Minimum number of edges that occur in odd cycles *

Andrzej Grzesik† Ping Hu‡ Jan Volec§

Abstract

If a graph G has $n \geq 4k$ vertices and more than $n^2/4$ edges, then it contains a copy of C_{2k+1}. In 1992, Erdős, Faudree and Rousseau showed even more, that the number of edges that occur in a triangle of such a G is at least $2\lfloor n/2 \rfloor + 1$, and this bound is tight. They also showed that the minimum number of edges in G that occur in a copy of C_{2k+1} for $k \geq 2$ suddenly starts being quadratic in n, and conjectured that for any $k \geq 2$, the correct lower bound should be $2n^2/9 - O(n)$. Very recently, Füredi and Maleki constructed a counterexample for $k = 2$ and proved an asymptotically matching lower bound, namely that for any $\varepsilon > 0$ graphs with $(1+\varepsilon)n^2/4$ edges contain at least $(2+\sqrt{2})n^2/16 \sim 0.2134n^2$ edges that occur in C_5.

In this paper, we use a different approach to tackle this problem and prove the following stronger result: Every n-vertex graph with at least $\lfloor n^2/4 \rfloor + 1$ edges has at least $(2+\sqrt{2})n^2/16 - O(n^{15/8})$ edges that occur in C_5. Next, for all $k \geq 3$ and n sufficiently large, we determine the exact minimum number of edges that occur in C_{2k+1} for n-vertex graphs with more than $n^2/4$ edges, and show it is indeed equal to $\left\lfloor \frac{n^2}{4} \right\rfloor + 1 - \left\lfloor \frac{n+4}{6} \right\rfloor \left\lfloor \frac{n+1}{6} \right\rfloor = 2n^2/9 - O(n)$. For both of these results, we give a structural description of all the large extremal configurations as well as obtain the corresponding stability results, which answers a conjecture of Füredi and Maleki.

The main ingredient of our results is a novel approach that combines the semidefinite method from flag algebras together with ideas from finite forcibility of graph limits, which may be of independent interest. This approach allowed us to keep track of the additional extra edge needed to guarantee even the existence of a single copy of C_{2k+1}, which a standard flag algebra approach would not be able to handle. Also, we establish the first application of the semidefinite method in a setting, where the set of tight examples has exponential size and arises from two very different constructions.

1 Introduction

A classical result in graph theory is Mantel's Theorem [28], which states that every triangle-free graph on n vertices has at most $\lfloor n^2/4 \rfloor$ edges, and this result is tight. In other words, a graph

*This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 648509). This publication reflects only its authors' view; the European Research Council Executive Agency is not responsible for any use that may be made of the information it contains.

†Faculty of Mathematics and Computer Science, Jagiellonian University, ul. St. Łojasiewicza 6, 30-348 Kraków, Poland. Email: Andrzej.Grzesik@uj.edu.pl. The author was partially supported by the National Science Centre grant 2013/08/T/ST1/00108.

‡Current address: School of Mathematics, Sun Yat-sen University, Guangzhou, 510275, China. This work has been carried out while at: Department of Computer Science, University of Warwick, Coventry, CV4 7AL, United Kingdom. Email: huping9@mail.sysu.edu.cn.

§Department of Mathematics and Statistics, McGill University, Burnside Hall, 805 Sherbrooke West, Montreal H3A 2K6, Canada. Former affiliation: Department of Mathematics, ETH, 8092 Zürich, Switzerland. Email: jan@ucw.cz. The author was partially supported by the SNSF grant 200021-149111 and CRM-ISM fellowship.

with n vertices and $\lfloor n^2/4 \rfloor + 1$ edges must contain a triangle. But can we guarantee something stronger than just one triangle? In 1941, Rademacher proved that such graphs contain at least $\lfloor n/2 \rfloor$ triangles, and in 1992, Erdős, Faudree and Rousseau [13] showed that such graphs have at least $2\lfloor n/2 \rfloor + 1$ edges that occur in a triangle. Both results are tight simply by adding one edge to the complete balanced bipartite graph.

Erdős [12] also considered analogous questions for longer odd cycles in graphs with n vertices and $\lfloor n^2/4 \rfloor + 1$ edges, where adding an extra edge into the complete balanced bipartite graph is not optimal. He showed that every such graph contains at least $2n^2/9$ edges that occur in some odd cycle. This number is best possible, and it can be achieved by the following construction.

Construction 1. Let G_1 be an n-vertex graph with the following two 2-connected blocks that overlap on exactly one vertex:

1. a complete graph on $\left\lfloor \frac{2n+4}{3} \right\rfloor$ vertices, and
2. a complete balanced bipartite graph on $\left\lfloor \frac{n+1}{3} \right\rfloor$ vertices.

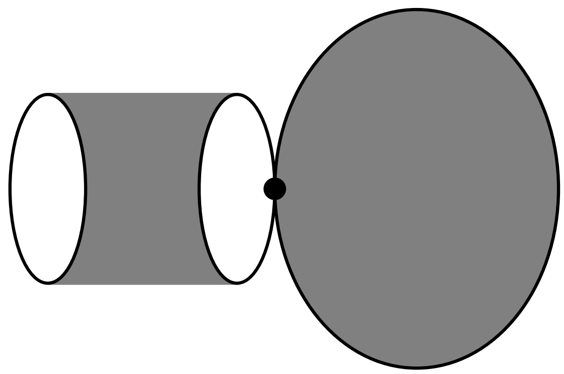


Figure 1: The graph G_1 from Construction 1. Gray areas represent all the possible edges.

Erdős, Faudree and Rousseau [13] conjectured that Construction 1 provides an extremal example also if we minimize the number of edges that occur only in copies of C_{2k+1} for a fixed $k \geq 2$. Again, we minimize over all n-vertex graphs with $\lfloor n^2/4 \rfloor + 1$ edges. The case of C_5 is Problem 11 in Erdős' paper [12] with interesting problems.

Conjecture 1.1 (Erdős-Faudree-Rousseau [13])**.** *Fix an integer $k \geq 2$. Every graph with n vertices and $\left\lfloor \frac{n^2}{4} \right\rfloor + 1$ edges contains at least $\frac{2}{9}n^2 - O(n)$ edges that occur in C_{2k+1}.*

Very recently, Füredi and Maleki [18] constructed the following n-vertex graph with $\lfloor n^2/4 \rfloor + 1$ edges, out of which only $\frac{2+\sqrt{2}}{16} \cdot n^2 + O(n) \approx 0.2134n^2$ occur in C_5, which disproves Conjecture 1.1 for $k = 2$.

Construction 2. Let G_2 be an n-vertex graph whose vertex-set is divided into four parts A, B, C and D of sizes $\frac{2-\sqrt{2}}{4} \cdot n$, $\frac{n}{4}$, $\frac{n}{4}$ and $\frac{\sqrt{2}}{4} \cdot n$, respectively. The edge-set of G_2 consists of all the edges between the parts A and B, B and C, C and D, and all the edges inside the part D. In other words, G_2 is a non-balanced blowup of a path on four vertices, where one of the endpoints of the path has a loop.

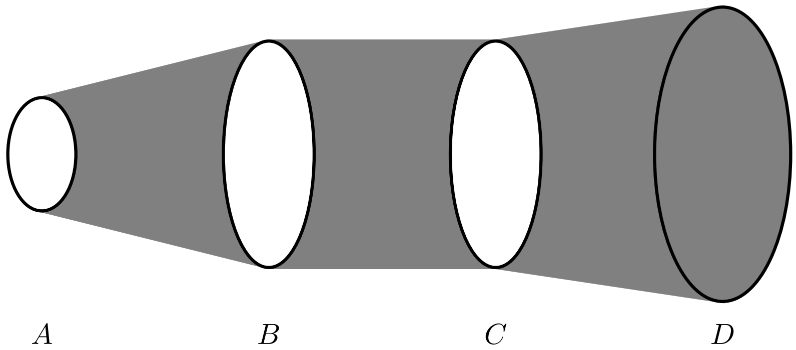


Figure 2: The graph G_2 from Construction 2. Gray areas represent all the possible edges. The respective sizes are $\frac{2-\sqrt{2}}{4} \cdot n$, $\frac{n}{4}$, $\frac{n}{4}$, and $\frac{\sqrt{2}}{4} \cdot n$.

Füredi and Maleki [18] developed a new version of Zykov's symmetrization method, and obtained the following asymptotic solution to this problem for all odd cycles of length at least five.

Theorem 1.2 ([18]). *For every $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that if a graph G on $n > n_0$ vertices has $\left(\frac{1}{4} + \varepsilon\right) n^2$ edges, then G contains at least $\frac{2+\sqrt{2}}{16} \cdot n^2$ edges that occur in C_5. Moreover, for any fixed $k \geq 3$, G contains at least $\frac{2}{9} \cdot n^2$ edges that occur in C_{2k+1}.*

Our first two results answer a conjecture of Füredi and Maleki that the assumption on the number of edges of G can be lowered to $\lfloor n^2/4 \rfloor + 1$, which is indeed best possible.

Theorem 1.3. *If an n-vertex graph has $\left\lfloor \frac{n^2}{4} \right\rfloor + 1$ edges, then it contains at least $\frac{2+\sqrt{2}}{16} \cdot n^2 - O\left(n^{15/8}\right)$ edges that occur in C_5.*

Theorem 1.4. *For every integer $k \geq 3$, if an n-vertex graph has $\left\lfloor \frac{n^2}{4} \right\rfloor + 1$ edges, then it contains at least $\frac{2}{9} \cdot n^2 - O(n)$ edges that occur in C_{2k+1}.*

In the case of odd cycles of length at least 7 and n sufficiently large, we determine the exact value of the number of edges that occur in C_{2k+1}, which indeed matches the value given by Construction 1, which answers another conjecture of Füredi and Maleki [17, Conjecture 8].

Theorem 1.5. *There exists $n_0 \in \mathbb{N}$ such that the following is true for any n-vertex graph G with $n \geq n_0$. If G has $\left\lfloor \frac{n^2}{4} \right\rfloor + 1$ edges, then it contains at least $\left\lfloor \frac{n^2}{4} \right\rfloor + 1 - \left\lfloor \frac{n+4}{6} \right\rfloor \left\lfloor \frac{n+1}{6} \right\rfloor$ edges that occur in C_{2k+1}.*

The main tool in our proofs is the semidefinite method from flag algebras, which we apply in a specific 2-edge-colored setting. This approach has the unfortunate by-product, that we lose track of the additional edge that is needed to guarantee even the existence of a single copy of C_{2k+1}. In order to overcome this, we apply a trick inspired by techniques used in the area of so-called finitely forcible graph limits. This allows us to obtain a tight bound from flag algebras conditioned by having a positive triangle density, and then handle the (almost) triangle-free case using a standard stability argument. A closely related difficulty of our approach arises from the fact that instead of most applications of the semidefinite method, where there is only one tight example, the flag algebra formulation of this problem has a significantly larger set of tight examples. Nevertheless, we were still able to obtain a tight result in this setting.

We guided our method to establish a slightly stronger flag algebra claims which yield also the corresponding stability results:

Theorem 1.6. *For every $\varepsilon > 0$ there exist $\delta > 0$ and $n_0 \in \mathbb{N}$ such that the following is true for any $n > n_0$. If G is an n-vertex graph with $\left(\frac{1}{4} \pm \delta\right) n^2$ edges out of which $\left(\frac{2+\sqrt{2}}{16} \pm \delta\right) n^2$ occur in C_5, then the edge-set of G can be modified on at most εn^2 pairs so that the resulting graph is isomorphic to Construction 2.*

Theorem 1.7. *Fix an integer $k \geq 3$. For every $\varepsilon > 0$ there exist $\delta > 0$ and $n_0 \in \mathbb{N}$ such that the following is true for any $n > n_0$. If G is an n-vertex graph with $\left(\frac{1}{4} \pm \delta\right) n^2$ edges out of which $\left(\frac{2}{9} \pm \delta\right) n^2$ occur in C_{2k+1}, then the edge-set of G can be modified on at most εn^2 pairs so that the resulting graph is isomorphic to Construction 1.*

Using the above stability results, we fully describe all the sufficiently large graphs that contain the minimum value of edges that occur in odd cycles of length at least 5. The description of the tight graphs in the case of pentagons is given by Theorem 6.1. For all the longer odd cycles, the description is provided by Theorem 7.1, which in turn proves both Theorems 1.4 and 1.5.

This paper is organized as follows. In Section 2, we describe the notation and introduce parts from the flag algebra framework we are going to use. In Section 3, we present our proof of Theorem 1.3, and in Section 4, we adapt the approach to cope with odd cycles of length at least 7. Section 5 is devoted to the corresponding stability results of the Constructions 1 and 2. Finally, in Sections 6 and 7 we provide the exact description of the tight extremal graphs. Section 8 concludes the paper with remarks and related open problems.

2 Notation and preliminaries

We start with the definition of the *induced density* of a k-vertex (small) graph F in an n-vertex (large) graph G, which we denote by $p(F, G)$. If $n \geq k$, then $p(F, G)$ is the probability that a randomly chosen k-vertex induced subgraph of G is isomorphic to F. In the case when $k > n$, the value of $p(F, G)$ is simply equal to zero.

In order to distinguish the edges that occur in some copy of C_5 (or more generally C_{2k+1} for some fixed $k \geq 2$) in graphs G in question, we will work with edge-colorings of G where the edges are colored using two colors – *red* and *blue*. With a slight abuse of notation, we will use G both to refer to the underlying graph and to the edge-colored graph, whenever it will be clear from the context which variant we intend to use. A graph G with edges colored by red and blue will be called a *red/blue-colored graph*. Through the whole paper, we will use the convention that none of the blue edges of G can occur in a copy of C_{2k+1} for a given $k \geq 2$. Let us emphasize that we do not put any restriction on the red edges of G, so in particular any graph G can be completely colored with red.

The definition of the induced density $p(F, G)$ naturally generalizes to the edge-colored setting. For convenience, we extend the definition of $p(F, G)$ also to graphs F where we allow the edges to be colored with three colors – red, blue or black (the edges of G will always be colored only with red and blue). The interpretation of an edge of F being black will be that we do not care whether G contains a copy of F where the edge is colored red or blue. Therefore, for a k-vertex

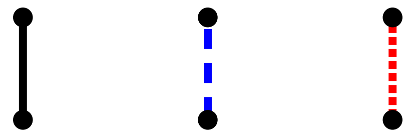


Figure 3: Our convention used for depicting black, blue and red edges – black edges are drawn with solid lines, blue edges with dashed lines and red edges with dotted lines.

red/blue/black-colored graph F and an n-vertex red/blue-colored graph G, the value of $p(F, G)$ is the probability that a random k-vertex subgraph of G is isomorphic to one of the graphs that can be obtained from F by recoloring each of its black edges to either red or blue.

We depict red/blue/black-colored graphs in the following way. We draw black edges using solid lines, for blue edges we used dashed lines, and finally red edges will be depicted using dotted lines; see Figure 3.

2.1 F-free graphs and sequences of almost F-free graphs

We will also use the following notion of *F-free graphs* and *sequences of almost F-free graphs*. For a k-vertex graph F and a graph G, we say that a graph G is *F-free*, if G does not contain F as a subgraph. For a sequence of graphs $(G_i)_{i\in\mathbb{N}}$, where the i-th graph G_i has n_i vertices, we say that $(G_i)_{i\in\mathbb{N}}$ is almost F-free, if G_i contains only $o\left(n_i^k\right)$ copies of F. This notion naturally generalizes to the red/blue-colored setting.

If $\mathcal{F}$ is a finite collection of graphs, we say that G is $\mathcal{F}$-free and $(G_i)_{i\in\mathbb{N}}$ is almost $\mathcal{F}$-free, if G is F-free for every $F \in \mathcal{F}$ and $(G_i)_{i\in\mathbb{N}}$ is almost F-free for every $F \in \mathcal{F}$, respectively. We also extend the notion of being F-free to red/blue/black-colored graphs F, where being F-free corresponds to being $\mathcal{F}(F)$-free, where $\mathcal{F}(F)$ denotes the family of red/blue-colored graphs consisting of all the possible recolorings of the black edges in F by red or blue. Analogously, we extend the notion of being almost F-free, and the notions of $\mathcal{F}$-free and almost $\mathcal{F}$-free for finite families $\mathcal{F}$ consisting of red/blue/black-colored graphs.

Now let us recall a classical generalization of the theorem of Kővari, Sós and Turán to r-uniform hypergraphs (or just *r-graphs* for short) which is due to Erdős [11].

Theorem 2.1. *If H is an r-graph on n vertices with no copy of the complete r-partite r-graph that has all the parts of size ℓ, then the number of r-edges in H is at most $O\left(n^{r-1/\ell^{(r-1)}}\right)$.*

A standard averaging argument together with Theorem 2.1 yields the following result on supersaturation in dense graphs, which will be one of the ingredients we will use in the proofs of Theorems 1.3 and 1.4.

Corollary 2.2. *Fix F an h-vertex red/blue-colored graph and a positive integer b. If G is an n-vertex red/blue-colored graph that does not contain the b-blowup of F as a subgraph, then the number of copies of F in G is $O\left(n^{h-1/b^{(h-1)}}\right)$.*

2.2 Flag Algebras

In this subsection, we describe parts of the flag algebra framework of Razborov [33] that will be relevant for our exposition. Flag algebras play a crucial role in our proofs of Theorems 1.3

and 1.4. We follow the notation from [33] with a few minor alternations specific to sequences of almost $\mathcal{F}$-free graphs. Flag algebras have been very successful in tackling various problems. To mention some of them: Caccetta-Häggkvist conjecture [21, 25, 36], various Turán-type problems in graphs [10, 20, 22, 24, 29, 31, 32, 34, 37, 39], hypergraphs [3, 15, 16, 19, 30] and hypercubes [2, 5], extremal problems in a colored environment [4, 9, 23, 26] and also to problems in geometry [27] or extremal theory of permutations [6]. For more details on these applications, see a survey of Razborov [35].

The central object of interest in flag algebras are so-called *convergent sequences* of finite discrete objects, for example finite graphs. In this paper, we apply the framework to sequences of red/blue-colored almost $\mathcal{F}$-free graphs, for two certain choices of $\mathcal{F}$ (the two families will be explicitly specified in Sections 3 and 4, respectively).

In the following, we describe a *flag algebra* $\mathcal{A}^\sigma$, where σ is a fixed vertex-labelled red/blue-colored graph, on all the red/blue-colored graphs with a fixed copy of a labelled graph σ. The graph σ is usually called a *type*. Note that we will use simply $\mathcal{A}$ to refer to the algebra $\mathcal{A}^\emptyset$, where $\emptyset$ is the empty type.

Fix a type σ. Let $\mathcal{H}^\sigma$ be the set of all finite red/blue-colored graphs with a fixed *embedding* of σ, i.e., an injective mapping θ from $V(\sigma)$ to $V(H)$ such that θ is an isomorphism between σ and $H[\mathrm{im}(\theta)]$. The elements of $\mathcal{H}^\sigma$ are called *σ-flags*, and the subgraph induced by $\mathrm{im}(\theta)$ is called the *root* of a σ-flag. For every $\ell \in \mathbb{N}$, we let $\mathcal{H}^\sigma_\ell$ to be the subset of $\mathcal{H}^\sigma$ containing all of its ℓ-vertex graphs. Let $\mathbb{R}\mathcal{H}^\sigma$ be the set of all formal linear combinations of the σ-flags with real coefficients, and $\mathcal{K}^\sigma$ the linear subspace of $\mathbb{R}\mathcal{H}^\sigma$ generated by all the combinations of the form

$$H - \sum_{H' \in \mathcal{H}^\sigma_{v(H)+1}} p(H, H') \cdot H'.$$

The algebra $\mathcal{A}^\sigma$ is defined as $\mathbb{R}\mathcal{H}^\sigma$ factored by $\mathcal{K}^\sigma$, and the element corresponding to $\mathcal{K}^\sigma$ in $\mathcal{A}^\sigma$ is the zero element of $\mathcal{A}^\sigma$. $\mathcal{A}^\sigma$ comes with a natural definition of the addition; the notion of multiplication is slightly more involved. Firstly, we describe a product of two σ-flags $H_1 \in \mathcal{H}^\sigma_k$ and $H_2 \in \mathcal{H}^\sigma_\ell$. For a σ-flag $H \in \mathcal{H}^\sigma_{k+\ell-v(\sigma)}$ with θ being the fixed embedding of σ, we define $p(H_1, H_2; H)$ to be the probability that a randomly chosen subset of $V(H) \setminus \theta(V(\sigma))$ of size $k - v(\sigma)$ and its complement in $V(H) \setminus \theta(V(\sigma))$ of size $\ell - v(\sigma)$ extend $\theta(V(\sigma))$ in H to σ-flags isomorphic to H_1 and H_2, respectively. We set

$$H_1 \times H_2 := \sum_{H \in \mathcal{H}^\sigma_{k+\ell-v(\sigma)}} p(H_1, H_2; H) \cdot H,$$

and extend the notion linearly to the elements of $\mathcal{A}^\sigma$. Note that the unique σ-flag of order $v(\sigma)$ is, modulo $\mathcal{K}^\sigma$, the neutral element of the product in $\mathcal{A}^\sigma$.

Fix a finite family $\mathcal{F}$ of red/blue-colored graphs and an $\mathcal{F}$-free type σ. The presented exposition of the flag algebra $\mathcal{A}^\sigma$ on red/blue-colored graphs naturally adapts to the setting of $\mathcal{F}$-free red/blue-colored graphs, simply by replacing the set $\mathcal{H}^\sigma$ with the set of all red/blue-colored $\mathcal{F}$-free σ-flags.

Now consider an infinite sequence $(G_i)_{i\in\mathbb{N}}$ of red/blue-colored almost $\mathcal{F}$-free graphs with increasing orders. We call the sequence *convergent* if the probabilities $p(H, G_i)$ converge for every $H \in \mathcal{H}$. By compactness, every sequence $(G_i)_{i\in\mathbb{N}}$ has a convergent subsequence. For the rest of this section, we will assume that $(G_i)_{i\in\mathbb{N}}$ is convergent. For $H \in \mathcal{H}$, we set $\phi(H) = \lim_{i\to\infty} p(H, G_i)$, and linearly extend ϕ to the elements of $\mathcal{A}$. We refer to the mapping ϕ as to the *limit* of the sequence. For every red/blue-colored graph H, it holds that $\phi(H) \geq 0$. Moreover, the sequence $(G_i)_{i\in\mathbb{N}}$ is almost $\mathcal{F}$-free, hence $\phi(F) = 0$ for all $F \in \mathcal{F}$. It follows that the restriction of ϕ to $\mathcal{F}$-free red/blue-colored graphs is in fact an algebra homomorphism from $\mathcal{A}_\mathcal{F}$ to $\mathbb{R}$. We let $\mathrm{Hom}^+(\mathcal{A}_\mathcal{F}, \mathbb{R})$ to be the set of all homomorphisms ψ from $\mathcal{A}_\mathcal{F}$ to $\mathbb{R}$ such that $\psi(H) \geq 0$ for every $\mathcal{F}$-free $H \in \mathcal{H}$.

For an $\mathcal{F}$-free type σ and its embedding θ in G_i, we define G_i^θ to be the red/blue-colored graph rooted on θ. For every $i \in \mathbb{N}$ and $H^\sigma \in \mathcal{H}^\sigma$, let $p_i^\theta(H^\sigma) = p(H^\sigma, G_i^\sigma)$. Picking θ at random gives rise to a probability distribution $\mathbf{P}_\mathbf{i}^\sigma$ on mappings from $\mathcal{A}^\sigma$ to $\mathbb{R}$. It holds that the sequence of $(\mathbf{P}_\mathbf{i}^\sigma)_{i\in\mathbb{N}}$ weakly converges to a Borel probability measure on $\mathrm{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$, see [33, Theorems 3.12 and 3.13]. In fact, for any σ such that $\phi(\sigma) > 0$, the homomorphism ϕ fully determines the limit probability distribution [33, Theorem 3.5]. We denote the limit of $(\mathbf{P}_\mathbf{i}^\sigma)_{i\in\mathbb{N}}$ by $\mathbf{P}_\phi^\sigma$. Furthermore, since $(G_i)_{i\in\mathbb{N}}$ is almost $\mathcal{F}$-free, any mapping ϕ^σ drawn from the support of the distribution $\mathbf{P}_\phi^\sigma$ is in fact an algebra homomorphism from $\mathcal{A}_\mathcal{F}^\sigma$ to $\mathbb{R}$ such that $\phi^\sigma(H^\sigma) \geq 0$ for any σ-flag H^σ.

The last notion we introduce is the *averaging operator* (also called the *downward operator*) $[\![\cdot]\!]_\sigma : \mathcal{A}_\mathcal{F}^\sigma \to \mathcal{A}_\mathcal{F}$. It is a linear operator defined on the σ-flags H^σ by

$$[\![H^\sigma]\!]_\sigma := p_H^\sigma \cdot H,$$

where H is the unlabelled red/blue-colored graph corresponding to H^σ, and p_H^σ is the probability that a random injection from $V(\sigma)$ to $V(H)$ yields a σ-flag isomorphic to H^σ. A key relation is

$$\forall A^\sigma \in \mathcal{A}_\mathcal{F}^\sigma, \quad \phi\left([\![A^\sigma]\!]_\sigma\right) = \phi\left([\![\sigma]\!]_\sigma\right) \cdot \int_{\phi^\sigma} \phi^\sigma(A^\sigma)\, d\mathbf{P}_\phi^\sigma\,. \tag{1}$$

If $\phi^\sigma(A^\sigma) \geq 0$ with probability one for some $A^\sigma \in \mathcal{A}_\mathcal{F}^\sigma$, then (1) yields that $\phi\left([\![A^\sigma]\!]_\sigma\right) \geq 0$. In particular, $\phi\left([\![A^\sigma \times A^\sigma]\!]_\sigma\right) \geq 0$ for every $\phi \in \mathrm{Hom}^+(\mathcal{A}_\mathcal{F}, \mathbb{R})$ and every $A^\sigma \in \mathcal{A}_\mathcal{F}^\sigma$.

3 Edges that occur in pentagons — proof of Theorem 1.3

We start the proof by formulating the statement of Theorem 1.3 into the language of red/blue-colored graphs. This statement is convenient for the flag algebra framework, which we intend to apply.

Theorem 3.1. *If G is a red/blue-colored graph on n vertices with $\left\lfloor \frac{n^2}{4} \right\rfloor + 1$ edges and no blue edge occur in C_5, then G contains at least $\frac{2+\sqrt{2}}{16} \cdot n^2 - O(n^{15/8})$ red edges.*

It is straightforward to check that the statements of Theorem 1.3 and Theorem 3.1 are equivalent. In the rest of the section, we give a proof of Theorem 3.1. We split the proof into the following two cases: either G contains many triangles and then we apply flag algebras, or, G contains only a small number of triangles in which case we use stability to show that G is close to the complete bipartite graph. Since G has more than $n^2/4$ edges, it follows that in the second case G must have many red edges (in fact, more than Theorem 3.1 asks for).

3.1 Flag algebra setting

We start with describing the precise setting of flag algebras we are going to use. Clearly, every G from Theorem 3.1 is B_5-free, where B_5 is the 5-cycle with one blue and four black edges (recall that an edge is black if it is either red or blue). But we can say more. Suppose F is a red/blue-colored

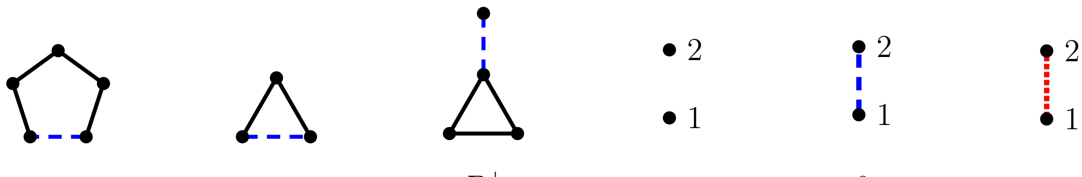


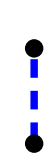
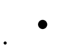
Figure 4: The graphs B_5, B_3 and B_3^+ used in the construction of the $\mathcal{F}_{\mathrm{C5}}$-free flag algebra $\mathcal{A}_{\mathrm{C5}}$, and the types λ, β and ρ.

graph such that the b-blowup of F, for some positive integer b, contains C_5 with at least one blue edge. Then, by Corollary 2.2, G can contain only $O\left(n^k\right)$ copies of such a graph F, where k is a rational strictly smaller than $v(F)$ and depends only on F and b. For example, since the 2-blowup of the graph B_3 depicted in Figure 4 contains C_5 with at least one blue edge, G contains only $O\left(n^{3-1/4}\right)$ copies of B_3. In other words, all but $O\left(n^{11/4}\right)$ triangles in G have only red edges. Analogously, the 2-blowup of the graph B_3^+, which is also depicted in Figure 4, contains C_5 with two blue edges. Therefore, G may contain only $O\left(n^{31/8}\right)$ copies of B_3^+.

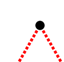
Let $\mathcal{F}_{\mathrm{C5}} := \left\{B_3, B_3^+, B_5\right\}$. For brevity, we will write $\mathcal{A}_{\mathrm{C5}}$ and $\mathcal{A}^\sigma_{\mathrm{C5}}$ instead of $\mathcal{A}_{\mathcal{F}_{\mathrm{C5}}}$ and $\mathcal{A}^\sigma_{\mathcal{F}_{\mathrm{C5}}}$. Next, let λ, β and ρ be the three red/blue-colored flag algebra types of size two with labels 1 and 2, where λ denotes the non-edge type, β the blue-edge type, and ρ the red-edge type; see Figure 4. We define $\mathcal{H}_4^\lambda$ to be the set of all the non-isomorphic 4-vertex λ-flags in $\mathcal{F}^\lambda_{\mathrm{C5}}$, $\mathcal{H}_4^\beta$ the set of the 4-vertex β-flags in $\mathcal{F}^\beta_{\mathrm{C5}}$, and $\mathcal{H}_4^\rho$ the set of the 4-vertex ρ-flags in $\mathcal{F}^\rho_{\mathrm{C5}}$. It holds that $\left|\mathcal{H}_4^\lambda\right| = 76$, $\left|\mathcal{H}_4^\beta\right| = 33$, and $|\mathcal{H}_4^\rho| = 43$. Let v_λ be the 76-dimensional vector in $\left(\mathcal{A}^\lambda_{\mathrm{C5}}\right)^{76}$ such that the i-th element of v_λ is equal to the i-th element of $\mathcal{H}_4^\lambda$. Analogously, we let v_β to be the 33-dimensional vector, where the coordinates correspond to the elements of $\mathcal{H}_4^\beta$, and v_ρ the 43-dimensional vector, where the coordinates correspond to the elements of $\mathcal{H}_4^\rho$. Finally, we define $\mathcal{H}_6$ to be the set of all 6-vertex red/blue-colored $\mathcal{F}_{\mathrm{C5}}$-free graphs; it can be checked that there are precisely 756 such graphs.

Theorem 3.1 is concerned with red/blue-colored graphs of density at least 1/2, which translates to the flag algebra framework as studying homomorphisms $\psi \in \mathrm{Hom}^+(\mathcal{A}_{\mathrm{C5}}, \mathbb{R})$ satisfying $\psi\big($  $+$  $\big) \geq 1/2$. As $\psi\big($  $+$  $+$  $\big) = 1$, the last condition is equivalent to $\psi\big($  $+$  $-$  $\big) \geq 0$.

Proposition 3.2. *There exist three positive-semidefinite matrices L, B and R with the entries from the field $\mathbb{Q}\left[\sqrt{2}\right]$, and non-negative numbers $a \in \mathbb{Q}\left[\sqrt{2}\right]$, $b \in \mathbb{Q}\left[\sqrt{2}\right]$ and $c_H \in \mathbb{Q}\left[\sqrt{2}\right]$ for $H \in \mathcal{H}_6$, such that in the algebra $\mathcal{A}_{\mathrm{C5}}$, the expression*

$$\left[\!\left[v_\lambda^{\mathrm{T}} L v_\lambda\right]\!\right]_\lambda + \left[\!\left[v_\beta^{\mathrm{T}} B v_\beta\right]\!\right]_\beta + \left[\!\left[v_\rho^{\mathrm{T}} R v_\rho\right]\!\right]_\rho + \Big($$  $+$  $-$ $$\Big) \times \Big(a \cdot$$  $+ b \cdot$  $$\Big) + \sum_{H \in \mathcal{H}_6} c_H \cdot H$$

is equal to

 $\times \Big(8 \cdot$  $- \left(2+\sqrt{2}\right) \cdot$ $\bullet$ $\Big).$

Moreover, if H is a 6-vertex red/blue-colored graph from $\mathcal{H}_6$ that contains an induced copy of any $P_5 \in \mathcal{P}_5$ or C_4^X, then $c_H > 0$.

Finding the positive-semidefinite matrices L, B and R, and the non-negative numbers a, b and c_H for $H \in \mathcal{H}_6$ such that the claimed identity holds can be expressed as a semidefinite program. We used an SDP solver called CSDP [8] together with a computer algebra software SAGE [40] to help us solving the corresponding semidefinite program. Since some of the numbers and the entries of the matrices are too large to be presented in a printed form, we created a webpage and uploaded all the corresponding data there. The URL of the webpage is `http://honza.ucw.cz/proj/EdgesInCycles/`.

We also prepared a short verification script in SAGE that checks the correctness of the claimed identity; see also Appendix A for the details about the formal verification. The script, as well as a description of all the data files, can be also found at the webpage. Note that the matrices L, B and R are not stored directly. Instead, they are decomposed as

$$L = M_\lambda^{\mathrm{T}} \cdot \widehat{L} \cdot M_\lambda, \quad B = M_\beta^{\mathrm{T}} \cdot \widehat{B} \cdot M_\beta, \quad \text{and} \quad R = M_\rho^{\mathrm{T}} \cdot \widehat{R} \cdot M_\rho,$$

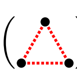
where $\widehat{L}$, $\widehat{B}$ and $\widehat{R}$ are positive definite matrices of sizes 50×50, 19×19 and 26×26, respectively, and M_λ, M_β and M_ρ are specific matrices of sizes 76×50, 33×19 and 43×26, respectively. Another advantage of this is that verifying whether a matrix is positive definite is faster from the practical point of view; see Appendix A.

3.2 Case 1 — Graphs with many triangles

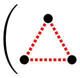
We first prove the theorem for graphs G that satisfy the assumptions of Theorem 3.1 and contain $\Omega\left(n^3\right)$ triangles. This will be the only case where we use flag algebra method, and the reason for that is the following. In order to apply flag algebra method, we pass the asymptotic statement to the limit. As we already mentioned in the introduction, an unfortunate consequence is that we completely lose control on having the additional edge that is needed to contain even a single copy of C_5. However, in the situation that G contains about $n^2/4$ edges and only a small number of triangles, a stability argument yields that G must be very close to the complete balanced bipartite graph. Such a situation will be analyzed in Section 3.3. Therefore, the statement we prove with flag algebras states that for every G that satisfies the assumptions of Theorem 3.1, at least one of the following is true:

1. G has at least $\frac{2+\sqrt{2}}{16} \cdot n^2 - O(n^{15/8})$ red edges, or,
2. G contains $o\left(n^3\right)$ triangles.

Suppose, for a contradiction, that Theorem 3.1 is false. Then there exists a sequence of red/blue-colored graphs $(G_i)_{i\in\mathbb{N}}$ of increasing orders n_i such that for i big enough every G_i has at most $\frac{2+\sqrt{2}}{16} \cdot n_i^2 - \omega\left(n_i^{15/8}\right)$ red edges. Without loss of generality, the sequence is convergent. Furthermore, the sequence $(G_i)_{i\in\mathbb{N}}$ is almost $\mathcal{F}_{\mathrm{C5}}$-free. Therefore, the sequence converges to a limit ϕ_0, which is an element of the set $\mathrm{Hom}^+\left(\mathcal{A}_{\mathrm{C5}}, \mathbb{R}\right)$. It is straightforward to check that the edge-density of ϕ_0 is equal to $1/2$. The following lemma states that such a limit ϕ_0 must have triangle density equal to zero.

Lemma 3.3. *Let $\delta > 0$ and $\phi \in \mathrm{Hom}^+\left(\mathcal{A}_{\mathrm{C5}}, \mathbb{R}\right)$. If* $\phi\Big($ $+$ $\Big) \geq \frac{1}{2}$ *and* $\phi\Big($$\Big) \geq \delta$*, then* $\phi\Big($$\Big) \geq \frac{2+\sqrt{2}}{8}$*. Moreover, if G is an n-vertex red/blue-colored graph with $\left\lfloor \frac{n^2}{4} \right\rfloor + 1$ edges, at least $\delta \cdot n^3$ triangles and no blue edge occur in C_5, then G contains at least $\frac{2+\sqrt{2}}{16} \cdot n^2 - O(n^{15/8})$ red edges.*

Proof. Proposition 3.2 yields that if $\psi \in \mathrm{Hom}^+(\mathcal{A}_{C5}, \mathbb{R})$ satisfies $\psi\Big($  $+$  $\Big) \geq 1/2$, then

$\psi\Big($  $\times \Big(8 \cdot$  $- 2 - \sqrt{2}\Big)\Big) \geq 0.$ (2)

It immediately follows that if the first factor of the product on the left-hand side, i.e., the triangle density in ψ, is at least $\delta > 0$, then inequality (2) yields that the second factor must be non-negative. In other words, the density of red edges is at least $\left(2+\sqrt{2}\right)/8$.

The moreover part of the lemma follows from a standard $O\left(n^{-1}\right)$ error estimate in the semidefinite method (for details, see, for example, [14]), and the $O\left(n^{-1/8}\right)$ estimate on the densities of B_3 and B_3^+ in G. □

Recall that $(G_i)_{i\in\mathbb{N}}$ is a sequence of n_i-vertex graphs with at most $\frac{2+\sqrt{2}}{16} \cdot n_i^2 - \omega\left(n_i^{15/8}\right)$ red edges. Applying Lemma 3.3 to (G_i) readily implies that G_i must contain $o\left(n_i^3\right)$ triangles.

3.3 Case 2 — Graphs with small number of triangles

It remains to verify Theorem 3.1 for graphs G that contain less than δn^3 triangles for an arbitrary $\delta > 0$. As we have already mentioned, in this case our plan is to use stability of triangle-free graphs to show that G must be close, in the so-called *edit-distance*, to a complete bipartite graph. Since the number of edges in G is strictly more than $n^2/4$, the graphs we are dealing with are essentially almost complete bipartite graphs plus an additional edge in one of the parts. Therefore, we will be able to show that nearly all the edges of G occur in C_5. This is summarized in the following lemma, which actually holds for any odd cycle of length at least five.

Lemma 3.4. *For every integer $k \geq 2$, there exists $\delta_k > 0$ such that the following is true. If G is an n-vertex graph with $\left\lfloor \frac{n^2}{4} \right\rfloor + 1$ edges and at most $\delta_k \cdot n^3$ triangles, then all but $o\left(n^2\right)$ edges of G occur in C_{2k+1}.*

Before proving Lemma 3.4, let us recall two classical results in extremal graph theory. The first one is the triangle removal lemma due to Ruzsa and Szemerédi [38].

Theorem 3.5. *If an n-vertex graph has $o(n^3)$ triangles, then it can be made triangle-free by removing at most $o(n^2)$ edges.*

Next, we recall a classical stability-type result for dense triangle-free graphs; for its proof, see, e.g., [7, Theorem VI.4.2].

Theorem 3.6. *If G is an n-vertex triangle-free graph with $n^2/4 - o(n^2)$ edges, then G contains an induced bipartite subgraph with minimum degree $n/2 - o(n)$.*

We are now ready to prove the main lemma of this subsection.

Proof of Lemma 3.4. Suppose that δ_k is sufficiently small. An application of Theorems 3.5 and 3.6 to G readily finds an induced bipartite subgraph on $n - o(n)$ vertices with minimum degree $\frac{n}{2} - o(n)$.

Let G_0 be an induced bipartite subgraph of G with maximum number of vertices that has the minimum degree at least $\frac{n}{2} - o(n)$. Let A and B be the parts of G_0 and let $L := V(G) \setminus V(G_0)$. Clearly, both A and B have sizes between $\frac{n}{2} \pm o(n)$ and $|L| = o(n)$.

The following claim states that G_0 has many edges between any two large subsets of A and B.

Claim 3.7. *If $A' \subseteq A$ and $B' \subseteq B$ are two sets of vertices of size $n/2 - o(n)$ each, then the number of edges in G_0 between the sets A' and B' is $n^2/4 - o(n^2)$.*

Proof. Let e be the number of edges between A' and B'. On one hand, G_0 has at least $n^2/4 - o(n^2)$ edges. On the other hand, $e(G_0) \leq e + |A \setminus A'| \cdot |B| + |B \setminus B'| \cdot |A| \leq e + o(n^2)$. □

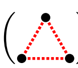
Our second claim states that any vertex with neighbors both in A and B allows us to find many edges that occur in C_{2k+1}.

Claim 3.8. *If some vertex $v_L \in L$ is adjacent both to a vertex $v_A \in A$ and a vertex $v_B \in B$, then all but $o(n^2)$ edges of G occur in C_{2k+1}.*

Proof. Let $V_{2k-1} \subseteq B$ be the neighborhood of v_A in G_0, and $v_B, v_2, \ldots, v_{2k-3}$ any path in G_0 that does not contain the vertex v_A. We set $V_{2k-2} \subseteq A$ to be the neighborhood of v_{2k-3} in G_0, $A' := V_{2k-2} \setminus \{v_2, \ldots, v_{2k-2}, v_A\}$ and $B' := V_{2k-1} \setminus \{v_B, v_3, \ldots, v_{2k-3}\}$.

It follows that both A' and B' have size at least $\frac{n}{2} - o(n)$. Therefore, the number of edges of the form $\{v_{2k-2}, v_{2k-1}\}$ between A' and B' is at least $n^2/4 - o(n^2)$ Each such an edge encloses a $(2k+1)$-cycle in G, which is of the form $v_L, v_B, v_2, \ldots, v_{2k-1}, v_A$. □

In order to finish the proof of the lemma, we simply need to find such a vertex v_L. Firstly, observe that $|L| \geq 1$, as otherwise G cannot have $\lfloor n^2/4 \rfloor + 1$ edges. Moreover, at least one $v_L \in L$ must have $\deg_G(v_L) \geq n/2$. If v_L would be adjacent only to A or only to B, then the subgraph $G_0 + v_L$ contradicts the choice of G_0. □

Now recall that if Theorem 3.1 would be false, then by Lemma 3.3 there exists a limit $\phi_0 \in \mathrm{Hom}^+(\mathcal{A}_{\mathrm{C5}}, \mathbb{R})$ such that $\phi_0\Big($  $\Big) = 0$ and $\phi_0\Big($  $\Big) \leq \frac{2+\sqrt{2}}{8}$. However, Lemma 3.4 yields that $\phi_0\Big($  $\Big) = \frac{1}{2}$. Therefore, there is no such ϕ_0, and the proof Theorem 3.1 is now finished.

4 Edges that occur in longer odd cycles — proof of Theorem 1.4

We adapt the approach presented in the previous section and give a proof of an asymptotic version of Theorem 1.4. The exact version will be obtained in Section 7, where we find a description of all the sufficiently large extremal constructions. We start with stating the main result of this section using the language of red/blue-colored graphs.

Theorem 4.1. *For every $\varepsilon > 0$ and integer $k \geq 3$, there exists $n_0 \in \mathbb{N}$ such that if G is a red/blue-colored graph on $n > n_0$ vertices with $\left\lfloor \frac{n^2}{4} \right\rfloor + 1$ edges and no blue edge occur in C_{2k+1}, then G contains at least $\left(\frac{2}{9} - \varepsilon\right) n^2$ red edges.*

The rest of this section is devoted to the proof of Theorem 4.1. First, we define a class of graphs $\mathcal{F}_{\mathrm{C7}}$ such that, for any fixed integer $k \geq 3$, the following will be true. If a sequence $(G_i)_{i \in \mathbb{N}}$ of red/blue-colored graphs is such that no blue edge occurs in C_{2k+1}, then (G_i) is almost $\mathcal{F}_{\mathrm{C7}}$-free. Analogously to the C_5 case, a k-blow of any graph $F \in \mathcal{F}_{\mathrm{C5}}$ contains a copy of C_{2k+1} with at least one blue edge. Similarly, a k-blowup of either the graph B_3^* or the graph B_5^+, which are both depicted in Figure 5, contains a copy of C_{2k+1} with at least one blue edge.

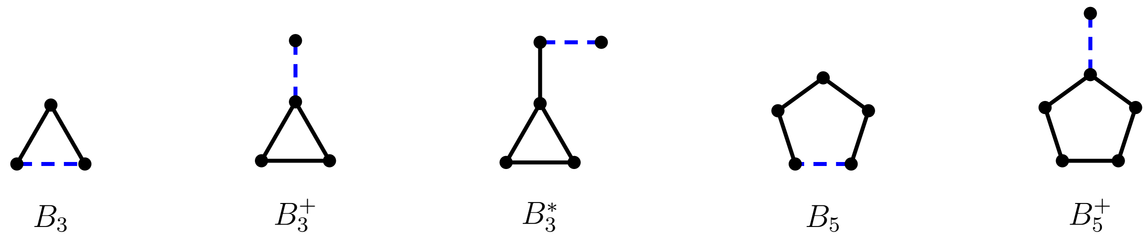


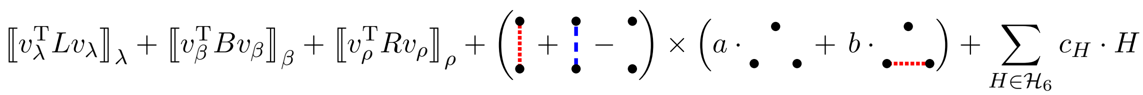
Figure 5: The family of graphs $\mathcal{F}_{\mathrm{C7}}$ used in the construction of the $\mathcal{F}_{\mathrm{C7}}$-free flag algebra $\mathcal{A}_{\mathrm{C7}}$.

Let $\mathcal{F}_{\mathrm{C7}} := \left\{B_3, B_3^+, B_3^*, B_5, B_5^+\right\}$. As we have just observed, any sequence of graphs satisfying the assumptions of Theorem 4.1 is almost $\mathcal{F}_{\mathrm{C7}}$-free. We use the class $\mathcal{F}_{\mathrm{C7}}$ to construct the corresponding flag algebras. Again, we refer to them $\mathcal{A}_{\mathrm{C7}}$ and $\mathcal{A}^{\sigma}_{\mathrm{C7}}$ instead of $\mathcal{A}_{\mathcal{F}_{\mathrm{C7}}}$ and $\mathcal{A}^{\sigma}_{\mathcal{F}_{\mathrm{C7}}}$.

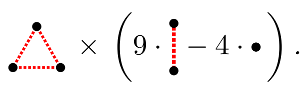
Recall the three types λ, β and ρ depicted in Figure 4. We define $\mathcal{H}_4^{\lambda}$, $\mathcal{H}_4^{\beta}$ and $\mathcal{H}_4^{\rho}$ to be the sets of all the 4-vertex λ-flags in $\mathcal{A}^{\lambda}_{\mathrm{C7}}$, β-flags in $\mathcal{A}^{\beta}_{\mathrm{C7}}$ and ρ-flags in $\mathcal{A}^{\rho}_{\mathrm{C7}}$, respectively. Since the set of all the 4-vertex flags in $\mathcal{A}_{\mathrm{C7}}$ is the same as the corresponding set in $\mathcal{A}_{\mathrm{C5}}$, we have again $\left|\mathcal{H}_4^{\lambda}\right| = 76$, $\left|\mathcal{H}_4^{\beta}\right| = 33$, and $|\mathcal{H}_4^{\rho}| = 43$. Let v_λ, v_β and v_ρ be the appropriate vectors, where the i-th element of $v_\lambda/v_\beta/v_\rho$ is equal to the i-th element of $\mathcal{H}_4^{\lambda}/\mathcal{H}_4^{\beta}/\mathcal{H}_4^{\rho}$.

This time, there are 741 non-isomorphic red/blue-colored $\mathcal{F}_{\mathrm{C7}}$-free graphs on 6 vertices. With a slight abuse of notation, we again denote the set of all such graphs by $\mathcal{H}_6$. An application of the flag algebra method for $\mathcal{A}_{\mathrm{C7}}$ yields the following:

Proposition 4.2. *There exist three positive-semidefinite matrices L, B and R with rational entries and non-negative rational numbers a, b and c_H, where $H \in \mathcal{H}_6$, such that in the algebra $\mathcal{A}_{\mathrm{C7}}$, the expression*



is equal to

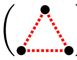


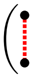
Moreover, if H is a 6-vertex red/blue-colored graph from $\mathcal{H}_6$ that contains an induced copy of any $P_4 \in \mathcal{P}_4$, then $c_H > 0$.

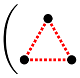
Again, we used CSDP and SAGE to find L, B, R, a, b and c_H. The webpage mentioned in Proposition 3.2 contains all the corresponding data, as well as a short SAGE script that verifies the claimed identity. As in Section 3, the matrices L, B and R are decomposed as

$$L = M_\lambda^{\mathrm{T}} \cdot \widehat{L} \cdot M_\lambda, \quad B = M_\beta^{\mathrm{T}} \cdot \widehat{B} \cdot M_\beta, \quad \text{and} \quad R = M_\rho^{\mathrm{T}} \cdot \widehat{R} \cdot M_\rho,$$

where $\widehat{L}$, $\widehat{B}$ and $\widehat{R}$ are positive definite matrices of sizes 58×58, 22×22 and 32×32, respectively, and M_λ, M_β and M_ρ are specific matrices of sizes 76×58, 33×22 and 43×32, respectively.

Corollary 4.3. *If $\delta > 0$ and $\phi \in \mathrm{Hom}^+(\mathcal{A}_{\mathrm{C7}}, \mathbb{R})$ such that* $\phi\Big($ $+$ $\Big) \geq \frac{1}{2}$ *and* $\phi\Big($$\Big) \geq \delta$, *then* $\phi\Big($$\Big) \geq \frac{4}{9}$.

Proof. By Proposition 4.2, any $\psi \in \mathrm{Hom}^+(\mathcal{A}_{\mathrm{C7}}, \mathbb{R})$ with $\psi\Big($ $+$ $\Big) \geq 1/2$ must satisfy

$\psi\Big($ $\times \Big(9 \cdot$  $- 4\Big)\Big) \geq 0.$ (3)

Since the triangle density in ϕ is at least $\delta > 0$, the density of red edges is at least $4/9$. □

Now if Theorem 4.1 would be false, then there exists some absolute constant $\varepsilon_0 > 0$ and a convergent sequence of red/blue-colored almost $\mathcal{F}_{\mathrm{C7}}$-free graphs $(G_i)_{i \in \mathbb{N}}$ of increasing orders (n_i) such that every G_i has at most $(2/9 - \varepsilon_0) \cdot (n_i)^2$ red edges. By Lemma 3.4, the limit of the triangle densities in the sequence must be positive. However, Corollary 4.3 yields that, for a sufficiently large i, the graph G_i has strictly more than $(2/9 - \varepsilon_0) \cdot (n_i)^2$ red edges; a contradiction.

5 Stability of Constructions 1 and 2

In this section, we show the corresponding stability for the extremal results presented in Sections 3 and 4, and prove Theorems 1.6 and 1.7. Let us start by recalling the following edge-colored variant of the induced graph removal lemma, which is a direct consequence of [1, Theorem 1.5].

Theorem 5.1. *For any $\varepsilon_{\mathrm{RL}} > 0$ and a finite family of red/blue-colored graphs $\mathcal{F}$, there exists $\delta_{\mathrm{RL}} > 0$ such that the following is true: If G is an n-vertex red/blue-colored graph with at most $\delta_{\mathrm{RL}} \cdot n^{v(F)}$ induced copies of F for all $F \in \mathcal{F}$, then the edge-set of G can be modified on at most $\varepsilon_{\mathrm{RL}} \cdot n^2$ pairs so that no induced subgraph of the resulting graph is isomorphic to an element of $\mathcal{F}$.*

Since the structure of Construction 1 is simpler than the structure of Construction 2, we begin with proving Theorem 1.7.

5.1 Odd cycles of length at least seven — stability of Construction 1

This whole subsection is devoted to the proof of Theorem 1.7. Recall that our task is, given an integer $k \geq 3$ and $\varepsilon > 0$, to find an integer n_0 and $\delta > 0$ so that for any graph G with $n \geq n_0$ vertices and $(1/4 \pm \delta)\, n^2$ edges out of which $(2/9 \pm \delta)\, n^2$ occur in C_{2k+1}, it holds that G is εn^2-close in the edit-distance to Construction 1. Since Construction 1 is $O(n)$-close to a disjoint union of $2n/3$-vertex clique and complete balanced bipartite graph on the remaining $n/3$ vertices, we show that G is εn^2-close to this construction.

Fix such a graph G. Following the notation from the previous sections, we color the edges of G that occur in some copy of C_{2k+1} red, and the other edges of G blue. Since G has only $(2/9 \pm \delta)\, n^2$ red edges, Lemma 3.4 yields that G contains at least $\delta_k \cdot n^3$ triangles. Without loss of generality, we may assume $\varepsilon \ll \delta_k$. Throughout the whole proof, we will use two auxiliary positive constants $\varepsilon_{\mathrm{RL}}$ and δ_{RL}, which will be determined during the proof, obeying the hierarchy $\delta \ll \delta_{\mathrm{RL}} \ll \varepsilon_{\mathrm{RL}} \ll \varepsilon$.

By Corollary 2.2, we can choose n_0 to be a large enough integer such that the graph G contains only $\delta n^{v(F)}$ copies of F for all $F \in \mathcal{F}_{\mathrm{C7}}$. We continue our exposition by showing that G

Figure 6: The family $\mathcal{P}_4$ containing all the 6 non-isomorphic red/blue-colorings of P_4.

cannot contain too many induced paths on four vertices. Let $\mathcal{P}_4$ be the set of all the six possible red/blue-colorings of the 4-vertex path; see also Figure 6. The following lemma directly follows from Proposition 4.2.

Lemma 5.2. *If* $\phi \in \mathrm{Hom}^+(\mathcal{A}_{\mathrm{C7}}, \mathbb{R})$ *satisfies* $\phi\Big($  $\Big) = \frac{4}{9}$ *and* $\phi\Big($  $\Big) = \frac{1}{18}$*, then* $\phi(P) = 0$ *for every* $P \in \mathcal{P}_4$*.*

Let n_0 be large enough so that the limit identity proven by flag algebras in Proposition 4.2 holds with an error of order $O(\delta)$ for any graph in question with at least n_0 vertices. Therefore, for any $F \in \mathcal{P}_4$, it holds that $p(F,G) = O(\delta) \ll \delta_{\mathrm{RL}}$. Set $\mathcal{F}$ to be the family containing

- all the red/blue-colored triangles with at least one blue edge, i.e, the graphs from B_3,
- all the 4-vertex red/blue-colored graphs that contain a copy of B_3^+,
- all the 5-vertex red/blue-colored graphs that contain a copy of B_3^*, and
- the six elements of $\mathcal{P}_4$.

Let δ_{RL} be the constant from Theorem 5.1 applied with the constant $\varepsilon_{\mathrm{RL}}$ and the family $\mathcal{F}$. Since $\delta \ll \delta_{\mathrm{RL}}$, the induced removal lemma yields a graph G' with no induced copy of F for all $F \in \mathcal{F}$, and differs from the original graph G on at most $\varepsilon_{\mathrm{RL}} \cdot n^2$ pairs. In other words, G' contains no induced path on 4 vertices and no (not necessarily induced) copy of B_3, B_3^+ or B_3^*. It follows that the number of edges in G' is $(1/4 \pm 2\varepsilon_{\mathrm{RL}})\,n^2$. By choosing $\varepsilon_{\mathrm{RL}}$ to be much smaller than ε, it is enough to show that G' is $\left(\varepsilon/2 \cdot n^2\right)$-close to Construction 1.

Let B be the set of vertices of G' that are incident to at least one blue edge or have a neighbor that is incident to a blue edge, H the subgraph induced by B, and A the vertices of G' that are not in B. Now we prove the following three claims describing the structure of G' in terms of A and B.

Claim 5.3. *The graph H is bipartite.*

Proof. Suppose for contradiction there is an odd cycle in H. Since H does not contain any induced path on four vertices, H must contain a triangle xyz. Since H is B_3-free, all three edges of the triangle must be red. Also, neither x nor y nor z is incident to a blue edge, because H is B_3^+-free. However, any of the three vertices, say x, is incident to a vertex w such that w is then incident to a blue edge so H fails to be B_3^*-free; a contradiction. □

Claim 5.4. *A has size at least $\delta_k \cdot n$.*

Proof. The number of triangles in G' is at least $(\delta_k - \varepsilon_{\mathrm{RL}})\,n^3$. Since triangles in G' can lie only inside the set A, $|A| \geq (\delta_k - \varepsilon_{\mathrm{RL}})^{1/3} \cdot n > \delta_k \cdot n$. □

Claim 5.5. *There are no edges between A and B.*

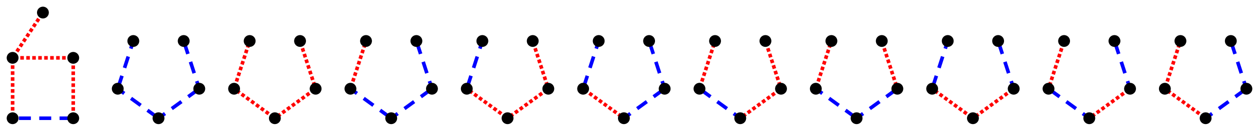


Figure 7: The red/blue-colored graph C_4^X and the family $\mathcal{P}_5$ containing all the 10 non-isomorphic red/blue-colorings of P_5.

Proof. Suppose there is an edge connecting two vertices $u \in A$ and $v \in B$. Since $u \notin B$, the vertex v is not incident to any blue edge, however, it has a neighbor w, which then has a neighbor x such that the edge $\{w, x\}$ has blue color. Since H is bipartite, x is not connected to v. Hence $\{u, v, w, x\}$ induces a 4-vertex path in G', which is a contradiction. $\square$

Let $a := |A|/n$ and $b := 1 - a = |B|/n$. In Claims 5.3-5.5, we have shown that the set B induces a bipartite graph and there are no edges in G' between A and B. Therefore, the edge-density of G' is bounded by a function $f(a) := a^2 + (1-a)^2/2$. The following observation directly follows from continuity of f, f having no local maximum on $(0, 1)$, and compactness of $[0, 1]$.

Observation 5.6. *The function $f(a)$ for $a \in [\delta_k, 2/3]$ has a unique maximum $1/2$ for $a = 2/3$. Moreover, if the value of the function for $a \in [\delta_k, 2/3 + O(\varepsilon_{\mathrm{RL}})]$ is close to $1/2$, then the value of a is close to $2/3$.*

Since the number of edges of G' is $(1/4 \pm 2\varepsilon_{\mathrm{RL}})\, n^2$, Observation 5.6 yields that a must be close to $2/3$. It follows that $|A| = (2/3 \pm O(\varepsilon_{\mathrm{RL}}))\, n$ and $|B| = (1/3 \pm O(\varepsilon_{\mathrm{RL}}))\, n$. Moreover, the bipartite graph H must have parts of sizes $(1/6 \pm O(\varepsilon_{\mathrm{RL}}))\, n$, and all but $O(\varepsilon_{\mathrm{RL}})n^2$ pairs between the parts are joined by an edge. Finally, the number of non-adjacent pairs with both endpoints in A is at most $O(\varepsilon_{\mathrm{RL}})n^2$. Since $\varepsilon_{\mathrm{RL}} \ll \varepsilon$, we can easily modify $\varepsilon/2 \cdot n^2$ pairs of G' in order to obtain Construction 1. This finishes the proof of Theorem 1.7.

5.2 The pentagon case — stability of Construction 2

We proceed very similarly as in the proof of Theorem 1.7, but this time, the arguments are tailored to Construction 2. The graph G has less than $0.22n^2$ red edges so Lemma 3.4 yields existence of at least $\delta_2 \cdot n^3$ triangles in G. Without loss of generality, we may assume $\varepsilon \ll \delta_2$. As in the previous subsection, we will use two constants $\varepsilon_{\mathrm{RL}} > 0$ and $\delta_{\mathrm{RL}} > 0$, and we assume they obey the hierarchy $\delta \ll \delta_{\mathrm{RL}} \ll \varepsilon_{\mathrm{RL}} \ll \varepsilon$.

Let C_4^X be the red/blue-colored 4-cycle with exactly one blue edge $\{u, v\}$ and a pendant red edge adjacent neither to u nor to v, and $\mathcal{P}_5$ the set of all the ten possible red/blue-colorings of the 5-vertex path; see also Figure 7. The moreover part of Proposition 3.2 yields the following lemma.

Lemma 5.7. *If $\phi \in \mathrm{Hom}^+(\mathcal{A}_{\mathrm{C5}}, \mathbb{R})$ that satisfy ϕ* () $= \frac{(2+\sqrt{2})}{8}$ *and ϕ* () $= \frac{2-\sqrt{2}}{8}$*, then $\phi(P) = 0$ for every $P \in \{C_4^X\} \cup \mathcal{P}_5$.*

Let n_0 be a large enough integer so that the graph G contains only $\delta n^{v(F)}$ copies of F for all $F \in \mathcal{F}_{\mathrm{C5}}$, and the limit identity proven by flag algebras in Proposition 3.2 holds with an error of order $O(\delta)$ for any graph with at least n_0 vertices. Set $\mathcal{F}$ to be the family containing

- all the red/blue-colored triangles with at least one blue edge,
- all the 4-vertex red/blue-colored graphs that contain a copy of B_3^+,
- all the 5-vertex red/blue-colored graphs that contain a copy of B_5,
- the red/blue-colored graph C_4^X and the ten elements of $\mathcal{P}_5$.

Let δ_{RL} be the constant from Theorem 5.1 applied with the constant $\varepsilon_{\mathrm{RL}}$ and the family $\mathcal{F}$. Since $\delta \ll \delta_{\mathrm{RL}}$, by induced removal lemma there is a graph G' differing from G on at most $\varepsilon_{\mathrm{RL}} \cdot n^2$ pairs that has no induced copy of F for all $F \in \mathcal{F}$. Clearly, G' has $(1/4 \pm 2\varepsilon_{\mathrm{RL}})\, n^2$ edges. It remains to show that G' is $\left(\varepsilon/2 \cdot n^2\right)$-close to Construction 2.

We begin with partitioning the vertices of G' into three parts X, Y, Z based on their distance to vertices incident to blue edges. Let X be the set of vertices of G' that are incident to at least one blue edge, Y the vertices that are incident only to red edges and have at least one neighbor in X, and Z the vertices of G' that are neither in X nor in Y. We define H to be the subgraph of G' induced by $X \cup Y$. Furthermore, let $X_0 \subseteq X$ be the set of all the vertices x such that the connected component of G' containing x contains no vertex from Z. Analogously, $Y_0 \subseteq Y$ are all the vertices such that their connected component does not contain any vertex from Z. Set $X_1 := X \setminus X_0$ and $Y_1 := Y \setminus Y_0$. Having in mind the aim is to prove that G' is close to Construction 2, we proceed with the following series of claims that describe the structure of G'.

Claim 5.8. *The graph H is bipartite.*

Proof. Suppose for contradiction H contains an odd cycle. Since H does not contain any induced P_5, H contains either a triangle, or an induced pentagon. In both cases, all the edges of the cycle must be red.

First, suppose that H contains a triangle u, v, w. If at least one of the three vertices is incident to a blue edge, we would have found a copy of B_3^+, which is not possible. Therefore, $\{u, v, w\} \subseteq Y$. Let $x_u \in X$ be a neighbor of u. If v would be a neighbor of x_u as well, then u, v, x_u and a blue edge going out from x_u would create a copy of B_3^+. Therefore, $\{x_u, v\}$ is not an edge. By the same reasoning, $\{x_u, w\}$ is not an edge and the vertex v has neighbor $x_v \in X$ such that neither $\{x_v, u\}$ nor $\{x_v, w\}$ are edges. Now let $x \in X$ be a vertex connected to x_u by a blue edge. Since H is B_3-free, x is not a neighbor of u, and since H is B_5-free, x is neither a neighbor of v nor x_v nor w. The path x, x_u, u, v, x_v cannot be induced and therefore there is an edge between x_u and x_v. But then the vertices w, v, x_v, x_u, x span an induced P_5, which is a contradiction. For the rest of the proof, we will assume that H is triangle-free.

Now suppose H has an induced pentagon u_1, u_2, u_3, u_4, u_5 so that one of its vertices, say u_1, is incident to a blue edge. Let $x_1 \in X$ be one of the neighbors of u_1 that is joined to u_1 by a blue edge. If x_1 would be joined by an edge either to u_2 or u_5, then we have found a copy of B_3. Since H is also B_5-free, the vertex x_1 cannot be joined by an edge to u_3 or u_4. Therefore, x_1, u_1, u_2, u_3, u_4 is an induced path of length four, a contradiction.

Finally, suppose there is an induced pentagon u_1, u_2, u_3, u_4, u_5 such that all the edges incident to the five vertices are red. The vertex u_1 must have a neighbor, say x_1, that is incident to a blue edge. We already know that H is triangle-free, so x_1 is adjacent neither to u_2, nor to u_5. Also, if x_1 would be a neighbor of u_3, then u_1, x_1, u_3, u_4, u_5 is a 5-cycle with one endpoint incident to a blue edge, which we already excluded in the previous paragraph. Analogously, x_1 is not adjacent to u_4, and hence x_1, u_1, u_2, u_3, u_4 is an induced path of length four; a contradiction. □

Claim 5.9. *Z has size at least $\delta_2/2 \cdot n$.*

Proof. As in Claim 5.4, the number of triangles in G' is at least $(\delta_2 - \varepsilon_{\mathrm{RL}})\, n^3$. Since every triangle has at least one vertex in Z, $|Z| \geq (\delta_2 - \varepsilon_{\mathrm{RL}}) \cdot n > \delta_2/2 \cdot n$. □

Let H_1 be the subgraph of H induced by X_1. We continue in our exposition and find a good bipartition of H_1.

Claim 5.10. *If u and v are two vertices from X_1 that are joined by a blue edge, then at most one of the two vertices has a neighbor in Y_1.*

Proof. Suppose for contradiction there are two such vertices u and v. Since H is bipartite and has no induced P_5, at least one of the two vertices is within distance exactly two to a vertex $z \in Z$. Without loss of generality, let u be the vertex, and let y_u be the middle vertex on a shortest path between u and z.

Let $y_v \in Y_1$ be a neighbor of v. Since H is bipartite, neither y_u is a neighbor of v, nor y_v is a neighbor of u. Also, G' is B_5-free, hence the vertex z is not a neighbor of y_v, and by definition, there are no edges between Z and X_1. So either y_u and y_v are not joined by an edge and y_v, v, u, y_u, z induces a path, which contradicts that G' does not contain an induced P_5. Or, $\{y_u, y_v\}$ is an edge, but then the vertices induces C_4^X; a contradiction. □

Claim 5.11. *Let $u \in X_1$ and $v \in X_1$ be two vertices from the same connected component of H_1. If both u and v have a neighbor in Y_1, then there exists a vertex $w \in V(H_1)$ such that both $\{u, w\}$ and $\{w, v\}$ are edges in H_1.*

Proof. Analogously to the previous claim, we may assume that one of the two vertices, say u, has a neighbor $y \in Y_1$ such that y is adjacent to a vertex $z \in Z$. On the other hand, since $v \in X_1$, it must have a neighbor $t \in X_1$ such that $\{t, v\}$ is blue. By Claim 5.10, $t \neq u$. If $\{t, u\}$ or $\{v, y\}$ is an edge, we are done by letting $w := t$ or $w := y$, respectively. For the rest of the proof, we assume that neither $\{t, u\}$ nor $\{v, y\}$ is an edge. Also, Claim 5.10 yields that t has no neighbor in Y_1, so in particular, $\{t, y\}$ is not an edge.

Now we show that u is not adjacent to v. Suppose there is an edge between u and v. By Claim 5.10, the edge must be red. Recall that the vertex y has a neighbor $z \in Z$. There are no edges between Z and X_1 so the vertices t, v, u, y, z induces P_5, which is a contradiction.

Suppose there is no $w \in V(H_1)$ such that u, w, v is a path of length two. Since H_1 does not contain any induced path of length four, there exist vertices $x_u \in V(H_1)$ and $x_v \in V(H_1)$ such that u, x_u, x_v, v is a path of length three. The vertex t must be connected to x_u, as otherwise u, x_u, x_v, v, t is an induced P_5. However, H is bipartite so the path v, t, x_u, u, y must be induced; a contradiction. □

The last claim immediately yields the following corollary.

Corollary 5.12. *There exists a partition of the set X_1 into two parts A_1 and B_1 such that both A_1 and B_1 are independent sets in G', and there are no edges between A_1 and Y_1.*

This also implies that the set Y_1 must be independent.

Claim 5.13. *The set Y_1 is an independent set in G'.*

Proof. Suppose there is an edge between two vertices $u \in Y_1$ and $v \in Y_1$. By definition, there exist two vertices $b_u \in B_1$ and $b_v \in B_1$ that are adjacent to u and v, respectively. The two vertices are distinct and none of them can be adjacent to both u and v. Let $a \in A_1$ be a neighbor of b_u such that $\{a, b_u\}$ is a blue edge. The vertex a cannot be adjacent to b_v, which yields that a, b_u, u, v, b_v is an induced path of length four; a contradiction. □

Now let (A_0, B_0) be the color classes of an arbitrary 2-coloring of the bipartite graph induced by $X_0 \cup Y_0$. We define the following four sets that partition the set $V(G')$: $A := A_0 \cup A_1$, $B := B_0 \cup B_1$, $C := Y_1$, and $D := Z$. Claims 5.8-5.13 yield that G' must have the following structure.

Corollary 5.14. *$\{A, B, C, D\}$ is a partition of the vertex-set of G', the sets A, B and C are independent sets in G', and every edge e of G' goes either between A and B, or B and C, or C and D, or inside D. Moreover, if e is blue, then e must go between A and B.*

Let $a := |A|/n$, $b := |B|/n$, $c := |C|/n$ and $d := |D|/n$. The edge-density of G' can be upper-bounded by $f(a, b, c, d) := 2ab + 2bc + 2cd + d^2$. Let us analyze the maximum value of f under constraints on a, b, c and d that we have already established. All of that is summarized in the following optimization problem:

$$\begin{aligned}
\textbf{maximize: } & 2ab + 2bc + 2cd + d^2 \\
\textbf{subject to: } & a \geq 0, \quad b \geq 0, \quad c \geq 0, \\
& d = 1 - a - b - c, \\
& ab \geq \left(2 - \sqrt{2}\right)/16, \\
& d \geq \delta_2/2.
\end{aligned}$$

Clearly, if the values of a, b, c and d are equal to those coming from Construction 2, then $f(a, b, c, d) = 1/2$. The following proposition shows that there is no other point $(a', b', c', d') \in \mathbb{R}^4$ that would satisfy the constraints and also attain the value $1/2$.

Claim 5.15. *The optimization problem has a unique solution at the point*

$$(a_m, b_m, c_m, d_m) = \left(\frac{1}{2} - \frac{\sqrt{2}}{4}, \frac{1}{4}, \frac{1}{4}, \frac{\sqrt{2}}{4}\right),$$

Moreover, if a point (a', b', c', d') satisfies all the constraints and $f(a', b', c', d')$ is close to $1/2$, then (a', b', c', d') is close to (a_m, b_m, c_m, d_m).

This claim immediately yields that G' is close to Construction 2, hence proving the claim finishes the proof of Theorem 1.6

Proof of Claim 5.15. Let $(a_0, b_0, c_0, d_0) \in \mathbb{R}^4$ be a point that satisfies the constraints and maximizes the objective function. In particular, $f(a_0, b_0, c_0, d_0) \geq 1/2$.

First, we show that $a_0 b_0 = \left(2 - \sqrt{2}\right)/16$. If $a_0 b_0 > \left(2 - \sqrt{2}\right)/16$, then let $\alpha := a_0 - \frac{2-\sqrt{2}}{16 \cdot b_0}$. It follows that

$$f(a_0 - \alpha, b_0, c_0 + \alpha, d_0) = f(a_0, b_0, c_0, d_0) + \alpha \cdot d_0,$$

and the point (a_0, b_0, c_0, d_0) was not an optimal solution. Next, we bound b_0 away from 1/2. Observe that

$$f(a,b,c,d) = 2ab + 2(b+d)(1-a-b-d) + d^2 = 2b(1-b) - d^2 + 2d(1-a-2b).$$

Therefore,

$$(1-a_0-2b_0) = \left(1 - \frac{2-\sqrt{2}}{16\cdot b_0} - 2b_0\right) > 0,$$

as otherwise $f(a_0,b_0,c_0,d_0) \le 1/2 - {d_0}^2 \le 1/2 - (\delta_2/2)^2 < 1/2$ contradicting $f(a_0,b_0,c_0,d_0) \ge 1/2$. Thus,

$$b_0 < \frac{2+2^{3/4}}{8} < 0.4603.$$

Moreover, the maximum value of $2b_0 \cdot (1-b_0)$ is at most $\left(6-\sqrt{2}+2^{7/4}\right)/16 < 0.497$. On the other hand, the maximum value of $\left(1-\frac{2-\sqrt{2}}{16\cdot b_0} - 2b_0\right)$ is at most $1-\sqrt{1-1/\sqrt{2}} < 0.46$. Since $f(a_0,b_0,c_0,d_0) \ge 1/2$, we conclude that $d_0 > 0.003$.

Suppose now that $b_0 \neq c_0$. If $b_0 < c_0$, then

$$f(a_0,b_0,c_0-\alpha,d_0+\alpha) - f(a_0,b_0,c_0,d_0) = 2(c_0-b_0)\alpha + \alpha^2,$$

where $\alpha = c_0 - b_0$; a contradiction. On the other hand, if $c_0 < b_0$, then

$$f(a_0,b_0,c_0+\alpha,d_0-\alpha) - f(a_0,b_0,c_0,d_0) = 2(b_0-c_0)\alpha - \alpha^2 \ge \alpha^2,$$

where this time $\alpha = \min(b_0-c_0, d_0-0.003)$. We conclude that

$$f(a_0,b_0,c_0,d_0) = \frac{2-\sqrt{2}}{8} + 2{c_0}^2 + 2c_0\cdot\left(1-\frac{2-\sqrt{2}}{16\cdot c_0} - 2c_0\right) + \left(1-\frac{2-\sqrt{2}}{16\cdot c_0} - 2c_0\right)^2. \quad (4)$$

Since swapping the values of a_0 and b_0 changes the objective function by $c_0(a_0-b_0)$, it holds that $b_0 \ge a_0$. In particular, $c_0 = b_0 \ge \left(\sqrt{2-\sqrt{2}}\right)/4 > 0.19$. The right-hand side of (4) depends only on c_0 and $c_0 \in [0.19, 0.4603]$. It is straightforward to check that in this range, the value of (4) is at most 1/2, and the unique point where the value is attained is $c_0 = 1/4$. Therefore, $b_0 = 1/4$, $a_0 = \frac{2-\sqrt{2}}{4}$ and $d_0 = \frac{\sqrt{2}}{4}$. By continuity of $f(a,b,c,d)$ and compactness of $[0,1]^4$, it also follows that if $f(a',b',c',d')$ is close to 1/2, then (a',b',c',d') is close to (a_0,b_0,c_0,d_0). □

6 Exact result for pentagons

For a graph G, we define $\mathcal{C}_5(G)$ to be the set of all edges of G that occur in a copy of C_5 in G. In other words,

$$\mathcal{C}_5(G) = \bigcup_{H\subseteq G, H\cong C_5} E(H).$$

Let $\mathcal{E}_n$ be the set of all n-vertex graphs with exactly $\lfloor n^2/4 \rfloor + 1$ edges, and define

$$F(n) := \min_{G\in\mathcal{E}_n} |\mathcal{C}_5(G)|.$$

For convenience, we set $\widetilde{F}(n) := \lfloor n^2/4 \rfloor + 1 - F(n)$.

Next, let $\mathcal{E}'_n$ be the set of all n-vertex graphs with at least $\lfloor n^2/4 \rfloor + 1$ edges. It immediately follows that for any $G \in \mathcal{E}'_n$, it holds $|\mathcal{C}_5(G)| \geq F(n)$, and if $|\mathcal{C}_5(G)| = F(n)$, then $G \in \mathcal{E}_n$. Finally we define $\mathcal{G}_n \subseteq \mathcal{E}'_n$ to be the set of all $G \in \mathcal{E}'_n$ with $|\mathcal{C}_5(G)| = F(n)$.

We call a quadruple of non-negative integers (a, b, c, d) n-extremal, if the following is satisfied:

- $a + b + c + d = n$,
- $a \cdot b = \widetilde{F}(n)$, and
- $a \cdot b + b \cdot c + c \cdot d + \binom{d}{2} > \frac{n^2}{4}$.

The main theorem of this section is the following:

Theorem 6.1. *There exists an integer n_0 such that the following holds for any $n \geq n_0$. If $G \in \mathcal{G}_n$, then $V(G)$ can be partitioned into four sets A, B, C and D such that*

- *the quadruple $(|A|, |B|, |C|, |D|)$ is n-extremal,*
- *A, B and C are independent sets of G,*
- *$\{u, v\} \in E(G)$ for any $u \in A$ and $v \in B$,*
- *$\{u, v\} \notin E(G)$ for any $u \in A$ and $v \in C \cup D$, and*
- *$\{u, v\} \notin E(G)$ for any $u \in B$ and $v \in D$.*

An immediate consequence of this theorem is that (a, b, c, d) is n-extremal if and only if it solves the following integer quadratic program:

$$\begin{aligned}
\textbf{maximize:}\quad & a \cdot b \\
\textbf{subject to:}\quad & a \in \mathbb{N}, \quad b \in \mathbb{N}, \quad c \in \mathbb{N}, \quad d \in \mathbb{N}, \\
& a \cdot b + b \cdot c + c \cdot d + \binom{d}{2} > \frac{n^2}{4}, \\
& a + b + c + d = n.
\end{aligned}$$

Since the exact solution of this maximization problem for a given integer n depends on errors in rounding expressions like $\sqrt{2}n/4$, we leave it in this form. Approximate values of a, b, c and d are indeed given by Construction 2.

Proof of Theorem 6.1. Theorems 1.3 and 1.6 immediately yield that for any $\varepsilon > 0$, there exists an integer n_0 so that if $n \geq n_0$, then by adding or removing εn^2 edges in a graph $G \in \mathcal{G}_n$ we obtain the graph from Construction 2. Moreover, the value of n_0 will be large enough so that Construction 2 yields that $F(n) = ((2 + \sqrt{2})/16 \pm \varepsilon)n^2$ and $\widetilde{F}(n) = ((2 - \sqrt{2})/16 \pm \varepsilon)n^2$ for every $n \geq n_0$.

Fix an integer $n \geq n_0$ and any graph $G \in \mathcal{G}_n$, and let V be the vertex-set of G. Clearly, for any $\varepsilon' > 0$, we can find $\varepsilon > 0$ small enough so that V can be partitioned into five sets A_0, B_0, C_0, D_0 and X such that

- $|A_0| = (1/2 - \sqrt{2}/4 \pm \varepsilon') \cdot n$, $|B_0| = (1/4 \pm \varepsilon') \cdot n$, $|C_0| = (1/4 \pm \varepsilon') \cdot n$, $|D_0| = (\sqrt{2}/4 \pm \varepsilon') \cdot n$,
- $0 \leq |X| \leq \varepsilon' \cdot n$,
- $\deg_{A_0}(u) \geq (1 - \varepsilon')|A_0|$ for every $u \in B_0$,
- $\deg_{B_0}(u) \geq (1 - \varepsilon')|B_0|$ for every $u \in A_0 \cup C_0$,
- $\deg_{C_0}(u) \geq (1 - \varepsilon')|C_0|$ for every $u \in B_0 \cup D_0$,
- $\deg_{D_0}(u) \geq (1 - \varepsilon')|D_0|$ for every $u \in C_0$, and
- the induced subgraph $G[D_0]$ has edge-density at least $1 - \varepsilon'$.

In other words, the stability result from Section 5 yields an approximate structure of G. In the following series of claims, we will show that the extremality of G allows us to "clean up" this description to the one claimed in the statement of the theorem. For the rest of the proof, we will assume $\varepsilon' > 0$ is sufficiently small ($\varepsilon' < 10^{-4}$ would be sufficient).

For a set $S \subseteq V$, we denote by $E(S)$ the set of edges of the subgraph induced by S, i.e., $E(S) = E(G[S])$. For two disjoint $X, Y \subseteq V$, we denote by $E(X, Y)$ the set of edges in G with exactly one endpoint in X and the other endpoint in Y.

First, let us observe that every graph with more than $n^2/4$ edges has at most $\widetilde{F}(n)$ edges that do not occur in C_5.

Claim 6.2. *There is no n-vertex graph $G' \in \mathcal{E}'_n$ with $|E(G') \setminus \mathcal{C}_5(G')| > \widetilde{F}(n)$.*

Proof. Indeed, otherwise remove from G' arbitrarily chosen $|E(G')| - \lfloor n^2/4 \rfloor - 1$ edges in $\mathcal{C}_5(G')$. The obtained graph has less than $\lfloor n^2/4 \rfloor + 1 - \widetilde{F}(n) = F(n)$ edges that occur in C_5, a contradiction. □

We continue with three simple claims that all the edges between the parts B_0 and C_0, C_0 and D_0, and inside D_0 occur in a copy of C_5.

Claim 6.3. $E(B_0, C_0) \subseteq \mathcal{C}_5(G)$.

Proof. Fix any $\{u, v\} \in E(B_0, C_0)$ with $u \in B_0$. Let $v' \in C_0$ be a neighbor of u in C_0 different from v. Since both v and v' have more than $|D_0|/2$ neighbors in D_0 and the edge-density of $G[D_0]$ is $(1 - \varepsilon')$, there exist a vertex $w \in D_0$ connected to v, and vertex $w' \in D_0$ connected to v' such that $\{w, w'\}$ is an edge. Therefore, $uvww'v'$ forms a C_5 in G. □

Claim 6.4. $E(C_0, D_0) \subseteq \mathcal{C}_5(G)$.

Proof. Fix any $\{u, v\} \in E(C_0, D_0)$ with $u \in C_0$. Let $v' \in D_0$ be a neighbor of u in D_0 different from v. Since the edge-density of $G[D_0]$ is $(1 - \varepsilon')$, there exists a path of length three between v and v' in $G[D_0]$. This path together with the edges $\{u, v\}$ and $\{u, v'\}$ forms a C_5. □

Claim 6.5. $E(G[D_0]) \subseteq \mathcal{C}_5(G)$.

Proof. Let u and v be two adjacent vertices from D_0. Since the edge-density of $G[D_0]$ is $(1 - \varepsilon')$, there is a path of length four between u and w, which together with $\{u, v\}$ forms a C_5. □

Since $|E(B_0, C_0) \cup E(C_0, D_0) \cup E(G[D_0])| \geq (2 + \sqrt{2} - 3\varepsilon')n^2$, we immediately conclude that

Corollary 6.6. $|E(A_0, B_0) \cap \mathcal{C}_5(G)| < 5\varepsilon' n^2$.

Let $E' := E(A_0, B_0) \cup E(B_0, C_0) \cup E(C_0, D_0) \cup E(G[D_0])$. Since most of the edges between A_0 and B_0 do not occur in any C_5, we now get much better control on the edges in $E \setminus E'$. We start with the following two claims.

Claim 6.7. *There is no vertex $z \in V$ adjacent both to $u \in A_0$ and $v \in B_0$.*

Proof. Suppose for contradiction there is such a vertex z, and let $u \in A_0$ and $v \in B_0$ be its two neighbors. Let $v' \in B_0$ be a neigbor of u different from v. Clearly, there are at least $(1 - \varepsilon')|A_0|$ ways of choosing v'. The vertices v and v' have at least $\deg_{A_0}(v) + \deg_{A_0}(v') - |A_0| \geq (1 - 2\varepsilon')|A_0|$ common neigbors $u' \in A_0$. Each such choice of u' and v' yields a copy of C_5 on the vertices $uzvu'v'$. In particular the edge $\{u', v'\} \in E(A_0, B_0) \cap \mathcal{C}_5(G)$. However, there are at least $(1 - 3\varepsilon')|A_0||B_0| > 0.03n^2$ choices of $\{u', v'\}$, which contradicts Corollary 6.6. □

Claim 6.8. *There is no vertex $z \in V$ adjacent both to $u \in B_0$ and $v \in C_0$.*

Proof. Suppose not, and let $u \in B_0$ and $v \in C_0$ be two neighbors of z. Let $u' \in B_0$ be any of the $(1 - \varepsilon')|B_0|$ neighbors of v different from u. Since $\deg_{A_0}(u) + \deg_{A_0}(u') - |A_0| \geq (1 - 2\varepsilon')|A_0|$, there are at least $(1 - 2\varepsilon')|A_0| \cdot (1 - \varepsilon')|B_0| > (1 - 3\varepsilon')|A_0||B_0|$ edges from $E(A_0, B_0)$ that occur in a C_5 (note that $uzvu'w$, where $w \in A_0$ is a common neighbor of u and u', form a C_5); a contradiction. □

A direct consequence of the last two claims is the following.

Corollary 6.9. *The sets A_0, B_0 and C_0 are independent, and $|E(A_0, C_0)| = |E(B_0, D_0)| = 0$.*

Now move our attention to paths of length at most two between A_0 and D_0. Let $Y \subseteq A_0$ be the set of vertices $u \in A_0$ such that there exist vertices $v \in V$ and $w \in D_0$ such that both $\{u, v\} \in E(G)$ and $\{v, w\} \in E(G)$.

Claim 6.10. $|Y| < 21\varepsilon' n$.

Proof. For each edge $\{y, v\}$ with $y \in Y$ and $v \in B_0$, consider vertices $z \in V \setminus \{y, v\}$ and $x \in N_{D_0}(z)$ such that yzx is a 3-vertex path in G. Note that such a path exists by the definition of Y. Since $|N_{C_0}(x) \cap N_{C_0}(v)| > |C_0|/2$, we conclude that $\{y, v\} \in \mathcal{C}_5(G)$. The edge $\{y, v\}$ can be chosen in at least $|Y| \cdot (1 - \varepsilon')|B_0|$ ways, so by Corollary 6.6 we conclude that

$$|Y| \leq \frac{5\varepsilon' n^2}{(1 - \varepsilon')|B_0|} < \frac{20\varepsilon' n}{1 - 2\varepsilon'} < 21\varepsilon' n.$$

□

We set $A'_0 := A_0 \setminus Y$ and $Z := X \cup Y$. We continue our exposition by establishing a lower bound on the minimum degree of G. We start with the following claim.

Claim 6.11. *There exists a vertex $u \in A'_0$ incident to at least $(1/4 - 191\varepsilon')n$ edges not in $\mathcal{C}_5(G)$, and a vertex $u' \in B_0$ incident to at least $((2 - \sqrt{2})/4 - 93\varepsilon')n$ edges that are not in $\mathcal{C}_5(G)$.*

Proof. There are at least $\widetilde{F}(n) - |Z|n \geq ((2 - \sqrt{2})/16 - 23\varepsilon')n^2$ edges between A'_0 and B_0 that do not occur in C_5. Since $|A'_0| \leq ((2 - \sqrt{2})/4 + \varepsilon')n$, there is a vertex $u \in A'_0$ incident to at least $(1/4 - 191\varepsilon')n$ such edges. Similarly, $|B_0| \leq (1/4 + \varepsilon')n$, which implies existence of $u' \in B_0$ incident to at least $((2 - \sqrt{2})/4 - 93\varepsilon')n$ edges in $E(G) \setminus \mathcal{C}_5(G)$. □

Claim 6.12. *For any $v \in V$, $\deg(v) \geq (1/4 - 191\varepsilon')n$.*

Proof. Otherwise consider the graph G' obtained from G by removing the vertex v and cloning the vertex u from the previous claim. G' has more edges that do not occur in C_5 than G and also $|E(G')| > |E(G)|$, a contradiction with Claim 6.2. □

Corollary 6.13. *There is no vertex $z \in Z$ such that $N(z) \subseteq A_0' \cup Z$.*

Proof. Indeed, any such z would have

$$\deg(z) \leq |A_0'| + |Z| = |A_0| + |X| \leq ((2-\sqrt{2})/4 + 2\varepsilon')n < 0.15n < (1/4 - 191\varepsilon')n,$$

which contradicts the previous claim. □

Now we are ready to split the vertices $z \in Z$ based on their adjacencies to the vertices outside of Z. First, let $Z' := \{z \in Z \mid \exists u \in B_0 : \{z,u\} \in E(G)\}$. Claims 6.7 and 6.8 yield that no vertex $z \in Z'$ has a neighbor in $A_0 \cup C_0$. We define

$$C_1 := \{z \in Z' \mid \exists v \in V \wedge \exists w \in D_0 : \{z,v\} \in E(G) \wedge \{v,w\} \in E(G)\},$$

and $A_1 := Z' \setminus C_1$. Note that $Y \subseteq C_1$, and if $z \in Z'$ has a neighbor in D_0, then $z \in C_1$. Let us first focus on the set A_1.

Claim 6.14. *For all $z \in A_1$, $\deg_{B_0}(z) > (1 - 214\varepsilon')|B_0|$.*

Proof. If there exist a vertex $z \in A_1$ with $\deg_{B_0}(z) \leq (1 - 214\varepsilon')|B_0|$, then its total degree in G is at most

$$\deg_{B_0}(z) + \deg_Z(z) < (1 - 214\varepsilon')(1/4 + \varepsilon')n + |Z| \leq (1/4 - 191\varepsilon')n,$$

a contradiction. □

Corollary 6.15. *A_1 is an independent set in G.*

Proof. If there is an edge in A_1, then this edge together with any edge in $E(A_0, B_0)$ are in a C_5, contradicts Corollary 6.6. □

We set $A := A_0' \cup A_1$, and continue our exposition by analyzing the vertices $B_1 := \{z \in Z \mid \exists u \in A : \{z,u\} \in E(G)\}$. Note that $B_1 \cap Z' = \emptyset$. By Claim 6.7 and definitions of the sets A_0' and A_1, we conclude that both $|E(B_1, B_0)| = 0$ and $|E(B_1, D_0)| = 0$. In the following two claims, we study the edges between B_1 and $C_0 \cup A$.

Claim 6.16. *For every $v \in B_1$, $\deg_{C_0}(v) > n/10$.*

Proof. We know that $v \in B_1$ can be adjacent only to the vertices from $A_0 \cup C_0 \cup X$. Therefore, Claim 6.12 yields that v has at least

$$(1/4 - 191\varepsilon')n - |A_0| - |X| \geq (\sqrt{2} - 1 - 193\varepsilon')n > n/10$$

neighbors in C_0. □

Claim 6.17. *For every $v \in B_1$, $\deg_A(v) \geq (1 - 800\varepsilon')|A|$.*

Proof. First observe that any edge $\{v, w\}$ with $w \in C_0$ is contained in $\mathcal{C}_5(G)$. Indeed, consider a neighbor $w' \in N_{C_0}(v) \setminus \{w\}$ and two adjacent vertices $x \in N_{D_0}(w)$ and $x' \in N_{D_0}(w')$.

Suppose for contradiction $\deg_A(v) < (1 - 800\varepsilon')|A| < |A| - 116\varepsilon' n$. In particular, v is incident to at most

$$\deg_A(v) + \deg_Z(v) < |A| - 94\varepsilon' n \leq \frac{2 - \sqrt{2}}{4} - 93\varepsilon' n$$

edges not in $\mathcal{C}_5(G)$. Let $u' \in B_0$ be the vertex from Claim 6.11 with at least $\left((2 - \sqrt{2})/4 - 93\varepsilon'\right) n$ incident edges that are not in $\mathcal{C}_5(G)$. Moreover, $\deg(u') \geq |A| + |C_0| - 2\varepsilon' n$. Therefore, removing the vertex v and adding a clone of the vertex u' yield an n-vertex graph with more than $n^2/4$ edges that contradicts Claim 6.2. ☐

Corollary 6.18. *B_1 is an independent set of G.*

Proof. If there is an edge in B_1, then this edge together with any edge in $E(A_0, B_0)$ are in a C_5, contradicts Corollary 6.6. ☐

Let $B := B_0 \cup B_1$. Recall the definition of C_1, i.e.,

$$C_1 = \{u \in Z \mid \deg_{B_0}(u) \geq 1 \wedge \exists v \in V, w \in D_0 : \{u, v\} \in E(G) \wedge \{v, w\} \in E(G)\}.$$

By Claim 6.8, there are no edges between C_1 and C_0, and by the definition of B_1, there are no edges between C_1 and A. Let us now show that vertices $v \in C_1$ have many neighbors in B_0.

Claim 6.19. *For every $u \in C_1$, $\deg_{B_0}(u) \geq n/25$.*

Proof. As noted above, u can be adjacent only to the vertices in $B_0 \cup D_0 \cup Z$. First observe that for every $z \in N_{D_0}(u)$, the edge $\{u, z\} \in \mathcal{C}_5(G)$. Indeed, let $x \in N_{B_0}(u)$ and $y \in N_{C_0}(x)$ be chosen arbitrarily. Since $y \in C_0$ and $z \in D_0$, there exist a common neighbor of y and z which encloses a C_5. Analogously, we show $\{u, x\} \in \mathcal{C}_5(G)$ for every $x \in N_{B_0}(u)$. Consider the vertices $v \in V$ and $w \in D_0$ with $\{u, v\} \in E(G)$ and $\{v, w\} \in E(G)$ witnessing that $u \in C_1$. Since $x \in B_0$ and $w \in D_0$, the two vertices must have a common neighbor which yields $\{u, x\} \in \mathcal{C}_5(G)$.

The last paragraph shows that u is incident to at most $|Z| < 22\varepsilon' n$ edges that do not occur in C_5. On the other hand, $|Z \cup D_0| < (\sqrt{2}/4 + 23\varepsilon')n$. So if $\deg_{B_0}(v) < n/25$, then

$$\deg(v) < 0.396n < \deg(u'),$$

where $u' \in B_0$ is the vertex from Claim 6.11. Therefore, the graph obtained by removing the vertex u and cloning the vertex u' contradicts Claim 6.2. ☐

Corollary 6.20. *C_1 is an independent set in G.*

Proof. Suppose for a contradiction there is an edge $\{u, u'\}$ with $u, u' \in C_1$. There are at least

$$\deg_{B_0}(u) \cdot \left(|A_0| - \varepsilon' n\right) > \frac{n}{25} \cdot \left(\frac{n}{4} - 23\varepsilon' n\right) > \frac{n^2}{101}$$

edges $\{v, w\}$ with $v \in N_{B_0}(u)$ and $w \in N_{A_0}(v)$. However, the vertices w and u' have a common neighbor in $B_0 \setminus \{v\}$ and hence $|E(A_0, B_0) \cap \mathcal{C}_5(G)| > n^2/101$; a contradiction with Corollary 6.6. ☐

We define $C := C_0 \cup C_1$, and $D_1 := Z \setminus (A_1 \cup B_1 \cup C_1)$. By the definition of the sets A, B_1 and C_1, every vertex $v \in D_1$ has no neigbors in $A \cup B_0$. We now concentrate on the edges between D_1 and C_0.

Claim 6.21. *For every $v \in D_1$, $\deg_{C_0}(v) \geq n/25$.*

Proof. The vertex v can be adjacent only to the vertices in $C_0 \cup D_0 \cup Z$, and clearly every edge $\{v, w\}$ with $w \in C_0 \cup D_0$ occurs in C_5. In particular, v is incident to at most $22\varepsilon' n$ edges that do not occur in C_5.

As in Claim 6.19, if $\deg_{C_0}(v) < n/25$ then $\deg(v) < |D_0| + |Z| + n/25 < 0.396n$. Therefore, removing the vertex v and cloning the vertex u' from Claim 6.11 result in a graph contradicting Claim 6.2. □

So the only possible edges that could be in G but not following the pattern of Construction 2 are those between B_1 and D_1. We rule them out in the following claim.

Claim 6.22. $|E(B_1, D_1)| = 0$.

Proof. Suppose for contradiction there is an edge $\{u, v\}$ with $u \in B_1$ and $v \in D_1$. Since any vertex $x \in B_0$ has $\deg_{C_0}(x) > |C_0| - n/25$, the vertices v and x have a common neighbor and hence

$$|E(A_0, B_0) \cap \mathcal{C}_5(G)| \geq \deg_{A_0}(u) \cdot (1 - \varepsilon')|B_0| > 0.03n^2,$$

which indeed contradicts Corollary 6.6. □

Let $D := D_0 \cup D_1$. Putting everything together, we conclude that the edges in G are as in Construction 2.

Corollary 6.23. $V(G) = A \,\dot\cup\, B \,\dot\cup\, C \,\dot\cup\, D$ *and* $E(G) \subseteq E(A, B) \,\dot\cup\, E(B, C) \,\dot\cup\, E(C, D) \,\dot\cup\, \binom{D}{2}$.

In particular, the set of edges $E(G) = \mathcal{C}_5(G) \,\dot\cup\, E(A, B)$. Since G is minimizing $|\mathcal{C}_5(H)|$ among all graphs in $H \in \mathcal{E}'_n$, we immediately conclude the following.

Claim 6.24. $|E(A, B)| = |A||B| = \widetilde{F}(n)$.

Therefore, the quadruple $(|A|, |B|, |C|, |D|)$ is n-extremal which finishes the proof of the theorem. □

7 Exact result for longer odd cycles

As in the previous section, for an integer $k \geq 3$ and a graph G we define $\mathcal{C}_{2k+1}(G)$ to be the set of all edges of G that occur in a copy of C_{2k+1} in G. In other words,

$$\mathcal{C}_{2k+1}(G) := \bigcup_{H \subseteq G, H \cong C_{2k+1}} E(H).$$

Recall $\mathcal{E}_n$ and $\mathcal{E}'_n$ are the sets of all n-vertex graphs with exactly $\lfloor n^2/4 \rfloor + 1$ edges and at least $\lfloor n^2/4 \rfloor + 1$ edges, respectively. For any $k \geq 3$, let

$$F_{2k+1}(n) := \min_{G \in \mathcal{E}_n} |\mathcal{C}_{2k+1}(G)|,$$

and $\widetilde{F}_{2k+1}(n) := \lfloor n^2/4 \rfloor + 1 - F_{2k+1}(n)$. Finally we define $\mathcal{G}_n^{2k+1} \subseteq \mathcal{E}'_n$ to be the set of all $G \in \mathcal{E}'_n$ with $|\mathcal{C}_{2k+1}(G)| = F_{2k+1}(n)$. As we will show, for any $k \geq \ell \geq 3$, there exists a sufficiently large $n_0 := n_0(k)$ such that $\mathcal{G}_n^{2k+1} = \mathcal{G}_n^{2\ell+1}$ for all $n \geq n_0$.

Theorem 7.1. *For any integer $k \geq 3$ there exists an integer n_0 such that the following holds for any $n \geq n_0$. If $G \in \mathcal{G}_n^{2k+1}$, then $V(G)$ can be partitioned into four sets A, B, C and D such that*

- $|A| = \lfloor \frac{n-2}{6} \rfloor$, $|B| = \lfloor \frac{n+1}{6} \rfloor$, $|C| = 1$ *and* $|D| = \lfloor \frac{2n+1}{3} \rfloor$.
- *A and B are independent sets of G,*
- *$\{u, v\} \in E(G)$ for any $u \in A \cup C$ and $v \in B$,*
- *$\{u, v\} \notin E(G)$ for any $u \in A$ and $v \in C \cup D$, and*
- *$\{u, v\} \notin E(G)$ for any $u \in B$ and $v \in D$.*

In particular, $$F_{2k+1}(n) = \begin{cases} 2n^2/9 + 1 & \text{for } n \equiv 0 \mod 6, \\ 2n^2/9 + (n+13)/18 & \text{for } n \equiv 1 \mod 6, \\ 2n^2/9 - (n-22)/18 & \text{for } n \equiv 2 \mod 6, \\ 2n^2/9 + 1 & \text{for } n \equiv 3 \mod 6, \\ 2n^2/9 + (n+22)/18 & \text{for } n \equiv 4 \mod 6, \\ 2n^2/9 - (n-13)/18 & \text{for } n \equiv 5 \mod 6. \end{cases}$$

Proof. Let $V := V(G)$. For any $\varepsilon' > 0$ there is a choice of $\varepsilon < \varepsilon'$ and a large enough constant $n_0 \in \mathbb{N}$ such that if $n \geq n_0$, then the stability result proven in Theorem 1.7 and the fact $G \in \mathcal{G}_n^{2k+1}$ imply that there exists a set of at most $\varepsilon' \cdot n$ vertices C' such that the subgraph $G[V \setminus C']$ is disconnected, and it contains a connected component D' with at least $(2/3 - \varepsilon') \cdot n$ vertices and minimum degree at least $(2/3 - \varepsilon') \cdot n$. This follows because, after removing at most $(\varepsilon'/2) \cdot n$ exceptionally behaving vertices C'_1, any pair of two vertex-disjoint edges $e_1 = x_1 y_1$ and $e_2 = x_2 y_2$, where x_1 and x_2 are from the nearly clique part and y_1 and y_2 from the nearly complete bipartite part, yields that at least $n/7$ edges incident to y_i, where $i \in \{1, 2\}$, occur in some C_{2k+1}. However, the nearly clique part contains at least $(4/9 - \varepsilon'/15) \cdot n^2$ edges of G and each such an edge occurs in some C_{2k+1}. Therefore, there must be less than $(\varepsilon'/2) \cdot n$ vertices C'_2 such that the nearly clique part in G forms a connected component of $G[V \setminus (C'_1 \cup C'_2)]$. We conclude that there is a partition of V into four parts A', B', C' and D' that satisfies

- $|A'| = (1/6 \pm \varepsilon') \cdot n$, $|B'| = (1/6 \pm \varepsilon') \cdot n$, $|C'| < \varepsilon' \cdot n$, $|D'| = (2/3 \pm \varepsilon') \cdot n$,
- $\forall v \in A' : \deg_{B'}(v) \geq (1/6 - \varepsilon') \cdot n$,
- $\forall v \in B' : \deg_{A'}(v) \geq (1/6 - \varepsilon') \cdot n$,
- $\forall v \in D' : \deg(v) \geq (2/3 - \varepsilon') \cdot n$, and
- $E(A' \cup B', D') = 0$,

Note that these properties yield that the induced subgraph $G[D']$ has edge-density at least $1 - \varepsilon'$, and $|E(A', B')| \geq |A'||B'| - 4\varepsilon' \cdot n^2$.

We start our exposition with a direct analogue of Claim 6.2.

Claim 7.2. *There is no n-vertex graph $G' \in \mathcal{E}'_n$ with $|E(G') \setminus \mathcal{C}_{2k+1}(G')| > \widetilde{F}_{2k+1}(n)$.*

Proof. As otherwise removing from G' arbitrarily chosen $|E(G')| - \lfloor n^2/4 \rfloor - 1$ edges in $\mathcal{C}_{2k+1}(G')$ yields an n-vertex graph with less than $F_{2k+1}(n)$ edges that occur in C_{2k+1}, a contradiction. □

We continue by showing that both A' and B' are in fact independent sets in G.

Claim 7.3. *No $v \in V$ is adjacent to $u \in A'$ and $w \in B'$.*

Proof. As otherwise, we will actually show that almost every edge of G occurs in some C_{2k+1}.

Firstly, recall that all the edges of $G[D']$ occur in C_{2k+1}. If there would be a vertex v adjacent to $u \in A'$ and $w \in B'$, then we can find a copy of C_{2k+1} containing any given edge $\{u', w'\}$ with $u' \in A' \setminus \{u\}$ and $w' \in B' \setminus \{w\}$ in the following way: let $u_0 \in A'$ be an arbitrary common neighbor of w and w', and let P be a $(2k-3)$-vertex path between u and u' disjoint from w, w' and u_0. Note that such a path exists because every vertex in A has at least $|B| - 2\varepsilon' n$ neighbors in B, and symmetrically every vertex in B has at least $|A| - 2\varepsilon' n$ neighbors in A. Therefore, $vwu_0w'P$ is a copy of C_{2k+1} in G containing the edge $\{u', w'\}$. It follows that $|E(G) \setminus \mathcal{C}_{2k+1}(G)| \leq \varepsilon' n^2$, which clearly contradicts the fact that $G \in \mathcal{G}_n^{2k+1}$. □

Corollary 7.4. *A' and B' are independent sets in G.*

Let $C \subseteq V$ be a minimum-size set so that $G - C$ is disconnected and one of its connected components is a bipartite graph (A, B) with minimum degree at least $n/7$. Clearly, this is well defined because C' has the bipartite graph (A', B') as one of the components. Moreover, among all such cuts C of the minimum size, we choose such a C that $|A| + |B|$ is as large as possible.

Let $D := V \setminus (A \cup B \cup C)$. Because we already have a partition (A', B', C', D') with (A', B') being a bipartite graph and D' of edge density at least $1 - \varepsilon'$, one can easily see that the partition $A \cupdot B \cupdot C \cupdot D$ of V behaves very similarly to the original partition $A' \cupdot B' \cupdot C' \cupdot D'$. In particular,

- $|A| = (1/6 \pm 2\varepsilon') \cdot n$,
- $|B| = (1/6 \pm 2\varepsilon') \cdot n$,
- $|C| < \varepsilon' \cdot n$,
- $|D \setminus D'| < \varepsilon' \cdot n$, and
- $E(D) \subseteq \mathcal{C}_{2k+1}(G)$.

Following the proof of Claim 7.3 we get also that there is no vertex $v \in V$ adjacent to $u \in A$ and $w \in B$.

Now let use an argument analogous to the one used in Section 6 to show that G must have a large minimum degree.

Claim 7.5. *There is a vertex $v \in A$ that is incident to at least $n/6 - 9\varepsilon' \cdot n$ edges not in $\mathcal{C}_{2k+1}(G)$.*

Proof. Suppose not, then the number of edges not in $\mathcal{C}_{2k+1}(G)$ is smaller than

$$|A| \cdot \left(\frac{n}{6} - 9\varepsilon' \cdot n\right) + |C| \cdot n \leq \left(\frac{n}{6} + 2\varepsilon' \cdot n\right) \cdot \left(\frac{n}{6} - 9\varepsilon' \cdot n\right) + \varepsilon' \cdot n^2 < n^2/36 - \varepsilon'/6 \cdot n^2.$$

Therefore, there are more than $2n^2/9 + \varepsilon' n^2/6$ edges in $\mathcal{C}_{2k+1}(G)$, contradicting the extremality of G since Construction 1 has at most $2n^2/9 + (n+22)/18$ edges that occur in C_{2k+1}. □

Corollary 7.6. *For any $v \in V$, $\deg(v) > n/6 - 9\varepsilon' \cdot n$.*

Proof. If there exist a vertex $w \in V$ of a smaller degree, then by removing w and adding a clone of the vertex v from the above claim, we improve the graph contradicting Claim 7.2. □

Claim 7.7. *Every $u, w \in D$ have a common neighbor in D.*

Proof. First observe that all pairs of vertices $u \in D'$ and $w \in D$ have a common neighbor in D'. Suppose for a contradiction there exist two vertices $u, w \in D \setminus D'$ with no common neighbor in D. Then consider the graph G' obtained from G by removing both u and w, adding a new vertex u' connected to the whole set $D' \cap D$, and adding a new vertex w' which will be a clone of the vertex v from Claim 7.5.

We removed at most $|D| + 2\varepsilon' n$ edges from G, and added $\deg(u') + \deg(w') \geq |D| + n/6 - 10\varepsilon' n$ new edges. Moreover, all the removed edges were in $\mathcal{C}_{2k+1}(G)$, so G' contradicts Claim 7.2. □

Now let us concentrate on the vertex-cut C. Firstly, we observe that C must be non-empty.

Claim 7.8. *G is a connected graph. In particular $|C| \geq 1$.*

Proof. If G is disconnected, take any two connected components of G and add one edge between them. Clearly, the added edge does not occur in any cycle contradicting $G \in \mathcal{G}_n^{2k+1}$. □

In the following series of claims, we will show that $|C| \leq 1$. In order to do so, we split the vertices of C based on their adjacencies to A and B (recall no vertex can be adjacent to both $u \in A$ and $w \in B$). Let $C_A := \{v \in C \mid \deg_A(v) > 0\}$ and $C_B := C \setminus C_A = \{v \in C \mid \deg_B(v) > 0\}$.

Claim 7.9. *$|E(C_A)| = |E(C_B)| = 0$.*

Proof. Suppose the claim is false. Without loss of generality, there is an edge $\{v_1, v_2\} \in E(C_A)$. Consider any two vertices $u_1 \in N_A(v_1)$ and $u_2 \in N_A(v_2)$, any vertex $w_1 \in N_B(u_1)$, any vertex $u_3 \in N_A(w_1) \setminus \{u_1, u_2\}$, and a $(2k-3)$-vertex path P between the vertices u_2 and u_3 with the internal vertices disjoint from u_1, v_1, v_2 and w_1. It follows that $v_1 u_1 w_1 P v_2$ yields a copy of C_{2k+1} in G. Therefore,

$$|E(A,B) \cap \mathcal{C}_{2k+1}(G)| \geq \deg_B(u_1) \cdot (\deg_A(w_1) - 2) > n^2/50,$$

and hence $|\mathcal{C}_{2k+1}(G)| > 2n^2/9 + (n+22)/18$; a contradiction. □

Next, we study the edges between the sets C and D.

Claim 7.10. *For any set $X \subseteq C$, $|N_D(X)| > |X|$. In particular, every vertex $v \in C$ have at least two neighbors in D.*

Proof. Suppose for contradiction that there exists $X \subseteq C$ with $|N_D(X)| \leq |X|$, and let $Y := N_D(X)$. By Corollary 7.6, $\deg_{A \cup B}(v) > n/6 - 9\varepsilon' n > n/7$ for any $v \in X$. Therefore, $(C \cup Y) \setminus X$ is a vertex-cut of size at most $|C|$ and $G[A \cup B \cup X]$ is a bipartite graph (from Claim 7.9) with a minimum degree at least $n/7$ contradicting the choice of C. □

Since every $v \in C$ has at least two neighbors in D, we conclude that every edge between C and D occurs in some $(2k+1)$-cycle, i.e., $E(C, D) \subseteq \mathcal{C}_{2k+1}(G)$.

Claim 7.11. *$|N_D(u_a) \cap N_D(u_b)| = 0$ for any $u_a \in C_A$ and $u_b \in C_B$.*

Proof. Suppose there exists $w \in N_D(u_a) \cap N_D(u_b)$. Let $v_a \in N_A(u_a)$ and $v_b \in N_B(u_b)$ be chosen arbitrarily, and consider the bipartite subgraph (A', B') with $A' := N_A(v_b)$ and $B' := N_B(v_a)$. It follows that $|E(A,B) \setminus E(A',B')| < 4\varepsilon' n^2$. On the other hand, any edge $\{x,y\} \in E(A',B')$ occurs in C_{2k+1} for all $k \geq 3$, a contradiction. □

Claim 7.12. $|C_A| \leq 1$ *and* $|C_B| \leq 1$.

Proof. By symmetry, it is enough to prove that $|C_A| \leq 1$. Suppose for contradiction that $|C_A| \geq 2$.

We first consider the case when there are two vertices $u_1, u_2 \in C_A$ and an edge $\{x_1, x_2\} \in E(D)$ with $x_1 \in N_D(u_1)$ and $x_2 \in N_D(u_2)$. In other words, there is a 4-vertex path with both of its endpoints in C_A. Let $W := N_A(\{u_1, u_2\})$. Note that $|W| \geq 2$ as otherwise $(C \cup W) \setminus \{u_1, u_2\}$ contradicts the minimality of C. Since any two vertices $w_1, w_2 \in A$ have more than $2n/7 - |B| > 4|B|/7$ common neighbors in B, we conclude that $|E(W,B) \setminus \mathcal{C}_{2k+1}(G)| < 3|B|/7 < n/13$. Also, $\deg(w) \leq |B| + |C| < n/5$ for any $w \in A$. It follows that the graph obtained from G by removing the vertex-set W, adding $|W|-1$ new vertices fully connected to D, and adding a clone of a vertex v from Claim 7.5 yields a graph G' with more than $n^2/4$ edges and $|E(G') \setminus \mathcal{C}_{2k+1}(G')| > |E(G) \setminus \mathcal{C}_{2k+1}(G)|$, a contradiction with Claim 7.2.

For the rest of the proof, we may assume there is no 4-vertex path with the endpoints in C_A. In particular, at most one vertex from C_A can have $\Omega(\varepsilon n)$ neighbors in D. Let us now focus on the edges between C_A and A that are not in $\mathcal{C}_{2k+1}(G)$. Clearly, there are at most $|A|$ of them since any two edges $e_1, e_2 \in E(C_A, A)$ with $e_1 \cap e_2 \in A$ occur in C_{2k+1}. Now suppose there exist two vertices $u_1, u_2 \in C_A$ that both have less than $n/24$ neighbors in D. By Corollary 7.6, it follows that $|N_A(u_1) \cap N_A(u_2)| > |A|/3$. On the other hand, $\deg(u_1) + \deg(u_2) < n/2$. Therefore, replacing the vertices u_1 and u_2 with one new vertex adjacent to every vertex in D and a clone of the vertex v from Claim 7.5 again yields a contradiction with Claim 7.2.

We conclude that if $|C_A| \geq 2$, then $C_A = \{u_1, u_2\}$ and $\deg_D(u_1) \geq n/24$. Note that u_1 or u_2 is incident to at most $|A|/2$ edges that do not occur in C_{2k+1}. Let $u \in C_A$ be this vertex and let $u' \in C_A$ be the other vertex. Since $\deg_D(u_2) \geq 2$ and hence $\deg_D(u') \geq 2$, there are at least $2 \cdot (\deg_D(u_1) - 2) > \deg_D(u)$ non-edges in D between $N_D(u)$ and $N_D(u')$. Therefore, removing the vertex u, adding all the edges $\{w, w'\}$ with $w \in N_D(u)$ and $w' \in N_D(u')$, and adding a clone of the vertex v from Claim 7.5 contradicts Claim 7.2, which finishes the proof of the claim. □

It remains to show that we cannot have both $|C_A| = 1$ and $|C_B| = 1$.

Claim 7.13. $|C| = 1$.

Proof. Suppose for contradiction there are vertices $u_a \in C_A$ and $u_b \in C_B$. Firstly, recall that $N_D(u_a) \cap N_D(u_b) = \emptyset$ by Claim 7.11.

Now let us prove that both $|N_A(u_a)|$ and $|N_B(u_b)|$ must have quite small sizes, say less than $n/24$. Suppose, without loss of generality, that $|N_A(u_a)| \geq n/24$. Our aim now is to show that any edge incident to u_b is contained in some C_{2k+1}. Consider any vertex $v \in N_B(u_b)$. Since $\deg_A(v) > |A| - 11\varepsilon' n$, the vertices v and u_a have a common neighbor $w \in A$. Therefore, $u_a w v u_b P$ gives a copy of C_{2k+1}, where P is a $(2k-3)$-vertex path in D between $x \in N_D(u_a)$ and $x' \in N_D(u_b) \setminus \{x\}$. We conclude that every edge incident to u_b is in $\mathcal{C}_{2k+1}(G)$. Moreover, there is at least one edge in $E(A,B) \cap \mathcal{C}_{2k+1}(G)$. But then consider a graph G' obtained from G by removing at most $|B|$ edges between u_b and B, and adding at least $|D| > |B|$ missing edges between C and D. Since no edge from $E(A,B)$ is in $\mathcal{C}_{2k+1}(G')$, the graph G' contradicts Claim 7.2.

It remains to consider the case when both $|N_A(u_a)|$ and $|N_B(u_b)|$ have sizes less than $n/24$. But then removing all the edges from, say, u_a to A, and adding all the missing edges between u_b and B yield a graph G' that again contradicts Claim 7.2. □

This gives us a complete information on the structure of the extremal graphs.

Corollary 7.14. $E(G) \subseteq E(A,B) \cup E(B,C) \cup E(C,D) \cup \binom{D}{2}$.

It follows that all the edges that do not occur in C_{2k+1} are incident to vertices in B.

Corollary 7.15. $E(A,B) = |A||B|$ *and* $|B| \geq |A|$*. Moreover,* $F_{2k+1}(n) = \left\lfloor \frac{n^2}{4} \right\rfloor + 1 - (|A|+1)|B|$.

Finally, knowing the structure, it is straightforward to get the fact that $G \in \mathcal{G}_n^{2k+1}$ yields that $|A| = \lfloor (n-2)/6 \rfloor$, $|B| = \lfloor (n+1)/6 \rfloor$ and $|D| = \lfloor (2n+1)/3 \rfloor$. □

8 Concluding remarks

For an n-vertex graph G with $\lfloor n^2/4 \rfloor + 1$ edges, we determined the asymptotic minimum number of the edges of G that occur in some copy of C_5 in G, and for any $k \geq 3$, the exact minimum number of the edges that occur in C_{2k+1}. Our results show that the pentagon case has a very different behavior compared to all the longer odd cycles. These results confirm a conjecture of Füredi and Maleki, who proved the optimal asymptotic bounds under a stronger assumption that G has $(1/4 + \varepsilon)n^2$ edges.

Our main tool was an application of techniques from finite forcibility in the setting of flag algebras, combined with stability results on triangle-free graphs. This was crucial for dealing with n-vertex graphs that have only $\lfloor n^2/4 \rfloor + 1$ edges. We believe that our approach can be adapted to various other scenarios, and we intend to investigate this direction further.

We were also able to guide flag algebras to give us additional structural information for extremal configurations which yielded the corresponding stability results. These stability results allowed us to fully describe the structure of all the sufficiently large tight constructions.

If G contains αn^2 edges for some $\alpha > 1/4$, then a standard averaging argument yields that G must contain much more edges that occur in C_{2k+1}, for k being fixed, than Theorems 1.3 and 1.4 guarantee for $\lfloor n^2/4 \rfloor + 1$ edges. However, the averaging argument yields only a weak improvement. Füredi and Maleki [18] determined an asymptotically optimal lower bound for this problem. Note that the corresponding approximate result for triangles was proven by Füredi and Maleki in [17].

Füredi and Maleki [18] also considered a more general question, where instead of minimizing the number of edges that occur in odd cycles of a fixed length, one minimizes the number of edges that occur in copies of F for some fixed graph F. If the graph F has chromatic number $\chi = 3$, they obtained an asymptotically tight solution to this question. However, for graphs F with chromatic number $\chi \geq 4$, these questions are widely open.

Acknowledgments.

The authors thank Zoltan Füredi and Zeinab Maleki for discussing the results of [18] and the relation to the results obtained in this paper, and to Shoham Letzter for her suggestions regarding the results presented in Sections 6 and 7. We also thank the anonymous referees for their valuable comments, and Jake Cooper and Dan Král' for fruitful discussions at the beginning of this project. All of these greatly improved the presentation of our results.

References

[1] T. Austin and T. Tao. Testability and repair of hereditary hypergraph properties. *Random Structures Algorithms*, 36(4):373–463, 2010.

[2] R. Baber. Turán densities of hypercubes. ArXiv e-prints, 1201.3587, 2012.

[3] R. Baber and J. Talbot. Hypergraphs do jump, 2011.

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

[5] 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 J. Combin.*, 35:75–85, 2014.

[6] J. Balogh, P. Hu, B. Lidický, O. Pikhurko, B. Udvari, and J. Volec. Minimum number of monotone subsequences of length 4 in permutations. *Combin. Probab. Comput.*, 24(4):658–679, 2015.

[7] B. Bollobas. *Extremal Graph Theory.* Dover Publications, Inc., New York, NY, USA, 2004.

[8] B. Borchers. CSDP, A C library for semidefinite programming. *Optimization Methods and Software*, 11(1-4):613–623, 1999.

[9] 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.

[10] 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.

[11] P. Erdős. On extremal problems of graphs and generalized graphs. *Isr. J. Math.*, 2(3):183–190, 1964.

[12] P. Erdős. Some recent problems and results in graph theory. *Discrete Math.*, 164(1–3):81–85, 1997.

[13] P. Erdős, R. Faudree, and C. Rousseau. Extremal problems involving vertices and edges on odd cycles. *Discrete Math.*, 101(1):23–31, 1992.

[14] V. Falgas-Ravry, E. Marchant, O. Pikhurko, and E. R. Vaughan. The codegree threshold for 3-graphs with independent neighborhoods. *SIAM J. Discrete Math.*, 29(3):1504–1539, 2015.

[15] V. Falgas-Ravry and E. R. Vaughan. Turán H-densities for 3-graphs. *Electron. J. Combin.*, 19:#P40, 2012.

[16] V. Falgas-Ravry and E. R. Vaughan. Applications of the semi-definite method to the Turán density problem for 3-graphs. *Combin. Probab. Comput.*, 22(1):21–54, 2013.

[17] Z. Füredi and Z. Maleki. The minimum number of triangular edges and a symmetrization method for multiple graphs. *Combin. Probab. Comput.*, 26(4):525–535, 2017.

[18] Z. Füredi and Z. Maleki. A proof and a counterexample for a conjecture of Erdős concerning the minimum number of edges on odd cycles. In preparation.

[19] R. Glebov, D. Kráľ, and J. Volec. A problem of Erdős and Sós on 3-graphs. *Isr. J. Math.*, 211:349–366, 2016.

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

[21] A. Grzesik. On the Caccetta-Häggkvist Conjecture with a Forbidden Transitive Tournament. *Electron. J. Combin.*, 24(2):#P2.19, 2017.

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

[23] H. Hatami, J. Hladký, D. Kráľ, S. Norine, and A. Razborov. Non-three-colourable common graphs exist. *Combin. Probab. Comput.*, 21(5):734–742, 2012.

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

[25] 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.

[26] D. Kráľ, C.-H. Liu, J.-S. Sereni, P. Whalen, and Z. B. Yilma. A new bound for the 2/3 conjecture. *Combin. Probab. Comput.*, 22(3):384–393, 2013.

[27] 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.

[28] W. Mantel. Problem 28. *Wiskundige Opgaven*, 10:60–61, 1907.

[29] V. Nikiforov. The number of cliques in graphs of given order and size. *Trans. Amer. Math. Soc.*, 363(3):1599–1618, 2011.

[30] O. Pikhurko. The minimum size of 3-graphs without a 4-set spanning no or exactly three edges. *European J. Combin.*, 32(7):1142–1155, 2011.

[31] O. Pikhurko and A. A. Razborov. Asymptotic structure of graphs with the minimum number of triangles. *Combin. Probab. Comput.*, 26(1):138–160, 2017.

[32] O. Pikhurko and E. R. Vaughan. Minimum number of k-cliques in graphs with bounded independence number. *Combin. Probab. Comput.*, 22(6):910–934, 2013.

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

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

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

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

[37] C. Reiher. The clique density theorem. *Ann. of Math.*, 184(3):683–707, 2016.

[38] I. Z. Ruzsa and E. Szemerédi. Triple systems with no six points carrying three triangles. *Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai*, 18:939–945, 1978.

[39] K. Sperfeld. The inducibility of small oriented graphs. ArXiv e-prints, 1111.4813, 2011.

[40] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 7.4)*, 2016. `http://www.sagemath.org`.

A Formal verification of correctness of Propositions 3.2 and 4.2

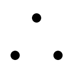
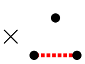
In order to verify the correctness of the claimed identities, given a *proof-certificate* consisting of matrices $\widehat{L}$, M_λ, $\widehat{B}$, M_β, $\widehat{R}$, M_ρ and two numbers $a, b > 0$, we perform the following 8 steps:

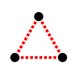
1) Generate all the non-isomorphic flags in the sets $\mathcal{H}_6$, $\mathcal{H}_4^\lambda$, $\mathcal{H}_4^\beta$, and $\mathcal{H}_4^\rho$,

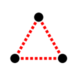
2) For every $F_1^\sigma, F_2^\sigma \in \mathcal{H}_4^\sigma$, where $\sigma \in \{\lambda, \beta, \rho\}$, express $[\![F_1^\sigma \times F_2^\sigma]\!]_\sigma$ as $\sum_{H \in \mathcal{H}_6} p_H^{F_1^\sigma, F_2^\sigma} \cdot H$,

3) Verify that the three matrices $\widehat{L}$, $\widehat{B}$, and $\widehat{R}$ are positive definite,

4) Express $\left[\!\!\left[v_\lambda^{\mathrm{T}} M_\lambda^{\mathrm{T}} \cdot \widehat{L} \cdot M_\lambda v_\lambda\right]\!\!\right]_\lambda + \left[\!\!\left[v_\beta^{\mathrm{T}} M_\beta^{\mathrm{T}} \cdot \widehat{B} \cdot M_\beta v_\beta\right]\!\!\right]_\beta + \left[\!\!\left[v_\rho^{\mathrm{T}} M_\rho^{\mathrm{T}} \cdot \widehat{R} \cdot M_\rho v_\rho\right]\!\!\right]_\rho$ as $\sum_{H \in \mathcal{H}_6} \zeta_H \cdot H$,

5) Express ( +  − ) ×  as $\sum_{H \in \mathcal{H}_6} \gamma_H \cdot H$ and ( +  − ) ×  as $\sum_{H \in \mathcal{H}_6} \gamma'_H \cdot H$,

6a) In the case of Proposition 3.2, express  × (8 ·  − $(2 + \sqrt{2})$ · •) as $\sum_{H \in \mathcal{H}_6} \kappa_H \cdot H$,

6b) In the case of Proposition 4.2, express  × (9 ·  − 4 · •) as $\sum_{H \in \mathcal{H}_6} \kappa_H \cdot H$,

7) For every $H \in \mathcal{H}_6$, verify that

$$\kappa_H \geq \zeta_H + a \cdot \gamma_H + b \cdot \gamma'_H \,, \tag{5}$$

8a) In the case of Proposition 3.2, verify that the inequality (5) is strict for every $H \in \mathcal{P}_5 \cup \{C_4^X\}$, and

8b) In the case of Proposition 4.2, verify that the inequality (5) is strict for every $H \in \mathcal{P}_4$.

In our verification scripts on the webpage `http://honza.ucw.cz/proj/EdgesInCycles/`, we implement the first and the second step by a simple exhaustive search over all the possibilities. The positive-definiteness of the given matrices is verified by finding their LDL^{T} decompositions and testing whether all diagonal entries of D are positive. Note that for the matrix decomposition, we use the corresponding function in SAGE which uses exact arithmetics.

Next, given a proof-certificate $\left(\widehat{L}, M_\lambda, \widehat{B}, M_\beta, \widehat{R}, M_\rho, a, b\right)$ and the values of $p_H^{F_1^\sigma, F_2^\sigma}$ computed in the second step, we directly compute the values of ζ_H, again using exact arithmetics implemented in SAGE. Note that to do so, we only need to perform summation and multiplication in $\mathbb{Q}\left[\sqrt{2}\right]$.

To find the values ζ_H, γ_H, γ'_H and κ_H from the steps 4-6, we again go exhaustively through all the possibilities. Finally, the steps 7 and 8 are verified by a direct computation in $\mathbb{Q}\left[\sqrt{2}\right]$.