Inducibility of directed paths

Ilkyoo Choi[*]   Bernard Lidický[†]   Florian Pfender[‡]

November 8, 2018

Abstract

A long standing open problem in extremal graph theory is to describe all graphs that maximize the number of induced copies of a path on four vertices. The character of the problem changes in the setting of oriented graphs, and becomes more tractable. Here we resolve this problem in the setting of oriented graphs without transitive triangles.

An oriented graph is a directed graph without 2-cycles. In this paper, both undirected graphs and oriented graphs are considered, and the following definitions apply to both classes. For a graph G, we use $|G|$ to denote the number of vertices of G. We use P_n to denote the path on n vertices. Given graphs G and H, the *density* of H in G, denoted $d_H(G)$, is defined to be

$$d_H(G) = \frac{\text{\# of induced copies of } H \text{ in } G}{\binom{|V(G)|}{|V(H)|}}.$$

Given a fixed graph H and a family $\mathcal{G}$ of graphs, investigating the maximum or minimum value of $d_H(G)$ over all graphs $G \in \mathcal{G}$ is an important area of research in extremal graph theory. This question was formulated by Pippenger and Golumbic [31], where they define the *(maximal) inducibility* of a given graph H, denoted $I(H)$, as

$$I(H) = \lim_{n \to \infty} \max_{|G|=n} d_H(G).$$

They initiated the study by considering the family of undirected graphs, and they proved that for a graph H, the value $\max_{|G|=n} d_H(G)$ is nondecreasing and the limit $I(H)$ always exists. A natural line of research is to to refine the question by considering an (infinite) family

1

$\mathcal{G}$ of graphs (instead of the family of all graphs), and define the *(maximal) inducibility of H in $\mathcal{G}$* as

$$I(H, \mathcal{G}) = \lim_{n \to \infty} \max_{|G|=n, G \in \mathcal{G}} d_H(G),$$

if the limit exists.

Given a graph H, natural candidate graphs for maximizing the number of induced copies of H are the iterated balanced blow-ups of H: Partition n vertices into $|V(H)|$ classes of sizes $\lceil \frac{n}{|V(H)|} \rceil$ and $\lfloor \frac{n}{|V(H)|} \rfloor$, corresponding to the vertices of H. Add all possible edges between any two classes corresponding to an edge of H. Now iterate this process inside each class. When H has k vertices, a simple calculation shows that a sequence of iterated balanced blow-ups of H gives $I(H) \geq k!/(k^k - k)$. In the original paper by Pippenger and Golumbic [31], they conjectured that for cycles of length at least 5, this bound is tight. This is still an open question for almost all values of k. A graph H is called a *fractalizer* if the iterated balanced blow-ups of H are the only graphs maximizing the number of induced copies of H for every n. In particular, for each fractalizer the bound is tight. Interestingly, Fox, Huang, and Lee [14, 15] showed that almost all graphs are fractalizers by considering random graphs. A similar result was recently published by Yuster [34].

Recently, investigating the inducibility of small graphs received much attention, thanks to the flag algebra method invented by Razborov [32]. With the notable exception of P_4, the inducibility of graphs on at most four vertices is well understood, see Even-Zoha and Linial [12]. For P_4, the known best lower bound on $I(P_4)$ is $1173/5824 \approx 0.2014$, provided by a construction from [12], and the best upper bound 0.204513, obtained by Vaughan [33] using flag algebras.

Inducibility of 5-vertex graphs is also not completely resolved. Recently, by proving $I(C_5) = \frac{1}{26}$, Balogh et al. [5] determined that the bound is tight for C_5. Before this result, Hatami et al. [20] and independently Grzesik [19] solved the Erdős pentagon problem, which asks for the value of $I(C_5, \mathcal{T})$, where $\mathcal{T}$ is the family of triangle-free graphs. In [30], this last problem is resolved for graphs of all orders. The main difference between the problems of determining $I(C_5, \mathcal{T})$ and $I(C_5)$ is the extremal construction. A balanced blow-up of C_5 is the extremal construction when considering triangle-free graphs, and an iterated balanced blow-up of C_5 is the extremal construction when there are no restrictions on the graphs under consideration. When determining $I(C_5, \mathcal{T})$, the flag algebra method gives the exact upper bound on $I(C_5, \mathcal{T})$. On the other hand, proving a tight upper bound on $I(C_5)$ by merely using flag algebras appears out of reach, and stability methods are used to improve the bound from flag algebra.

In this paper, we consider inducibility of oriented graphs. Hladký, Kráľ, and Norin [23] announced that $I(\vec{P}_3) = \frac{2}{5}$ and the extremal construction is an iterated blow-up of $\vec{C}_4$. We conjecture that this generalizes to longer oriented paths, namely, the number of induced copies of $\vec{P}_k$ is maximized by an iterated blow-up of $\vec{C}_{k+1}$.

Conjecture 1. *The number of induced copies of $\vec{P}_k$ over all oriented graphs on n vertices*

is maximized by an iterated balanced blow-up of $\vec{C}_{k+1}$. As a consequence,

$$I(\vec{P}_k) = \frac{k!}{(k+1)^{k-1} - 1}.$$

Note that Conjecture 1 states that the graph maximizing the number of induced copies of $\vec{P}_k$ is the same graph as the graph conjectured to maximize the number of induced copies of $\vec{C}_{k+1}$. The statement regarding $I(\vec{C}_5)$ is a consequence of a result by Balogh et al. [5] on $I(C_5)$. Note that Hu et al. [24] resolved $I(\vec{C}_4)$, where the extremal example is an iterated blow-up of $\vec{C}_4$. This last construction is not extremal in the undirected case.

Let $\vec{T}_3$ denote the transitive tournament on three vertices. Similar to triangle-free graphs in the class of undirected graphs, $\vec{T}_3$-free oriented graphs do not include iterated blow-ups of small graphs. Therefore, extremal graphs often have simpler structure. In this vain, we attack Conjecture 1 first by considering the same inducibility parameter but for $\vec{T}_3$-free oriented graphs. We formulate the following conjecture.

Conjecture 2. *The number of induced copies of $\vec{P}_k$ over all $\vec{T}_3$-free oriented graphs on n vertices is maximized by a balanced blow-up of $\vec{C}_{k+1}$. As a consequence,*

$$I(\vec{P}_k, \vec{\mathcal{T}}) = \frac{k!}{(k+1)^{k-1}},$$

where $\vec{\mathcal{T}}$ is the family of $\vec{T}_3$-free oriented graphs.
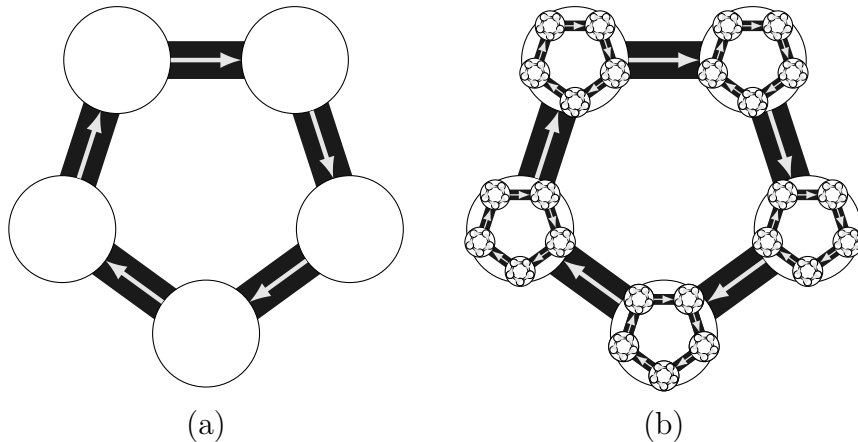

Figure 1: (a) The blow-up of $\vec{C}_5$ and (b) the iterated blow-up of $\vec{C}_5$.

In this paper, we prove Conjecture 2 for $\vec{P}_4$, and we also show the uniqueness of the extremal construction for sufficiently large graphs.

Theorem 3. *Let $\vec{\mathcal{T}}$ be the family of oriented graphs without $\vec{T}_3$.*

$$I(\vec{P}_4, \vec{\mathcal{T}}) = \frac{24}{125}.$$

Moreover, for n that is either sufficiently large or divisible by 5, the balanced blow-up of $\vec{C}_5$ is the only oriented n-vertex graph that maximizes the number of induced copies of $\vec{P}_4$ in $\vec{\mathcal{T}}$.

Our proof uses the flag algebra method. The method was developed by Razborov [32] and it has been successfully applied in various settings, see [2–4,6,7,9–11,13,16–18,22,25–27,29]. The method has been already described in many previous papers, so we do not describe it here and merely use it as a black box. For an accessible introduction to flag algebras, see [8].

In Section 1, we first show the stability of the extremal construction, and then we obtain the exact result in Section 2. Utilizing a tool developed by Pippenger and Golumbic [31] and Král', Norin, and Volec [28] in order to study the value of $I(C_k)$, we discuss upper bounds on $I(\vec{P}_k)$ and $I(\vec{P}_k, \vec{\mathcal{T}})$, where $\vec{\mathcal{T}}$ are $\vec{T}_3$-free oriented graphs, for all k, in Section 3.

From now on, we will use $\vec{C}_5(n)$ to denote the balanced blow-up of $\vec{C}_5$ on n vertices, see Figure 1(a).

1 Stability

This section is devoted to proving the following stability lemma.

Lemma 4. *For every $\varepsilon > 0$, there exist n_0 and $\varepsilon' > 0$ such that every oriented $\vec{T}_3$-free graph G of order $n \geq n_0$ with $d_{\vec{P}_4}(G) \geq \frac{24}{125} - \varepsilon'$ is isomorphic to $\vec{C}_5(n)$ after adding and/or removing at most εn^2 edges.*

Our main tools to prove Lemma 4 are flag algebras and a removal lemma. We use the following removal lemma, which follows from a more general theorem by Aroskar and Cummings [1].

Lemma 5 (Infinite Induced Oriented Graph Removal Lemma [1]). *Let $\mathcal{F}$ be a (possibly infinite) set of oriented graphs. For every $\epsilon_R > 0$, there exist n_0 and $\delta > 0$ such that for every oriented graph G of order $n \geq n_0$, if G contains at most $\delta n^{v(H)}$ induced copies of H for each H in $\mathcal{F}$, then there exists G' of order n such that G' is induced H-free for all H in $\mathcal{F}$ and G' can be obtained from G by adding/removing/reorienting at most $\epsilon_R n^2$ edges.*

Let $\mathcal{F}$ be the oriented 4-vertex graphs depicted in Figure 2; we call them the *forbidden oriented graphs*. A standard flag algebra calculation shows that the forbidden oriented graphs rarely appear in extremal examples.

Lemma 6. *For every $\delta > 0$, there exist n_0 and $\varepsilon' > 0$ such that every oriented $\vec{T}_3$-free graph G of order $n \geq n_0$ with $d_{\vec{P}_4}(G) \geq \frac{24}{125} - \varepsilon'$ contains at most δn^4 induced copies of an oriented graph in $\mathcal{F}$. Furthermore, G contains at most δn^3 directed triangles.*
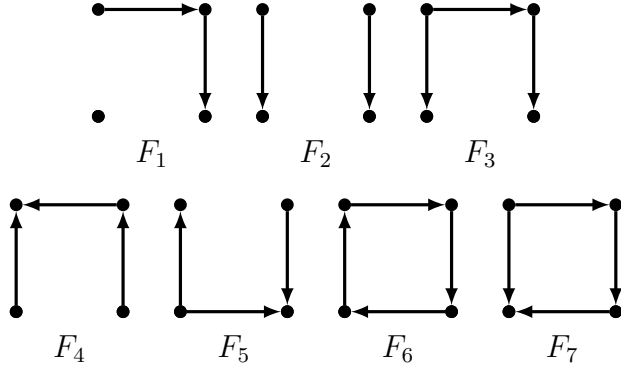
4

Figure 2: Forbidden induced subgraphs.

Proof. We perform a calculation using the plain flag algebra framework. We obtain that if $(G_k)_{k\in\mathbb{N}}$ is a convergent sequence of oriented graphs, then $\lim\limits_{k\to\infty} d_{\vec{P}_4}(G_k) \le \frac{24}{125}$. Moreover, if $\lim\limits_{k\to\infty} d_{\vec{P}_4}(G_k) = \frac{24}{125}$, then for every $F \in \mathcal{F}$, it follows that $\lim\limits_{k\to\infty} d_F(G_k) = 0$. It was sufficient to execute the calculation with flags on 4 vertices and two types. Rounding was performed as described in [7]. All technical details of the calculation, including rounded solution matrices, are available at `http://lidicky.name/pub/P4noT3` and on the arXiv. □

Proof of Lemma 4. Fix $\varepsilon > 0$. We use positive numbers δ and ε_R that depend on Lemmas 5 and 6. We specify the dependency later in the proof together with n_0 and ε'.

Let G be an oriented graph of order $n \ge n_0$ with $d_{\vec{P}_4}(G) \ge \frac{24}{125} - \varepsilon'$. Notice that

$$d_{\vec{P}_4}(G) \ge \frac{24}{125} - \varepsilon' = d_{\vec{P}_4}(\vec{C}_5(n)) - \varepsilon' + o(1). \tag{1}$$

By Lemma 6, G contains at most δn^4 induced copies of oriented graphs in $\mathcal{F}$, and at most δn^3 triangles. By Lemma 5, there exists an oriented $\vec{T}_3$-free graph G' (on the same vertex set as G) differing from G in at most $\varepsilon_R n^2$ pairs that avoids all oriented graphs in $\mathcal{F}$ and all triangles.

Claim 7. *G' is a (not necessarily balanced) blow-up of $\vec{C}_5$.*

Proof. Let $P = v_1, v_2, v_3, v_4$ form an induced $\vec{P}_4$ in G'. We call a vertex x a *clone* of v_i if x and v_i have exactly the same in-neighbors and out-neighbors on P. Let $X_1, \ldots, X_5$ be pairwise disjoint subsets of $V(G')$, where

$$X_i = \{x : x \text{ is a clone of } v_i \text{ in } P\} \qquad \text{for } i \in \{1, 2, 3, 4\},$$
$$X_5 = \{x : N^+(x) \cap V(P) = \{v_1\}, N^-(x) \cap V(P) = \{v_4\}\}.$$

Using the list $\mathcal{F}$ of forbidden oriented graphs, we show that $X_1, \ldots, X_5$ is a partition of $V(G')$. Let $y \in V(G) \setminus \{v_1, v_2, v_3, v_4\}$. By F_1, y has at least one neighbor on P. By $\vec{T}_3$ and

5

$\vec{C}_3$, y cannot have two consecutive neighbors on P. In particular, y cannot have three or four neighbors on P.

Assume y has exactly two neighbors s and t on P. If $\{s,t\} = \{v_1, v_3\}$, then $y \in X_2$ by F_6 and F_7. If $\{s,t\} = \{v_2, v_4\}$, then $y \in X_3$ by F_6 and F_7. If $\{s,t\} = \{v_1, v_4\}$, then $y \in X_5$ by F_3 and F_4.

Assume y has exactly one neighbor z on P. By F_3 and F_4, we know $z \notin \{v_1, v_4\}$. If $z = v_2$, then $y \in X_1$ by F_3, and if $z = v_3$, then $y \in X_4$ by F_4. $\qquad\square$

Next, we show G' is close to being a balanced blow-up of $\vec{C}_5$.

Claim 8. *For every $\varepsilon_B > 0$, there exist $\varepsilon_R > 0$ and $\varepsilon' > 0$ such that if n_0 is sufficiently large and G and G' differ in at most $\varepsilon_R n^2$ pairs and $d_{\vec{P}_4}(G) \geq \frac{24}{125} - \varepsilon'$, then $\vec{C}_5(n)$ and G' differ in at most $\varepsilon_B n^2$ pairs.*

Proof. Given an oriented graph H, let $\vec{P}_4(H)$ denote the number of induced copies of $\vec{P}_4$ in H. Since G' was obtained from G by changing at most $\varepsilon_R n^2$ pairs, $\vec{P}_4(G')$ is large:

$$\vec{P}_4(G') \geq \vec{P}_4(G) - \varepsilon_R n^4. \tag{2}$$

Notice that $d_{\vec{P}_4}(G) \geq \frac{24}{125} - \varepsilon'$ implies that for sufficiently large n,

$$\vec{P}_4(G) \geq \left(\frac{24}{125} - \varepsilon'\right)\binom{n}{4} \geq 5 \cdot \left(\frac{n}{5}\right)^4 - \varepsilon' n^4. \tag{3}$$

By evaluating $\vec{P}_4(G')$ and combining it with (2) and (3) we obtain:

$$\left(\prod_i |X_i|\right) \cdot \left(\sum_i \frac{1}{|X_i|}\right) \geq \vec{P}_4(G') \geq 5\left(\frac{n}{5}\right)^4 - \varepsilon' n^4 - \varepsilon_R n^4 = \left(\frac{1}{5^3} - \varepsilon' - \varepsilon_R\right)n^4.$$

The product on the left is maximized when $|X_i| = \frac{n}{5}$ for each $i \in \{1,2,3,4,5\}$, and the maximum value is $\frac{n^4}{5^3}$. Hence, for every $\varepsilon_B > 0$, there exist $\varepsilon_R > 0$ and $\varepsilon' > 0$ such that if $\varepsilon' + \varepsilon_R$ is small enough, then $\left(\frac{1-\varepsilon_B}{5}\right)n \leq |X_i| \leq \left(\frac{1+\varepsilon_B}{5}\right)n$. Therefore, in order to obtain $\vec{C}_5(n)$ from G', we need to move at most $\varepsilon_B n$ vertices between parts, which means changing at most $\varepsilon_B n^2$ pairs. $\qquad\square$

Let $\varepsilon_B = \varepsilon/2$. Let $\varepsilon_R \leq \varepsilon/2$ be small enough such ε_R, ε_B, ε', and n_0 satisfy Claim 8. Let $\delta > 0$ be small enough to satisfy Lemma 5 with ε_R. Finally, let ε' and n_0 be small and big, respectively, enough also for Lemma 6 when applied with δ. These choices will guarantee that G is different from $\vec{C}_5(n)$ in at most $(\varepsilon_B + \varepsilon_R)n^2 \leq \varepsilon n^2$ pairs. $\qquad\square$

2   Exact Result

This section contains the proof of Theorem 3. The proof follows the following outline. We start with an extremal example G with order n and use Lemma 4 to conclude that G is almost $\vec{C}_5(n)$. We first put "unruly vertices" aside and argue that the rest of G is exactly a (not necessarily balanced) blow-up of $\vec{C}_5$. We then argue that the "unruly vertices" have drastically different sets of neighbors compared to the rest of the vertices in G. Finally, we show that if there is a unruly vertex, then it would be in too few copies of an induced $\vec{P}_4$. Hence, there are no "unruly vertices", and we finish the proof by showing that G is a balanced blow-up of $\vec{C}_5$.

Given an oriented graph H and a set of vertices $A \subset V(H)$, let $\vec{P}_4(H, A)$ be the number of induced $\vec{P}_4$'s in H containing all vertices in A. If $A = \{a\}$, then we simplify the notation and write $\vec{P}_4(H, a)$ instead of $\vec{P}_4(H, \{a\})$.

Proof of Theorem 3. For simplicity, we fix $\varepsilon = 0.0000005$. Let n_0 be big enough to apply Lemma 4 with ε such that every extremal oriented graph H of order at least n_0 satisfies $d_{\vec{P}_4}(H) \geq \frac{24}{125} - \varepsilon$.

Let G be an extremal oriented graph of order $n \geq n_0$. By Lemma 4, the vertices of G can be partitioned into five parts $X_1, \ldots, X_5$ with sizes as equal as possible such that by changing the adjacencies of at most εn^2 pairs of vertices, the graph can be turned into $\vec{C}_5(n)$. Call a pair of vertices where the adjacency needs to be changed *funky*. Use f to denote the number of funky pairs in G, and we know that

$$f \leq \varepsilon n^2 = 0.0000005 n^2.$$

For a vertex $v \in V(G)$, let $f_p(v)$ denote the number of funky pairs containing v. Move every vertex with $f_p(v) \geq 0.001n$ to a new set X_0, so we know the following two inequalities hold:

$$|X_0| \leq 2f/0.001n = 0.001n$$
$$f_p(v) \leq 0.001n \text{ for all } v \in X_1 \cup \cdots \cup X_5.$$

Let x_{min} and x_{max} be a lower bound and an upper bound, respectively, on the size of X_i for all $i \in \{1, 2, 3, 4, 5\}$. Since we started with a balanced partition and $|X_0| \leq 0.001n$, we may use

$$x_{min} = 0.198n \leq |X_i| \leq \lceil 0.2n \rceil = x_{max} \quad \text{for } i \in \{1, 2, 3, 4, 5\}.$$

Claim 9. $G - X_0$ *is a blow-up of* $\vec{C}_5$.

Proof. Let uv be a funky pair in $G - X_0$. Obtain G_{uv} from G by making uv not funky. All induced $\vec{P}_4$'s either in G_{uv} but not in G or in G but not in G_{uv} must contain both u and v. Recall that $\vec{P}_4(H, \{u, v\})$ is the number of induced $\vec{P}_4$'s in H containing vertices u and v.

First, we get an easy lower bound

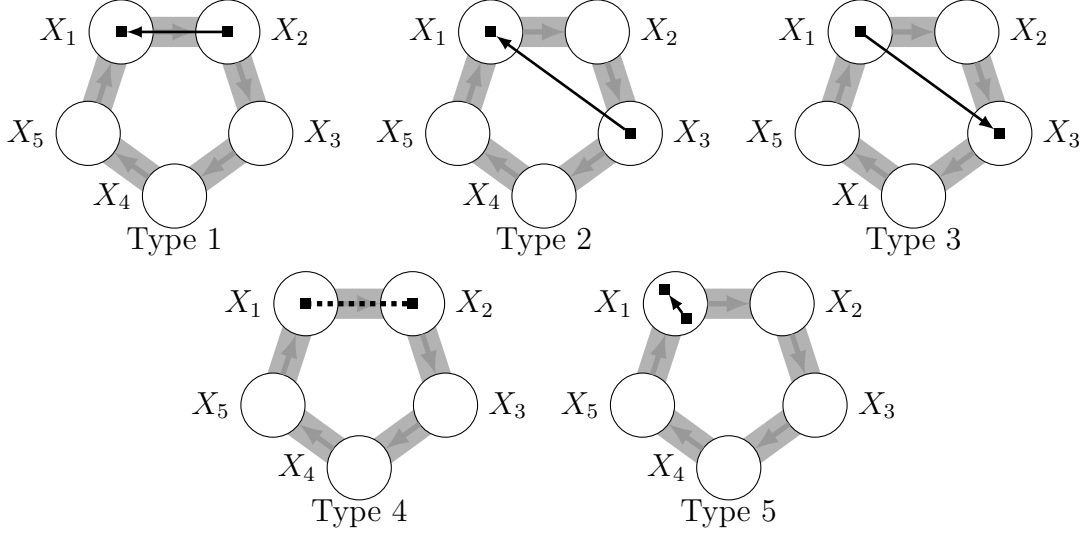$$\vec{P}_4(G_{uv}, \{u, v\}) \geq 3 \cdot x_{min}^2 - n \cdot f_p(u) - n \cdot f_p(v) - f.$$

Figure 3: Possible funky pairs.

For G, note that every induced $\vec{P}_4$ containing the funky pair uv must contain either a vertex from X_0 or another funky pair. Hence

$$\vec{P}_4(G, \{u, v\}) \leq n \cdot f_p(u) + n \cdot f_p(v) + n \cdot |X_0| + f.$$

By the extremality of G, we get

$$\begin{aligned}
\vec{P}_4(G, \{u, v\}) &\geq \vec{P}_4(G_{uv}, \{u, v\}) \\
0 &\geq 3 \cdot x_{min}^2 - 2n \cdot f_p(u) - 2n \cdot f_p(v) - n \cdot |X_0| - 2f \\
0 &\geq \left(3 \cdot 0.198^2 - 5 \cdot 0.001 - 0.000001\right) n^2 > 0.1n^2,
\end{aligned}$$

which is a contradiction. □

Claim 10. *For $i \in \{1, 2, 3, 4, 5\}$ and a vertex $x \in X_0$, if x is moved to X_i, then $f_p(x) \geq 0.17n$.*

Proof. Fix a vertex $x \in X_0$ and without loss of generality let $i = 1$. There are five possible types of funky pairs containing x, see Figure 3.

If x is in a funky pair of Type 3 and Type 5, then x is not adjacent to a vertex in X_2 and X_5, respectively, because $\vec{T}_3$ is forbidden. Hence x has funky degree at least $x_{min} = 0.198n$.

Therefore, we may assume that all funky pairs involving x are of Type 1, 2, or 4. Let xy be a funky pair of Type 1, 2, or 4. Obtain G_{xy} from G by making xy not funky. Recall that Claim 9 implies that all funky pairs in $G - (X_0 - x)$ contain x. This allows us to use slightly better estimates when counting the possible induced $\vec{P}_4$'s containing x and y. We get

$$\vec{P}_4(G, \{x, y\}) \leq x_{max} f_p(x) + \binom{f_p(x)}{2} + |X_0| \cdot n$$

8

and
$$\vec{P_4}(G_{xy}, \{x, y\}) \geq 3x_{min}^2 - 2x_{max}f_p(x).$$

By the extremality of G, we get

$$\vec{P_4}(G, \{x, y\}) \geq \vec{P_4}(G_{xy}, \{x, y\})$$

$$x_{max}f_p(x) + \binom{f_p(x)}{2} + |X_0| \cdot n \geq 3x_{min}^2 - 2x_{max}f_p(x)$$

$$\frac{f_p(x)^2}{2} + 3x_{max}f_p(x) - 3x_{min}^2 + |X_0| \cdot n \geq 0$$

This gives $f_p(x) \geq 0.17n$ and finishes the proof of Claim 10. $\qquad \square$

Claim 11. *For every vertex x, $\vec{P_4}(G, x) \geq 0.19\binom{n}{3}$.*

Proof. First we show that all vertices of G are in approximately the same number of induced $\vec{P_4}$'s. Suppose to the contrary that x and y are two vertices such that $\vec{P_4}(G, x) - \vec{P_4}(G, y) > n^2$. Obtain G' from G by deleting y and adding a clone of x, denoted by x'. Note that there is no edge between x and x' in G'. If x' was in a $\vec{T_3}$, then x would also be in a $\vec{T_3}$ since x and x' are not adjacent to each other. Hence G' is $\vec{T_3}$-free. The only induced $\vec{P_4}$'s that are different in G and G' are the ones containing both y and x'. Hence

$$\vec{P_4}(G') - \vec{P_4}(G) = \vec{P_4}(G, x) - \vec{P_4}(G, y) - \vec{P_4}(G, \{x, y\}) > 0,$$

since $\vec{P_4}(G, \{x, y\}) < n^2$. This contradicts that $\vec{P_4}(G)$ is maximum. Hence for two arbitrary vertices x and y, $|\vec{P_4}(G, x) - \vec{P_4}(G, y)| \leq n^2$. Since $\vec{P_4}(G) = 0.192\binom{n}{4} + o(n^4)$ and every $\vec{P_4}$ contains four vertices, the average number of $\vec{P_4}$'s containing one fixed vertex is $0.192\binom{n-1}{3} + o(n^3) = 0.192\binom{n}{3} + o(n^3)$. Therefore, $\vec{P_4}(G, x) \geq 0.19\binom{n}{3}$ for every vertex x when n is sufficiently large. $\qquad \square$

Claim 12. $|X_0| = 0$.

Proof. Suppose to the contrary that x is a vertex in X_0. We will show that x violates Claim 11.

For $j \in [5]$, let i_j, o_j, and n_j denote the number of in-neighbors, out-neighbors, and non-neighbors, respectively, of x in X_j divided by n. This allows us to count the number of induced $\vec{P_4}$'s containing x and no other vertex from X_0. To simplify the notation, for all $j > 5$ we define $i_j = i_{j-5}$, $o_j = o_{j-5}$, and $n_j = n_{j-5}$. The following program provides an upper bound on the number of induced $\vec{P_4}$'s containing x divided by n^3.

$$(P) \begin{cases} \text{maximize} & \sum_{j=1}^5 \left(o_j n_{j+1} n_{j+2} + i_j o_{j+2} n_{j+3} + n_j i_{j+1} o_{j+3} + n_j n_{j+1} i_{j+3} \right) \\ \text{subject to} & i_j + o_j + n_j \leq 0.21 \text{ for } j \in [5], \\ & o_j + i_j + n_{j+1} + i_{j+1} + i_{j+2} + o_{j+2} + i_{j+3} + o_{j+3} + n_{j+4} + o_{j+4} \geq 0.17 \text{ for } j \in [5], \\ & i_j, o_j, n_j \geq 0 \text{ for } j \in [5]. \end{cases}$$

The objective in (P) counts the number of induced $\vec{P_4}$'s containing x. The first set of constraints count relations between x and vertices in each X_j. Notice that we used a very generous upper bound on $|X_i|$. The second set of constraints comes from Claim 10, where we count the number of funky pairs containing x if x was in X_j.

We aim to provide an upper bound on the value of an optimal solution of (P). We do this by sampling points in the space of feasible solutions of (P) and then upper bounding the maximum by using first derivatives. Unfortunately, the program has ten variables, which seems to be too many for generating a sufficiently refined grid.

Fortunately, the presence of some edges incident with x blocks presence of other edges. If there are no edges from x to $\bigcup_{i \in [5]} X_i$, then we can reverse all edges of G. If there are still no edges from x to $\bigcup_{i \in [5]} X_i$, then all neighbors of x are in X_0 and x is in at most $|X_0| \binom{n}{2} < 0.001 n^3$ induced $\vec{P_4}$'s, which contradicts Claim 11.

By symmetry, assume there is an edge directed from x to a vertex in X_1. This already prevents all edges between x and X_5, and also edges from x to X_2, since G is $\vec{T_3}$-free.

All the possible combinations of allowed edges are depicted in Figure 4, possible to verify by case analysis. In each of them, there are only four variables. We examine them separately, run a mesh optimization program, and show in the following paragraphs that v is in at most $0.08 \binom{n}{3}$ induced $\vec{P_4}$'s, which contradicts Claim 11.

The mesh optimization program works in the following way. For each variable, it samples 100 points uniformly distributed in $[0, 0.21]$. That means examining 100^4 points. For each of the points, we test if it is a feasible solution to (P) and if yes, then we remember the solution with the highest value of the objective function of (P). The optimal solution of (P) must be in each coordinate at distance at most $0.21/100$ from some point we sampled[1]. The largest value among the sampled feasible points is less than 0.04.

The first partial derivative of the objective function in any variable is at most $6 \cdot 0.21^2 = 2.52$. Hence the difference between the point and the optimum is at most $4 \cdot 2.52 \cdot \frac{0.21}{100} < 0.03$, and the value of the optimum solution is at most 0.07.

This implies that x is in at most $0.07 \binom{n}{3}$ induced $\vec{P_4}$'s that avoid X_0. There are at most $|X_0| \binom{n}{2}$ other induced $\vec{P_4}$'s containing x. Hence there are at most $0.08 \binom{n}{3}$ induced $\vec{P_4}$'s containing x, which contradicts Claim 11.

$\square$

Claim 13. *G is a balanced blow-up. That is, $||X_i| - |X_j|| \leq 1$ for all $i, j \in [5]$.*

Proof. Suppose to the contrary that there exist i and j such that $|X_i| - |X_j| \geq 2$. Let $\{a, b, c\} = [5] \setminus \{i, j\}$.

We will obtain a contradiction by deleting a vertex in X_i and duplicating a vertex in X_j, and show that this will increase the number of induced $\vec{P_4}$'s. Let $y \in X_i$ and $z \in X_j$. Obtain G' from G by deleting y and adding a clone of z, denoted by z'. If z' was in a $\vec{T_3}$, then z

---

[1] We actually also sample points that slightly violate the constraints of (P) to make sure our grid captures the optimal solution if it is on the boundary of feasible solutions of (P).
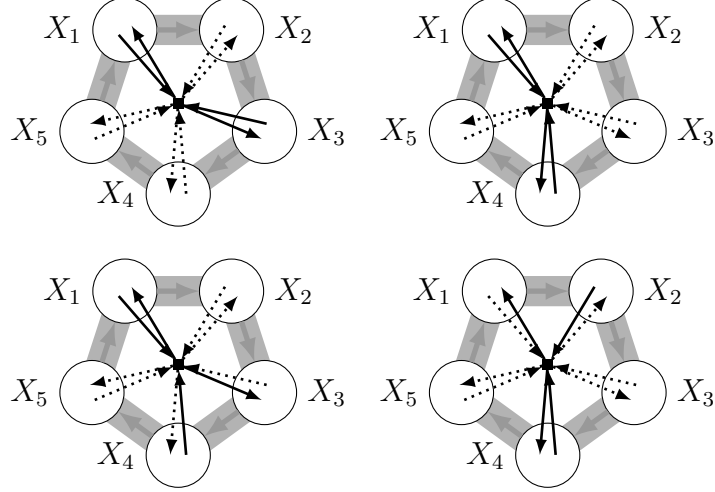
Figure 4: Possible neighborhoods of a vertex x, depicted in the center, in X_0. If solid edges are present, then dashed edges are not present. The solid edges cover all options if there is an edge from x to a vertex in X_1.

would also be in a $\vec{T}_3$ as z and z' are not adjacent to each other. Hence G' is $\vec{T}_3$-free. For $w \in [5]$, let $x_w = |X_w|$. By checking all possible embeddings of an induced $\vec{P}_4$, we calculate

$$\vec{P}_4(G, z) = x_a x_b x_c + x_a x_b x_i + x_a x_i x_c + x_i x_b x_c$$
$$\vec{P}_4(G, y) = x_a x_b x_c + x_a x_b x_j + x_a x_j x_c + x_j x_b x_c$$
$$\vec{P}_4(G, \{z, y\}) = x_a x_b + x_a x_c + x_b x_c$$

Notice that the induced $\vec{P}_4$'s containing only z' or y contribute to the difference of the number of induced $\vec{P}_4$'s in G' and G. In particular,

$$\begin{aligned}
\vec{P}_4(G') - \vec{P}_4(G) &= \vec{P}_4(G, z) - \vec{P}_4(G, y) - \vec{P}_4(G, \{z, y\}) \\
&= x_i(x_a x_b + x_a x_c + x_b x_c) - x_j(x_a x_b + x_a x_c + x_b x_c) - x_a x_b - x_a x_c - x_b x_c \\
&= (x_i - x_j - 1)(x_a x_b + x_a x_c + x_b x_c) \\
&\geq x_a x_b + x_a x_c + x_b x_c > 0,
\end{aligned}$$

which contradicts that $\vec{P}_4(G)$ is maximum. $\qquad\square$

The last remaining part of the proof of Theorem 3 is to show that it holds for all n divisible by 5. Assume that $n = 5\ell$ for some $\ell \geq 1$ and G is a graph maximizing the number of induced $\vec{P}_4$ among all n-vertex graphs. Our goal is to show that G is isomorphic to $\vec{C}_5(n)$, which is a balanced blow-up of $\vec{C}_5$ on n vertices. By the extremality of G, we get

$$\vec{P}_4(G) \geq \vec{P}_4(\vec{C}_5(n)) = 5\ell^4.$$

11

Now consider a blow-up B of G, where every vertex of G is replaced by j vertices. That is, B has $5j\ell$ vertices. Every $\vec{P}_4$ in G yields j^4 copies of $\vec{P}_4$ in B. Hence $\vec{P}_4(B) = 5j^4\ell^4$. If j is sufficiently large, we have already proved, that $\vec{C}_5(5j\ell)$ is the unique extremal construction. Hence

$$5j^4\ell^4 = \vec{P}_4(\vec{C}_5(5j\ell)) \geq \vec{P}_4(B) = 5j^4\ell^4.$$

Therefore, B is isomorphic to $\vec{C}_5(5j\ell)$. Since B was obtained as a blow-up of G, we conclude that G is isomorphic to $\vec{C}_5(n)$. This finishes the proof of Theorem 3. □

3 Longer directed paths

We use methods developed for determining the inducibility of cycles in non-oriented graphs in order to obtain bounds for oriented paths of arbitrary length. The first general upper bound shown in Lemma 14 utilizes an approach by Pippenger and Golumbic [31]. Hefetz and Tyomkyn [21] developed a more complicated approach, and Kráľ, Norin, and Volec [28] recently improved the result via a simple counting argument. We use the method from [28] for $\vec{T}_3$-free graphs, as proven in Lemma 15.

Lemma 14.

$$I(\vec{P}_k) \leq \frac{k!}{(k-1)^{k-1}}$$

Proof. Let G be an oriented graph on n vertices. We try to build a path $v_1, \ldots, v_k$ by starting at v_1 and trying to append one vertex at a time. We can choose v_1 to be any of the n vertices. Now in each step, let w_i be the number of candidates for v_i. That is, $w_1 = n$, $w_2 = |N^+(v_1)|$, $w_3 = |N^+(v_2) \setminus N(v_1)|$, and so on. Then, the total number of choices to build a path on k vertices is

$$\prod_{i=1}^{k} w_i = n \cdot \prod_{i=2}^{k} w_i \leq n \left(\frac{n}{k-1} \right)^{k-1} = \frac{n^k}{(k-1)^{k-1}}.$$

Therefore,

$$I(\vec{P}_k) \leq \lim_{n \to \infty} \frac{\frac{n^k}{(k-1)^{k-1}}}{\binom{n}{k}} = \frac{k!}{(k-1)^{k-1}}.$$

□

In the above proof of Lemma 14, the worst case of the calculation is achieved when $w_i = \frac{n}{k-1}$ for all i. Instead of naively building the path, we will consider different orderings of the path (this is a trick inspired by [28]) in order to modify the worst case to be $w_i = \frac{n}{k}$ for all i. This gives a further improvement on the bound, but it falls short of the best known construction, a blow-up of a $\vec{C}_{k+1}$.

Lemma 15.

$$I(\vec{P}_k, \vec{\mathcal{T}}) \leq \frac{k!}{k^{k-1}},$$

where $\vec{\mathcal{T}}$ is the family of $\vec{T}_3$-free oriented graphs.

Proof. This proof follows the approach developed in [28]. Let G be an oriented graph on n vertices. Let $T = (z_1, \ldots, z_k)$ be a k-tuple of vertices of G. We will consider $D_1(T), \ldots, D_k(T)$, where $D_i(T)$ denotes the following permuted k-tuple of T:

$$z_i, z_{i-1}, \ldots, z_3, z_2, z_1, z_{i+1}, z_{i+2}, \ldots, z_k.$$

Intuitively, we will think of the sequence $z_1, \ldots, z_k$ as an order of picking the vertices and $D_i(T)$ as an order in which these vertices form a copy of $\vec{P}_k$. We define a weight w as

$$w(D_i(T)) = \prod_{j=1}^{k} \frac{1}{n_{i,j}},$$

where $n_{i,1} = n$ and $n_{i,j}$ is the number of possible candidates for z_j subject to $z_1, \ldots, z_{j-1}$ are already chosen and the copy of $\vec{P}_k$ is being built according to $D_i(T)$.

For a fixed D_i, we call a k-tuple T *good*, if $D_i(T)$ induces a copy of $\vec{P}_k$ in the same order. Let D_i be fixed. By using reverse induction on m, the sum of the weights of all good k-tuples $(z_1, \ldots, z_k)$ with respect to D_i that starts with $(z_1, \ldots, z_m)$ is at most $\prod_{j=1}^{m} \frac{1}{n_{i,j}}$. Hence, the total sum of weight of all good k-tuples with respect to D_i is at most 1.

By summing over all $i \in \{1, \ldots, k\}$, we conclude that the sum of all weights of all k-tuples that are good for at least one D_i is at most k.

Let $v_1, v_2, \ldots, v_k$ be an induced $\vec{P}_k$ in G. For $i \in \{1, \ldots, k\}$, let

$$T_i = (v_i, v_{i-1}, \ldots, v_2, v_1, v_{i+1}, v_{i+2}, \ldots, v_k).$$

Notice that T_i is a good k-tuple for D_i. We will later show that

$$\frac{k^k}{n^k} \le w(D_1(T_1)) + w(D_2(T_2)) + \cdots + w(D_k(T_k)). \tag{4}$$

Since the contribution to the sum of the weights of all good k-tuples is at least $\frac{k^k}{n^k}$ for each $\vec{P}_k$ in G, and the total sum is at least k, we conclude that the number of induced $\vec{P}_k$'s is at most $\frac{n^k}{k^{k-1}}$. By considering the limit, we get

$$I(\vec{P}_k) \le \lim_{n \to \infty} \frac{\frac{n^k}{k^{k-1}}}{\binom{n}{k}} = \frac{k!}{k^{k-1}}.$$

It remains to prove (4). We will use the AM-GM inequality twice. The first use is

$$\left(\prod_{i=1}^{k} w(D_i(T_i)) \right)^{\frac{1}{k}} \le \frac{w(D_1(T_1)) + \cdots + w(D_k(T_k))}{k}. \tag{5}$$

13

The second comes in

$$\left(\prod_{i=1}^{k} \frac{1}{w(D_i(T_i))}\right)^{\frac{1}{k(k-1)}} = \left(\prod_{i=1}^{k} n \cdot n_{i,2} \cdot n_{i,3} \cdots n_{i,k}\right)^{\frac{1}{k(k-1)}}$$

$$= n^{\frac{1}{k-1}} \cdot \left(\prod_{i=1}^{k} n_{i,2} \cdot n_{i,3} \cdots n_{i,k}\right)^{\frac{1}{k(k-1)}}$$

$$\leq \frac{n^{\frac{1}{k-1}}}{k(k-1)} \cdot \left(\sum_{i=1}^{k} (n_{i,2} + \cdots + n_{i,k})\right) \tag{6}$$

Our next goal is to show that every vertex x of G can contribute at most one to $n_{i,2} + \cdots + n_{i,k}$ for each i in (6), and moreover, that there is one i, where x does not contribute at all. This would give that the big sum in (6) is upper bounded by $n(k-1)$.

If x has no neighbors among $v_1, \ldots, v_k$, then it does not contribute at all. Let a be the smallest index such that v_a and x are adjacent. If $xv_a \in E(G)$, then x does not contribute to $w(D_1(T_1))$. Let b be the largest index such that v_b and x are adjacent. If $v_b x \in E(G)$, then x does not contribute to $w(D_k(T_k))$. Hence assume $v_a x \in E(G)$ and $xv_b \in E(G)$. Since G is $\vec{T}_3$-free, x is not adjacent to v_{a+1} and hence it does not contribute to $w(D_{a+1}(T_{a+1}))$. Notice that if G was not $\vec{T}_3$-free, then it might be the case that $b = a+1$ and x would contribute to $w(D_{i+1}(T_{i+1}))$ for all $i \in \{1, \ldots, k\}$.

By replacing the big sum in (6) by its upper bound $n(k-1)$ we obtain

$$\left(\prod_{i=1}^{k} \frac{1}{w(D_i(T_i))}\right)^{\frac{1}{k(k-1)}} \leq \frac{n^{\frac{1}{k-1}}}{k(k-1)} \cdot n(k-1) = \frac{n^{\frac{k}{k-1}}}{k}$$

$$\left(\prod_{i=1}^{k} \frac{1}{w(D_i(T_i))}\right)^{\frac{1}{k}} \leq \frac{n^k}{k^{k-1}} \tag{7}$$

By combining the reciprocal of (7) and multiplying (5) by k, we obtain

$$\frac{k^k}{n^k} \leq k \cdot \left(\prod_{i=1}^{k} w(D_i(T_i))\right)^{\frac{1}{k}} \leq w(D_1(T_1)) + w(D_2(T_2)) + \cdots + w(D_k(T_k)),$$

which proves (4) and finishes the proof of Lemma 15. $\qquad\square$

4 Conclusion

Flag algebra calculations support Conjectures 1 and 2 for other small values of k. For Conjecture 1, we compute the following bounds:

$$0.1935483870 \approx \frac{6}{31} \leq I(\vec{P}_4) \leq 0.19356$$

$$0.092664092 \approx \frac{24}{259} \leq I(\vec{P}_5) \leq 0.092676$$

$$0.0428418421 \approx \frac{120}{2801} \leq I(\vec{P}_6) \leq 0.04323$$

When restricted to $\vec{T}_3$-free oriented graphs, we get the following exact results for paths on five and six vertices, respectively:

Theorem 16. *For the family $\vec{\mathcal{T}}$ of $\vec{T}_3$-free oriented graphs,*

$$I(\vec{P}_5, \vec{\mathcal{T}}) = \frac{5!}{6^4} \qquad\qquad I(\vec{P}_6, \vec{\mathcal{T}}) = \frac{6!}{7^5}$$

For $\vec{P}_7$, we compute a numerical upper bound matching Conjecture 2 when $n \to \infty$. This means that we successfully rounded numerical solution by flag algebras for $\vec{P}_5$ and $\vec{P}_6$, but fell short to do so for $\vec{P}_7$. We expect that the approach we used for $\vec{P}_4$ in this paper could also work for stability and exactness of $\vec{P}_5$ and $\vec{P}_6$.

Acknowledgements

References

[1] A. Aroskar and J. Cummings. Limits, Regularity and Removal for Finite Structures. To appear in Journal of Symbolic Logic. Preprint available as arXiv:1412.8084, 2014.

[2] R. Baber and J. Talbot. Hypergraphs do jump. *Combin. Probab. Comput.*, 20:161–171, 2011.

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

[4] 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.

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

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

[7] 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.*, 24:658–679, 2015.

[8] L. N. Coregliano, R. F. Parente, and C. M. Sato. On the maximum density of fixed strongly connected subtournaments. Preprint available as arXiv:1505.05200, 2015.

[9] L. N. Coregliano and A. A. Razborov. On the density of transitive tournaments. *J. Graph Theory*, 85:12–21, 2017.

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

[11] 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*, 103:344–373, 2013.

[12] C. Even-Zohar and N. Linial. A note on the inducibility of 4-vertex graphs. *Graphs Combin.*, 31:1367–1380, 2015.

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

[14] J. Fox, H. Huang, and C. Lee. A solution to the inducibility problem for almost all graphs. Manuscript.

[15] J. Fox, H. Huang, and C. Lee. Packing problems, 2015. `http://videos.birs.ca/2015/15w5008/201508240900-Fox.mp4`.

[16] E. Gethner, L. Hogben, B. Lidický, F. Pfender, A. Ruiz, and M. Young. On crossing numbers of complete tripartite and balanced complete multipartite graphs. *J. Graph Theory*, 84:552–565, 2017.

[17] R. Glebov, D. Král', and J. Volec. A problem of Erdős and Sós on 3-graphs. *Israel J. Math.*, 211:349–366, 2016.

[18] X. Goaoc, A. Hubard, R. de Joannis de Verclos, J.-S. Sereni, and J. Volec. Limits of order types. In *31st International Symposium on Computational Geometry*, volume 34 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages 300–314. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2015.

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

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

[21] D. Hefetz and M. Tyomkyn. On the inducibility of cycles. *J. Combin. Theory Ser. B*, 133:243–258, 2018.

[22] J. Hladký, D. Král', and S. Norin. Counting flags in triangle-free digraphs. *Combinatorica*, 37:49–76, 2017.

[23] J. Hladký, D. Ráľ, and S. Norin. Private communication.

[24] P. Hu, B. Lidický, F. Pfender, and J. Volec. Maximum density of induced oriented C_4 is achieved by iterated blow-ups. Manuscript.

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

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

[27] D. Král' and O. Pikhurko. Quasirandom permutations are characterized by 4-point densities. *Geom. Funct. Anal.*, 23:570–579, 2013.

[28] D. Ráľ, S. Norin, and J. Volec. A bound on the inducibility of cycles. *J. Combin. Theory Ser. A*, 161:359–363, 2019.

[29] B. Lidický and F. Pfender. Semidefinite programming and Ramsey numbers. Preprint available as arXiv:1704.03592, 2017.

[30] B. Lidický and F. Pfender. Pentagons in triangle-free graphs. *European J. Combin.*, 74:85–89, 2018.

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

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

[33] E. R. Vaughan. Flagmatic 2.0, 2012. `http://jakubsliacan.eu/flagmatic/`.

[34] R. Yuster. On the exact maximum induced density of almost all graphs and their inducibility. In press in J. Combin. Theory Ser. B, 2018.