2}}{2} \le \frac{1}{2}n^2 - \frac{1}{4}(1-\beta)^2 n^2 = \left(\frac{1}{4} + O(\beta)\right) n^2$$

and so we must have $k = 2$ since throughout the proof we assume ϵ (and therefore α and β) are sufficiently small. Now if $||X_1| - |X_2|| > \beta^{1/3} n$, say WLOG $X_1 > X_2 + \beta^{1/3} n$ then the number of edges in G' is at most

$$\begin{aligned} |X_1||X_2| + \beta n(|X_1| + |X_2|) + \binom{\beta n}{2} &\le \left(\frac{n}{2} + \frac{1}{2}\beta^{1/3} n\right)\left(\frac{n}{2} - \frac{1}{2}\beta^{1/3} n\right) + \beta n^2 + \binom{\beta n}{2} \\ &= \left(\frac{1}{4} - \frac{1}{4}\beta^{2/3} + O(\beta)\right) n^2, \end{aligned}$$

which is a contradiction for small ϵ since G' has at least $\binom{n}{2}p - \alpha n^2$ edges (where $p = \frac{1}{2}$ since $k = 2$) and $\frac{1}{4}\beta^{2/3} + O(\beta) > \alpha$ for small ϵ. To summarize, G' is a complete partite graph that has two large parts X_1, X_2 which differ in size by at most $\beta^{1/3} n$, and together the rest of the parts make up at most βn vertices. It is easy to see then that G' can be changed into a balanced complete bipartite graph by editing $O(\beta^{1/3} n^2)$ edges. □

Thus we henceforth assume that for any $i \neq j$, $|X_i| + |X_j| \le (1 - \beta)n$.

Claim 5. *For all i, j, if $|X_i|, |X_j| \ge \beta n$, then $||X_i| - |X_j|| \le \beta n$.*

Proof. Suppose for contradiction that there are two parts (WLOG say X_1, X_2) such that $|X_1|, |X_2| \geq \beta n$ and $||X_1| - |X_2|| > \beta n$. We will derive a contradiction by arguing that G' can be modified by Lemma 3 to form another graph G^* of asymptotically the same edge density but with significantly smaller C_5-density than G'.

We apply Lemma 3 with $J = G'$, $X = X_1 \cup X_2$, $\epsilon' = \frac{1}{2}\beta^6$ and

$$q = \frac{x_1 x_2}{\binom{x}{2}} = (1 + o(1))\frac{2x_1 x_2}{x^2}$$

where $|X_i| = x_i$ and $x = x_1 + x_2$. Let us check the conditions of the lemma. Clearly we have

$$\beta n \leq x \leq (1 - \beta)n,$$

and X is complete to the rest of the graph (since X is composed of two parts of a complete partite graph). Finally, the number of copies of P_4 in X is

$$\nu_{P_4}(X) = 2\binom{x_1}{2} \cdot 2\binom{x_2}{2} = (1 + o(1))x_1^2 x_2^2$$

which we compare to

$$\frac{1}{2}x^4 q^3 = (1 + o(1))\frac{1}{2}x^4\left(\frac{2x_1 x_2}{x^2}\right)^3 = (1 + o(1))\frac{4x_1^3 x_2^3}{x^2}.$$

From here we can see that

$$\begin{aligned}
\nu_{P_4}(X) - \frac{1}{2}x^4 q^3 &\geq (1 + o(1))\left(x_1^2 x_2^2 - \frac{4x_1^3 x_2^3}{x^2}\right) \\
&\geq \frac{1}{2} \cdot \frac{x_1^2 x_2^2}{x^2}(x^2 - 4x_1 x_2) \\
&= \frac{1}{2} \cdot \frac{x_1^2 x_2^2}{x^2}(x_1 - x_2)^2 \\
&\geq \frac{1}{2}\frac{(\beta n)^4}{n^2}(\beta n)^2 = \frac{1}{2}\beta^6 n^4 \geq \frac{1}{2}\beta^6 x^4
\end{aligned}$$

and so Lemma 3 applies, implying that $J = G'$ must have C_5-density at least

$$d_{C_5}(p') + \frac{1}{2}\left(\frac{1}{2}\beta^6\right)^6 = d_{C_5}(p') + \frac{1}{128}\beta^{36}.$$

But then from (6), we have

$$d_{C_5}(p') + \frac{1}{128}\beta^{36} \leq d_{C_5}(G') \leq d_{C_5}(p') + O(\epsilon + \alpha),$$

a contradiction for small ϵ since $\beta = (\epsilon + \alpha)^{1/100}$. $\square$

WLOG say that $|X_1|, \ldots, |X_\ell| \geq \beta n$ and $|X_i| < \beta n$ for any $i > \ell$. By Claim 5, there is some value x such that $|X_i| \in [(x-\beta)n, (x+\beta)n]$ for $1 \leq i \leq \ell$. Then the number of edges in G' is at most

$$\begin{aligned}\binom{n}{2} - \sum \binom{|X_i|}{2} &\leq \binom{n}{2} - \ell \binom{(x-\beta)n}{2} \\ &= \frac{1}{2} n^2 (1 - \ell x^2 + O(\beta)).\end{aligned}$$

We will now show a lower bound matching the above upper bound. Since for any numbers $a \geq b$ and $\delta > 0$, we have $(a+\delta)^2 + (b-\delta)^2 > a^2 + b^2$ the following holds. Since $\sum_{i>\ell} |X_i| \leq n$, and for $i > \ell$ we have $|X_i| \leq \beta n$, the maximum possible value of $\sum_{i>\ell} |X_i|^2$ occurs when all the terms are either 0 or $(\beta n)^2$, meaning that the number of positive terms would be at most $\frac{1}{\beta}$, so we have

$$\sum_{i>\ell} |X_i|^2 \leq \frac{1}{\beta} \cdot (\beta n)^2 = \beta n^2$$

the number of edges in G' is then at least

$$\begin{aligned}\binom{n}{2} - \sum \binom{|X_i|}{2} &\geq \binom{n}{2} - \ell \binom{(x+\beta)n}{2} - \frac{1}{2}\beta n^2 \\ &= \frac{1}{2} n^2 (1 - \ell x^2 + O(\beta)).\end{aligned}$$

But we know G' has edge density $p' = 1 - \frac{1}{k} + O(\alpha) = 1 - \ell x^2 + O(\beta)$ and so we get

$$x = \frac{1}{\sqrt{k\ell}} + O(\beta)$$

and in particular $\ell \leq k$ since otherwise $|X_1| + \ldots + |X_\ell| \geq (\ell x + O(\beta))n > n$. To summarize, at this point we know that the graph must have $\ell \leq k$ "large" parts which each have about $\frac{1}{\sqrt{k\ell}} n$ vertices, and the rest of the parts are "small" and each have at most βn vertices. We would like to show that $\ell = k$, so assume for contradiction that $\ell < k$.

Claim 6. *$\sum_{i>\ell} |X_i| > \beta n$.*

Proof. Observe that

$$\sum_{i>\ell} |X_i| = n - \sum_{i\leq\ell} |X_i| = n - \ell \left(\frac{1}{\sqrt{k\ell}} + O(\beta) \right) n = \left(1 - \frac{\sqrt{\ell}}{\sqrt{k}} + O(\beta) \right) n > \beta n$$

since $\ell < k$ and we may assume $\beta > 0$ is arbitrarily small. □

Now we will use Lemma 3 on $J = G'$ and X being X_1 together with several of the small X_is, which will finish the proof. Recall we have $|X_1|$ of size $\left(\frac{1}{\sqrt{kl}} + O(\beta)\right) n$. We know $|X_i| < \beta n$ for all $i > \ell$ and at the same time $|\cup_{i>\ell} X_i| > \beta n$. Hence there exists an integer

z such that $\beta n \leq |\cup_{z\geq i>\ell} X_i| \leq 2\beta n$. Let $Y = \cup_{z\geq i>\ell} X_i$. In order to apply Lemma 3 to $X = X_1 \cup Y$, we need to count the number of copies of P_4 in X, the other assumptions of Lemma 3 are clearly satisfied. Notice that $\nu_{P_4}(X)$ is bounded from below by the number of copies of P_4 that alternate vertices in X_1 and in Y, which gives

$$\nu_{P_4}(X) \geq |X_1|^2|Y|^2 \geq |X_1|^2(\beta n)^2 = \frac{\beta^2}{kl}n^4 + O(\beta^3)n^4. \tag{7}$$

Denote $|X|$ by x. Notice that

$$x = |X_1| + |Y| = \left(\frac{1}{\sqrt{kl}} + O(\beta)\right) n.$$

Let e be the number of edges in X. It can be bounded from above by pretending that Y is a complete graph, which gives

$$e \leq |X_1| \cdot |Y| + |Y|^2/2 \leq \frac{2\beta n^2}{\sqrt{kl}} + O(\beta^2)n^2.$$

This gives

$$q = \frac{2e}{x^2} \leq 4\beta\sqrt{kl} + O(\beta^2).$$

Hence X satisfies of Lemma 3*(iii)* with $\epsilon' = \frac{\beta^2 kl}{2}$, since

$$\frac{1}{2}x^4q^3 \leq \frac{32\beta^3}{\sqrt{kl}}n^4 + O(\beta^4)n^4$$

is significantly smaller than $\nu_{P_4}(X)$ (see (7)) and $\epsilon' x^4 \leq \frac{\beta^2}{2kl}n^4 + O(\beta^4)n^4$ is about $\frac{1}{2}\nu_{P_4}(X)$. Hence Lemma 3 implies

$$d_{C_5}(G') \geq d_{C_5}(p') + \frac{\beta^{12}(kl)^6}{2^7} > d_{C_5}(p') + \beta^{19}.$$

Combining this with (6) gives the final contradiction

$$d_{C_5}(p') + \beta^{19} \leq d_{C_5}(G') \leq d_{C_5}(p') + O(\epsilon + \alpha)$$

for a small ϵ since $\beta = (\epsilon + \alpha)^{1/100}$. □

Summarizing, we just showed that G can be transformed into the Turán graph T_n^k by adding or deleting at most $o(n^2)$ edges. Unfortunately, our stability result hinges on the list $\mathcal{L}$ containing the only graphs of nonzero density, which we were not able to prove for arbitrary k.

4 Remarks on the case $p \neq 1 - \frac{1}{k}$

Our general upper bound construction is as follows. Suppose that p is a constant satisfying $1 - \frac{1}{k} < p < 1 - \frac{1}{k+1}$. Partition the vertices into $k-1$ sets $X_1, \dots, X_{k-1}$ of size xn and one more set Y of size yn. Each X_i is an independent set. For $1 \leq i \neq j \leq k-1$ we have that X_i is complete to X_j. Finally, $G[Y]$ is any graph such that for some parameter $0 < \rho < \frac{1}{2}$ we have

(i) $G[Y]$ has asymptotically $\frac{1}{2}y^2n^2\rho$ edges, $\frac{1}{2}y^3n^3\rho^2$ paths of length 2 (that means on 3 vertices), and $\frac{1}{2}y^4n^4\rho^3$ paths of length 3;

(ii) $G[Y]$ has $o(n^5)$ copies of C_5.

(See the end of this subsection for discussion on which graphs are suitable for $G[Y]$). We assume that

$$(k-1)x + y = 1$$

so we have n vertices total. The edge density in this construction is

$$\frac{\binom{k-1}{2}(xn)^2 + (k-1)(xn)(yn) + (\frac{1}{2} + o(1))y^2n^2\rho}{\binom{n}{2}},$$

which tends to

$$g(x, y, \rho) = (k-1)_2x^2 + 2(k-1)xy + \rho y^2$$

as $n \to \infty$. So we also assume that the parameters x, y, ρ satisfy $g(x, y, \rho) = p$.

Now we consider the ratio $f(x, y, \rho) = \lim_{n\to\infty} \frac{\nu_G(C_5)}{n^5}$. We claim that

$$\begin{aligned} f(x, y, \rho) = &\left[\frac{1}{10}(k-1)_5 + \frac{1}{2}(k-1)_4 + \frac{1}{2}(k-1)_3\right] x^5 \\ &+ \left[\frac{1}{2}(k-1)_4 + \frac{3}{2}(k-1)_3 + \frac{1}{2}(k-1)_2\right] x^4y \\ &+ \left[\left(\frac{1}{2} + \frac{1}{2}\rho\right)(k-1)_3 + \left(1 + \frac{1}{2}\rho\right)(k-1)_2\right] x^3y^2 \\ &+ \left[\left(\frac{1}{2}\rho + \frac{1}{2}\rho^2\right)(k-1)_2 + \frac{1}{2}\rho(k-1)\right] x^2y^3 \\ &+ \frac{1}{2}\rho^3(k-1)xy^4. \end{aligned}$$

Note that we have grouped the terms of $f(x, y, \rho)$ according to powers of x and y, and then according to falling factorials of $(k-1)$. To understand our formula, it helps to think of the powers of x, y as specifying how many vertices come from sets of size xn, yn, and the falling factorial $(k-1)$ as specifying how many distinct sets of size xn are involved. For example, the first term $\frac{1}{10}(k-1)_5\, x^5$ is there because there are $\frac{1}{10}(k-1)_5(xn)^5$ many copies of C_5 having vertices $v_1, \dots, v_5$ all in different parts of size xn. Now let us justify a more complicated term like say the second term in the third line, $\left(1+\frac{1}{2}\rho\right)(k-1)_2\ x^3y^2$. This term counts the copies of C_5 that have vertices $v_1,\ldots,v_5$ such that v_1 and v_2 come from Y, v_3 and v_4 are in the same set of size xn, and v_5 is in some other set of size xn (and $v_1,\ldots,v_5$ may be in any order on the cycle). The case where v_1 and v_2 are consecutive in the cycle contributes $\frac{1}{2}(k-1)_2\rho(yn)^2(xn)^3$, and the other case contributes $(k-1)_2(yn)^2(xn)^3$.

Now for a given integer $k\geq 2$ and a real number $1-\frac{1}{k}<p<1-\frac{1}{k+1}$ we define an optimization problem (P):

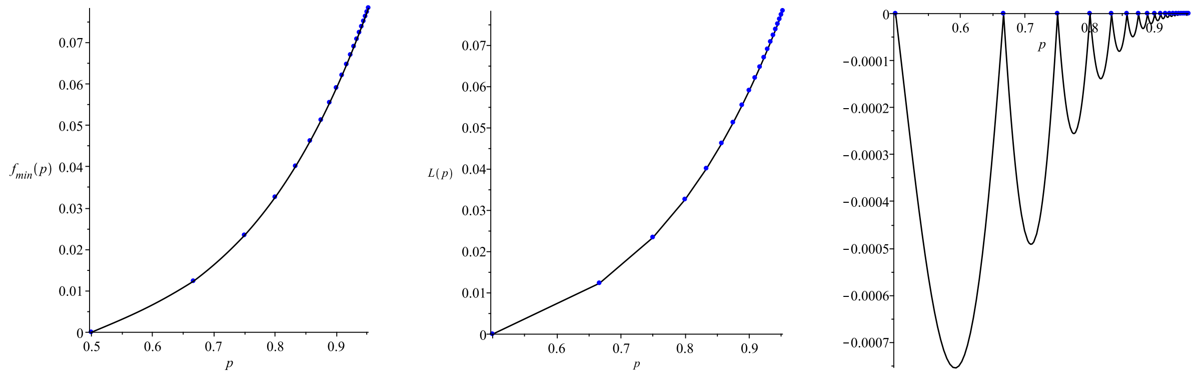
$$\begin{aligned} \texttt{Minimize}\quad & f(x,y,\rho)\\ \texttt{subject to:}\quad & (k-1)x+y=1,\\ & g(x,y,\rho)=p,\\ & x,y\geq 0. \end{aligned}$$

Let us denote its solution by $f_{min}(p)=f(x_0,y_0,\rho_0)$. Clearly, $d_{C_5}(p)\leq f_{min}(p)$. For some certain values of k and p we verified that $120\cdot f_{min}(p)$ numerically matches the lower bound on $d_{C_5}(p)$ given by the flag algebras. In particular, when we calculated with unlabeled flags of order ℓ, we were getting numerically matching bounds for $p\leq 1-\frac{1}{\ell-2}$ and we observed a gap in the bounds for $p>1-\frac{1}{\ell-2}$ different from Turán densities. Since computer calculations can be performed with current computers in a reasonable time only for $\ell\leq 8$, a simple straightforward use of computer is unlikely to provide a numerical match of $d_{C_5}(p)$ and $f_{min}(p)$ for all p. Unfortunately, we were unable to convert the numerical match to a formal proof. The main problem is that (P) has no closed solution. For example, for $k=2$ and $\frac{1}{2}<p<\frac{2}{3}$ we can plug into the objective function $y=1-x$ and $\rho=(p-x^2-2xy)/y^2$ obtaining

$$f(2,x,1-x,(p-x^2-2xy)/y^2)=\frac{x(2x^2-2x+p)(3x^4-5x^3+(1+4p)x^2+(1-4p)x+p^2)}{2\,(x-1)^2}.$$

Now it is not difficult to show that there exists a local minimum for some $\frac{1}{3}<x<\frac{1}{2}$. Unfortunately, it looks like this minimum can be only found numerically. There might be a different parametrization of the problem that would make it possible to solve (P) and formally show a match with flag algebra calculations for some range of p. On Figure 1 we present the shape of $f_{min}(p)$. We conjecture that $d_{C_5}(p)=f_{min}(p)$ for any p.

We now address what graphs are suitable for $G[Y]$, i.e. what graphs satisfy (i) and (ii). Note first that some such choice of $G[Y]$ exists, for example it can be a random bipartite graph with two parts of size $\frac{1}{2}yn$ and edge probability 2ρ. Now we claim that $G[Y]$ satisfies (i) if and only if $G[Y]$ is *almost $yn\rho$-regular,* or more formally, all but $o(n)$ vertices in $G[Y]$ have degree $(1+o(1))yn\rho$. Indeed, if $G[Y]$ is almost $yn\rho$-regular then it is easy to verify the edge and path counts in (i). Conversely, suppose (i) holds, and let the random variable Z represent the degree of a random vertex in $G[Y]$. Then we have $\mathbb{E}[Z]=(1+o(1))yn\rho$ and



(a) (b) (c)

Figure 1: (a) A graph of $f_{min}(p)$ based on numerical calculations. Blue points correspond to the Turán densities (i.e. $p = 1 - 1/k$). (b) Secant lines between Turán densities. (c) A graph of $f_{min}(p) - L(p)$.

since $\sum_{v \in Y} \binom{deg(v)}{2}$ is the number of paths of length 2 we can calculate

$$\mathbb{E}[Z^2] = \frac{1}{yn} \sum_{v \in V(Y)} deg(v)^2 = \frac{1}{yn} \cdot 2(1+o(1))\frac{1}{2}y^3n^3\rho^2 = (1+o(1))y^2n^2\rho^2 = (1+o(1))\mathbb{E}[Z]^2$$

so Z is concentrated by Chebyshev's inequality (see, e.g., Lemma 20.3 in [7]). In other words, $G[Y]$ is almost $yn\rho$-regular.

We believe that we have described all optimal graphs. Specifically, we believe that any graph with edge density p and C_5-density $d_{C_5}(p) + o(1)$ can be transformed by adding or deleting at most $o(n^2)$ edges into a graph with a vertex partition $X_1, \ldots, X_{k-1}, Y$ where $|X_i| = xn, |Y| = yn$, all X_i are independent, all X_i and Y are complete to each other, and $G[Y]$ is $yn\rho$-regular where x, y, ρ are a solution to the optimization problem (P).

References

[1] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy, *Efficient testing of large graphs*, Combinatorica **20** (2000), no. 4, 451–476.

[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** (2016), no. part A, 47–58.

[3] B. Borchers, *CSDP, a C library for semidefinite programming*, Optimization Methods and Software **11** (1999), no. 1-4, 613–623.

[4] D. Conlon and J. Fox, *Graph removal lemmas*, Surveys in combinatorics 2013, London Math. Soc. Lecture Note Ser., vol. 409, Cambridge Univ. Press, Cambridge, 2013, pp. 1–49.

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

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

[7] A. Frieze and M. Karoński, *Introduction to random graphs*, Cambridge University Press, Cambridge, 2016.

[8] A. Grzesik, *On the maximum number of five-cycles in a triangle-free graph*, J. Combin. Theory Ser. B **102** (2012), no. 5, 1061–1066.

[9] H. Hatami, J. Hladký, D. Král', S. Norine, and A. Razborov, *On the number of pentagons in triangle-free graphs*, J. Combin. Theory Ser. A **120** (2013), no. 3, 722–732.

[10] B. Lidický, `https://orion.math.iastate.edu/lidicky/pub/minC5`.

[11] B. Lidický and F. Pfender, *Pentagons in triangle-free graphs*, `arXiv:1712.08869`.

[12] H. Liu, O. Pikhurko, and K. Staden, *The exact minimum number of triangles in graphs of given order and size*, `arXiv:1712.00633`.

[13] L. Lovász and M. Simonovits, *On the number of complete subgraphs of a graph. II*, Studies in pure mathematics, Birkhäuser, Basel, 1983, pp. 459–495.

[14] W. Mantel, *Problem 28*, Winkundige Opgaven **10** (1907), 60–61.

[15] J. W. Moon and L. Moser, *On a problem of Turán*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **7** (1962), 283–286.

[16] V. Nikiforov, *The number of cliques in graphs of given order and size*, Trans. Amer. Math. Soc. **363** (2011), no. 3, 1599–1618.

[17] E. A. Nordhaus and B. M. Stewart, *Triangles in an ordinary graph*, Canad. J. Math. **15** (1963), 33–41.

[18] O. Pikhurko and A. Razborov, *Asymptotic structure of graphs with the minimum number of triangles*, Combin. Probab. Comput. **26** (2017), no. 1, 138–160.

[19] N. Pippenger and M. C. Golumbic, *The inducibility of graphs*, J. Combinatorial Theory Ser. B **19** (1975), no. 3, 189–203.

[20] A. Razborov, *Flag algebras*, J. Symbolic Logic **72** (2007), no. 4, 1239–1282.

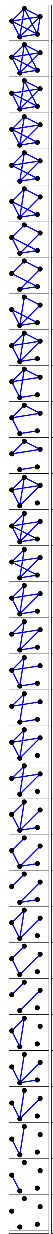
[21] ———, *On the minimal density of triangles in graphs*, Combin. Probab. Comput. **17** (2008), no. 4, 603–618.

[22] Ch. Reiher, *The clique density theorem*, Ann. of Math. (2) **184** (2016), no. 3, 683–707.

[23] A.F. Sidorenko, *Inequalities for functionals generated by bipartite graphs*, Diskret. Mat. **3** (1991), no. 3, 50–65.

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

A Appendix



F																																		
c_F^{OPT}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	2	2	4	6	12
$p(K_2,F)$	0	1	2	3	4	3	2	3	4	3	4	5	4	5	4	5	5	6	6	7	6	4	5	6	7	6	5	6	7	7	8	8	9	10
$[\![X_1 \times X_1]\!]_1$	30	12	4	0	0	0	4	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$[\![X_1 \times X_2]\!]_1$	0	3	4	3	0	6	0	1	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$[\![X_1 \times X_3]\!]_1$	0	6	4	3	0	0	8	2	0	6	2	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$[\![X_1 \times X_4]\!]_1$	0	0	2	6	12	0	0	2	2	0	3	4	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$[\![X_1 \times X_5]\!]_1$	0	0	1	0	0	0	0	2	0	0	1	0	4	0	1	0	2	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$[\![X_1 \times X_6]\!]_1$	0	0	0	0	0	3	0	0	2	0	0	2	0	2	0	1	0	2	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$[\![X_2 \times X_2]\!]_1$	0	0	0	0	0	0	2	2	2	0	0	0	4	4	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
$[\![X_2 \times X_3]\!]_1$	0	0	0	0	0	0	4	2	1	4	2	0	0	0	2	2	0	0	0	0	0	6	2	1	0	0	0	0	0	0	0	0	0	0
$[\![X_2 \times X_4]\!]_1$	0	0	0	0	0	0	0	0	0	2	2	2	0	0	2	2	2	2	0	0	0	0	1	2	3	0	0	0	0	0	0	0	0	0
$[\![X_2 \times X_5]\!]_1$	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	1	0	0	0	0	0	1	0	0	0	5	2	0	1	0	0	0	0
$[\![X_2 \times X_6]\!]_1$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	1	0	4	0	1	2	0	1	0	0	0
$[\![X_3 \times X_3]\!]_1$	0	0	4	0	0	12	0	4	4	4	0	0	0	0	6	2	0	0	0	0	0	12	4	0	0	0	10	2	0	0	0	0	0	0
$[\![X_3 \times X_4]\!]_1$	0	0	0	0	0	0	0	2	2	0	2	4	8	4	2	0	4	2	0	0	0	0	4	2	0	8	0	2	2	0	0	0	0	0
$[\![X_3 \times X_5]\!]_1$	0	0	0	3	0	0	0	0	2	0	2	0	0	2	0	2	1	0	0	0	0	0	2	2	0	0	0	2	0	2	0	0	0	0
$[\![X_3 \times X_6]\!]_1$	0	0	0	0	0	0	0	0	1	0	0	0	0	4	0	2	0	2	0	0	12	0	0	3	6	0	0	0	2	0	2	0	0	0
$[\![X_4 \times X_4]\!]_1$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	4	4	12	12	0	0	0	0	0	0	10	6	4	4	4	0	0	0
$[\![X_4 \times X_5]\!]_1$	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	2	1	6	0	0	0	0	2	0	0	0	2	2	4	0	4	0	0
$[\![X_4 \times X_6]\!]_1$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	0	0	2	2	6	4	8	6	0
$[\![X_5 \times X_5]\!]_1$	0	0	0	0	6	0	0	0	0	0	0	4	0	0	0	0	0	2	0	0	0	0	0	0	0	4	0	0	2	0	0	2	0	0
$[\![X_5 \times X_6]\!]_1$	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	6	0	0	0	0	3	0	0	0	1	0	4	0	3	0
$[\![X_6 \times X_6]\!]_1$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	0	4	4	12	30

Table 1: All entries corresponding to $p(K_2, F)$ are multiplied by 10 and all entries corresponding to $[\![X_i \times X_j]\!]_1$ are multiplied by 30.

B Appendix

This Maple code computes c_F coefficients. Matrices A, B and M are defined in Subsection 2.2.2. X is a matrix of size 21×34 and it is defined in Appendix A (rows correspond to $[\![X_i \times X_j]\!]_1$). Vectors cFOPT, pF, cFM and cF (each of size 34) correspond to $c_F^{OPT}, p(K_2, F), c_F^M$ and c_F, respectively. Constant a corresponds to α.

```
restart:
with(LinearAlgebra):
A := Matrix([[32*k^2-96*k+96, 0, 4*k^2-16*k],
[0, 10*k^4-30*k^3-8*k^2+96*k-96, -10*k^4+35*k^3-4*k^2-80*k+96],
[4*k^2-16*k, -10*k^4+35*k^3-4*k^2-80*k+96, 10*k^4-40*k^3+24*k^2+64*k-96]]):
B := Matrix([[k-1, 1, k-2, 0, k-3, -1],
[0, 2, k-2, 0, 2*k-4, -2],
[0, 0, k-1, -1, 2*k-2, -2]]):
M:= (3/(2*k^4))*Matrix(Multiply(Transpose(B), Multiply(A, B))):
X:=(1/30)*Matrix([[30,12,4,0,0,0,4,2,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0],
[0,3,4,3,0,6,0,1,2,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0],
[0,6,4,3,0,0,8,2,0,6,2,0,0,0,2,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0],
[0,0,2,6,12,0,0,2,2,0,3,4,0,0,0,2,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0],
[0,0,1,0,0,0,0,2,0,0,1,0,4,0,1,0,2,0,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0],
[0,0,0,0,0,3,0,0,2,0,0,2,0,2,0,1,0,2,0,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0],
[0,0,0,0,0,0,2,2,2,0,0,0,4,4,0,0,0,0,0,0,6,0,0,0,0,0,0,0,0,0,0,0,0,0],
[0,0,0,0,0,0,4,2,1,4,2,0,0,0,2,2,0,0,0,0,0,6,2,1,0,0,0,0,0,0,0,0,0,0],
[0,0,0,0,0,0,0,0,0,2,2,2,0,0,2,2,2,2,0,0,0,0,1,2,3,0,0,0,0,0,0,0,0,0],
[0,0,0,0,0,0,0,0,0,1,0,0,0,0,2,0,1,0,0,0,0,0,1,0,0,0,5,2,0,1,0,0,0,0],
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,3,2,1,0,4,0,1,2,0,1,0,0,0],
[0,0,4,0,0,12,0,4,4,4,0,0,0,0,6,2,0,0,0,0,0,12,4,0,0,0,10,2,0,0,0,0,0,0],
[0,0,0,0,0,0,0,2,2,0,2,4,8,4,2,0,4,2,0,0,0,0,4,2,0,8,0,2,2,0,0,0,0,0],
[0,0,0,3,0,0,0,0,2,0,2,0,0,2,0,2,1,0,0,0,0,0,2,2,0,0,0,2,0,2,0,0,0,0],
[0,0,0,0,0,0,0,0,1,0,0,0,0,4,0,2,0,2,0,0,12,0,0,3,6,0,0,0,2,0,2,0,0,0],
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,4,4,12,12,0,0,0,0,0,0,10,6,4,4,4,0,0,0],
[0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,2,2,1,6,0,0,0,0,2,0,0,0,2,2,4,0,4,0,0],
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,2,3,0,0,2,2,6,4,8,6,0],
[0,0,0,0,6,0,0,0,0,0,0,4,0,0,0,0,0,2,0,0,0,0,0,0,0,4,0,0,2,0,0,2,0,0],
[0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,2,0,6,0,0,0,0,3,0,0,0,1,0,4,0,3,0],
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,0,2,0,4,4,12,30]]):
cFM := Vector(34):
k_ind := 0:
printlevel := 2:
for i to 6 do
    for j from i to 6 do
        k_ind := k_ind+1;
        if i = j then cFM := cFM+M(i, j)*Transpose(Row(X, k_ind));
                 else cFM := cFM+2*M(i, j)*Transpose(Row(X,k_ind));
        end if;
    end do;
end do:
cFOPT := Vector([0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,2,2,4,6,12]):
pF := (1/10)*Vector([0,1,2,3,4,3,2,3,4,3,4,5,4,5,4,5,5,6,6,7,6,4,5,6,7,6,5,6,7,7,8,8,9,10]):
```

```
a := (1/(5k^4))*(30*k^3 - 120*k^2 + 180*k - 96):
cF := Vector(34):
for i to 34 do
    cF(i) := cFOPT(i)-10*k*a*pF(i)-cFM(i)+(10*(k-1))*a
end do:
for i to 34 do
    printf("5*k^4*cF(%d) = %s\n", i, convert(expand(5*k^4*cF(i)), string))
end do
```