

# Upper bounds on the extremal number of the 4-cycle

Jie Ma      Tianchi Yang

## Abstract

We obtain some new upper bounds on the maximum number  $f(n)$  of edges in  $n$ -vertex graphs without containing cycles of length four. This leads to an asymptotically optimal bound on  $f(n)$  for a broad range of integers  $n$  as well as a disproof of a conjecture of Erdős from 1970s which asserts that  $f(n) = \frac{1}{2}n^{3/2} + \frac{1}{4}n + o(n)$ .

## 1 Introduction

Let  $\text{ex}(n, C_4)$  denote the maximum number of edges in an  $n$ -vertex  $C_4$ -free graph.<sup>1</sup> The study of this extremal number can be dated back to Erdős [2] eighty years ago and has a rich, lasting influence on the development of extremal graph theory (see [8]).

It is well-known (see Kővári-Sós-Turán [12] and Reiman [13]) that for any positive integer  $n$ ,

$$\text{ex}(n, C_4) \leq \frac{n}{4}(1 + \sqrt{4n - 3}) = \frac{1}{2}n^{3/2} + \frac{n}{4} - O(n^{1/2}). \quad (1)$$

Notably, the well-known friendship theorem of Erdős-Rényi-Sós [5] shows that an equation can never hold in (1). On the other hand, using polarity graphs defined by finite projective planes, Brown [1] and Erdős-Rényi-Sós [5] independently proved the following famous lower bound that

$$\text{ex}(q^2 + q + 1, C_4) \geq \frac{1}{2}q(q + 1)^2 \text{ for all prime powers } q. \quad (2)$$

These two results together with some basic property on the distribution of prime numbers imply the asymptotic formula that  $\text{ex}(n, C_4) = (\frac{1}{2} + o(1))n^{3/2}$ .

Motivated by above results, Erdős raised several conjectures in 1970s to enrich the understanding on the extremal number  $\text{ex}(n, C_4)$ . In [3] Erdős conjectured that the lower bound (2) is sharp for prime powers  $q$ . This was fully resolved by Füredi in [6, 7], where he proved that

$$\text{ex}(q^2 + q + 1, C_4) \leq \frac{1}{2}q(q + 1)^2 \text{ for all integers } q \geq 14. \quad (3)$$

Results related to (3) can be found in [9, 11]; it is also worth noting that the bipartite version of the extremal number of  $C_4$  has been well studied in [10] (see Section 3.1 of [8]). If one substitutes  $n = q^2 + q + 1$  for prime powers  $q$  in (2), then it yields that  $\text{ex}(n, C_4) \geq \frac{1}{2}n^{3/2} + \frac{n}{4} - O(n^{1/2})$  for such integers  $n$ . Note that this meets the upper bound (1) up to the error term  $O(n^{1/2})$ . Erdős [3, 4] made the following tempting conjecture.

---

School of Mathematical Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China. Partially supported by the National Key R and D Program of China 2020YFA0713100, National Natural Science Foundation of China grant 12125106, and Anhui Initiative in Quantum Information Technologies grant AHY150200. Emails: jiema@ustc.edu.cn, ytc@mail.ustc.edu.cn.

<sup>1</sup>Throughout this paper, a graph is called  $C_4$ -free if it does not contain a cycle of length four as a subgraph.

**Conjecture 1.1** (Erdős [3, 4]). *It holds that*

$$\text{ex}(n, C_4) = \frac{1}{2}n^{3/2} + \frac{1}{4}n + o(n).$$

He also commented in [4] that “it is not impossible that the error term is  $O(n^{1/2})$ .”

The present paper aims to establish new upper bounds on  $\text{ex}(n, C_4)$  (see Theorems 1.3 and 1.5 below). Contrary to the supportive evidences, these bounds imply the following result, which shows that the above conjecture of Erdős does not hold in a strong sense.

**Theorem 1.2.** *There exist some real  $\epsilon > 0$  and a positive density of integers  $n$  such that*

$$\text{ex}(n, C_4) \leq \frac{1}{2}n^{3/2} + \left(\frac{1}{4} - \epsilon\right)n.$$

This will follow by our upper bounds on  $\text{ex}(n, C_4)$ . Our computation shows that  $\epsilon$  can be taken as any positive real less than 0.075. This number is unlikely to be tight, so we did not try to optimize our calculation as well as the constants appearing in the forthcoming results; see the concluding remarks for more discussion on the real  $\epsilon$ . To proceed, we introduce some notation. For integers  $q \geq 0$ , let  $I_q$  denote the set of  $2q + 1$  consecutive integers  $\{q^2 + 1, \dots, (q + 1)^2\}$ ; let  $I_q^- = \{q^2 + 1, \dots, q^2 + q\}$  and  $I_q^+ = \{q^2 + q + 2, \dots, (q + 1)^2\}$  so that  $I_q = I_q^- \cup \{q^2 + q + 1\} \cup I_q^+$ . Note that these  $I_q$ 's form a partition of the set of positive integers.

Our first result on the upper bound of  $\text{ex}(n, C_4)$  focuses on integers  $n$  from the sets  $I_q^-$ .

**Theorem 1.3.** *Let  $n = q^2 + q + 1 - r$  be an integer in  $I_q^-$  such that  $r \leq 0.01q$  is sufficiently large. Then*

$$\text{ex}(n, C_4) = \text{ex}(q^2 + q + 1 - r, C_4) \leq \frac{1}{2}q(q + 1)^2 - 0.92rq.$$

We point out that this bound can be further improved to the form (13) (see the remark after the proof of Theorem 1.3 in Section 3). As a corollary, this yields the following asymptotic bound.

**Corollary 1.4.** *Let  $q$  be a prime power and  $r = o(q)$  be sufficiently large. Then*

$$\text{ex}(q^2 + q + 1 - r, C_4) = \frac{1}{2}q(q + 1)^2 - (r + o(1))q.$$

Our second result on the upper bound of  $\text{ex}(n, C_4)$  considers integers  $n$  belonging to the sets  $I_q^+$ .

**Theorem 1.5.** *Let  $n = q^2 + q + 1 + r$  be a sufficiently large integer in  $I_q^+$  such that  $r \leq 0.6q$ . Then*

$$\text{ex}(n, C_4) = \text{ex}(q^2 + q + 1 + r, C_4) \leq \frac{1}{2}(q^2 + q + 1 + \max\{r, 2r - 0.3q\})(q + 1).$$

See Figure 1 for an illustration of the improvements on  $\text{ex}(n, C_4)$  in Theorems 1.3 and 1.5.<sup>2</sup>

The rest of the paper is structured as follows. In Section 2, we use Theorems 1.3 and 1.5 to prove Theorem 1.2. In Section 3, we introduce the key terminology, establish two useful lemmas and give a sketch for the proofs of the upper bounds on  $\text{ex}(n, C_4)$ . In Section 4, we show Theorem 1.3 and Corollary 1.4. In Section 5, we finish the proof of Theorem 1.5. Finally, in Section 6, we make some concluding remarks.

---

<sup>2</sup>The curve  $F(n) = \frac{n}{4}(1 + \sqrt{4n - 3})$  in the interval  $[q^2 - q + 1, q^2 + q + 1]$  is quite close to a straight line with slope  $0.75q$  (i.e., the line 1). The upper bound of Theorem 1.3 is indicated by the line 2 with some slope close to  $q$ , while the upper bounds of Theorem 1.5 are indicated by the line 3 and line 4 with slopes  $0.5q$  and  $q$ , respectively.

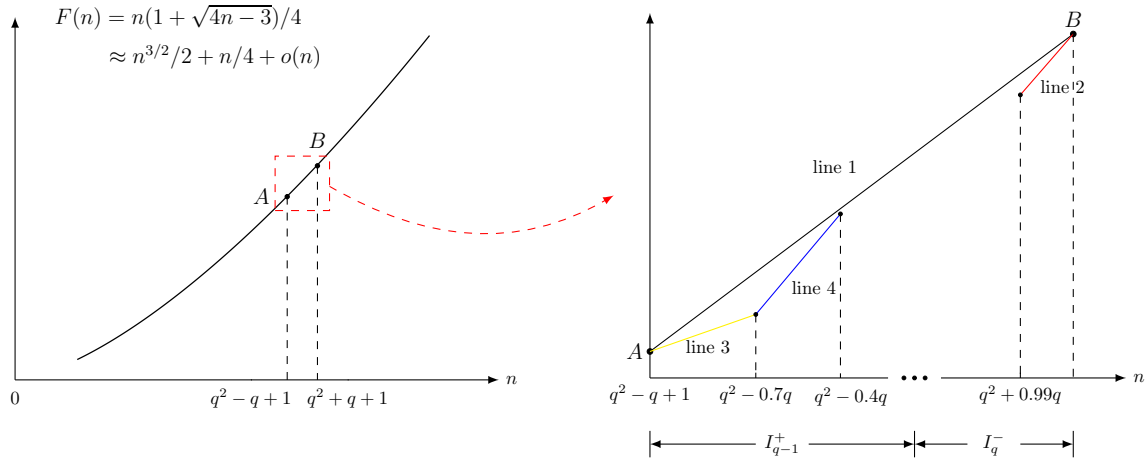


Figure 1: The old and new upper bounds on  $\text{ex}(n, C_4)$ .

## 2 Proof of Theorem 1.2

Assuming Theorems 1.3 and 1.5, we now present a short proof of Theorem 1.2. (In fact, either Theorem 1.3 or Theorem 1.5 suffices to derive the statement of Theorem 1.2.)

**Proof of Theorem 1.2.** Let  $\epsilon > 0$  be some small absolute constant (i.e.,  $\epsilon = 0.001$  suffices). Let  $n = q^2 + q + 1 + r \in I_q$ . So we have  $-q \leq r \leq q$ . By some straightforward calculations, it holds that

$$\frac{1}{2}n^{3/2} + \frac{1}{4}n = \frac{1}{2}q(q+1)^2 + \frac{3}{4}rq + O(q). \quad (4)$$

Let  $N_1 = \{q^2 + q + 1 - r : 6\epsilon \leq r/q \leq 0.01\}$ . Consider any sufficiently large integer  $n = q^2 + q + 1 - r \in N_1$ . So  $q$  and (thus)  $r$  are sufficiently large as well. We claim that every such  $n \in N_1$  satisfies that  $\text{ex}(n, C_4) \leq \frac{1}{2}n^{3/2} + (\frac{1}{4} - \epsilon)n$ . To see this, using Theorem 1.3, we can derive that

$$\text{ex}(n, C_4) \leq \frac{1}{2}q(q+1)^2 - 0.92rq = \left(\frac{1}{2}n^{3/2} + \frac{1}{4}n\right) - 0.17rq + O(q) \leq \frac{1}{2}n^{3/2} + \left(\frac{1}{4} - \epsilon\right)n,$$

where the equation holds by (4) and the last inequality holds since  $r \geq 6\epsilon q$  and  $q$  is sufficiently large. We note that the set  $N_1$  has a positive density  $\lim_{n \rightarrow \infty} \frac{|N_1 \cap [n]|}{n} = (0.01 - 6\epsilon)/2 > 0$ .

Consider another set of integers  $N_2 = \{q^2 + q + 1 + r : 5\epsilon \leq r/q \leq 0.3\}$ , which has a positive density  $(0.3 - 5\epsilon)/2 > 0$ . We claim that any sufficiently large  $n \in N_2$  also has  $\text{ex}(n, C_4) \leq \frac{1}{2}n^{3/2} + (\frac{1}{4} - \epsilon)n$ . Indeed, by Theorem 1.5 and the above inequality (4), we have

$$\text{ex}(n, C_4) \leq \frac{1}{2}(q^2 + q + 1 + r)(q+1) = \left(\frac{1}{2}n^{3/2} + \frac{1}{4}n\right) - \frac{1}{4}rq + O(q) \leq \frac{1}{2}n^{3/2} + \left(\frac{1}{4} - \epsilon\right)n,$$

where the last inequality holds because  $r \geq 5\epsilon q$ . This completes the proof of Theorem 1.2.  $\square$

## 3 Preliminaries

First we introduce the key notation on  $C_4$ -free graphs  $G$  for the coming proofs.

**Definition 3.1.** Let  $n \in I_q$  and  $G$  be an  $n$ -vertex  $C_4$ -free graph. As usual, we write  $d(v)$  for the degree of a vertex  $v$  in  $G$ . The deficiency of a vertex  $v$  in  $G$  is defined by

$$f(v) = q + 1 - d(v).$$

The deficiency of a subset  $A \subseteq V(G)$  is given by  $f(A) = \sum_{v \in A} f(v)$ . For each integer  $i \geq 0$ , we denote  $S_i$  by the set of vertices of degree  $i$  in  $G$ . Finally, we let  $S = \cup_{i \leq q} S_i$  and  $S^+ = \cup_{j \geq q+2} S_j$ .

Note that  $V(G) = S \cup S_{q+1} \cup S^+$ , and a vertex  $v$  is in  $S$  (or in  $S^+$ ) if and only if it has positive (or negative) deficiency. Next we prove some estimations on vertex degrees in  $C_4$ -free graphs.

**Lemma 3.2.** Let  $n \in I_q$  and  $G$  be an  $n$ -vertex  $C_4$ -free graph. If we write  $N(v)$  for the neighborhood of a vertex  $v$  in  $G$ , then we have  $f(N(v)) \geq qd(v) - n + 1$ .

*Proof.* As  $G$  is  $C_4$ -free, all subsets  $N(v_i) \setminus \{v\}$  for  $v_i \in N(v)$  are disjoint. So we have  $\sum_{v_i \in N(v)} (d(v_i) - 1) = \sum_{v_i \in N(v)} |N(v_i) \setminus \{v\}| \leq n - 1$ . This implies that

$$f(N(v)) = \sum_{v_i \in N(v)} (q + 1 - d(v_i)) = qd(v) - \sum_{v_i \in N(v)} (d(v_i) - 1) \geq qd(v) - n + 1,$$

as desired.  $\square$

The coming lemma provides a crucial technical tool for later sections. Its proof idea is rooted in a lemma of [6].

**Lemma 3.3.** Let  $G$  be an  $n$ -vertex  $C_4$ -free graph with  $m$  edges. Let  $v_i$ 's for  $i \in [n]$  be vertices of  $G$  of degree  $d_i$ . If there exists some vertex  $v$  such that  $\{v_i\}_{i \in I} \subseteq N(v)$  for  $I \subseteq [n]$  with  $|I| = k$ , then

$$\binom{n - \sum_{i \in I} d_i + k - 1}{2} \geq (n - k) \binom{\frac{2m - \sum_{i \in I} d_i + (k-1)d(v) - nk + k}{n-k}}{2}.$$

*Proof.* We say  $v_i v_j v_l$  is a 2-path if  $v_i v_j, v_j v_l \in E(G)$ . Let  $X = \bigcup_{i \in I} N(v_i)$  and we count the number  $M$  of 2-paths with both end-points in  $V(G) \setminus X$ . Since  $G$  is  $C_4$ -free, all  $N(v_i) \setminus \{v\}$  are disjoint, implying that  $|X| = \sum_{i \in I} (|N(v_i)| - 1) + 1 = \sum_{i \in I} d_i - k + 1$ . So it is evident that

$$M \leq \binom{n - |X|}{2} = \binom{n - \sum_{i \in I} d_i + k - 1}{2}.$$

On the other hand, for any  $j \in [n]$  there are  $\binom{|N(v_j) \setminus X|}{2}$  many counted 2-paths with the middle-point  $v_j$ . For any  $j \notin I$  we have  $|N(v_j) \setminus X| \geq d_j - k$  and moreover, for  $v_j \in N(v) \setminus \{v_i\}_{i \in I}$ , we have  $|N(v_j) \setminus X| = |N(v_j) \setminus \{v\}| = d_j - 1$ . So it follows by Jensen's inequality that

$$M = \sum_{j \notin I} \binom{|N(v_j) \setminus X|}{2} \geq \sum_{v_j \in N(v) \setminus \{v_i\}_{i \in I}} \binom{d_j - 1}{2} + \sum_{v_j \notin N(v)} \binom{d_j - k}{2} \geq (n - k) \binom{\frac{L}{n-k}}{2},$$

where  $L = \sum_{v_j \in N(v) \setminus \{v_i\}_{i \in I}} (d_j - 1) + \sum_{v_j \notin N(v)} (d_j - k) = \sum_{j \notin I} d_j - (d(v) - k) - k(n - d(v)) = 2m - \sum_{i \in I} d_i + (k - 1)d(v) - nk + k$ . Putting the above together, we finish the proof.  $\square$

Lastly, we would like to give an outline of the proof of Theorem 1.3 (the proof of Theorem 1.5 can be proved by a similar approach). The proof is inspired by the work of Füredi [6, 7]. There are several new ingredients as well. The first ingredient comes from Lemma 4.1, which roughly says that one only needs to consider  $C_4$ -free graphs with very large minimum degree. Another ingredient is the deficiency  $f(\cdot)$  in Definition 3.1.<sup>3</sup> Under this definition, we have  $f(V(G)) = O(rq)$  and we can maximize the benefits of this estimate. Putting these ingredients together with the technical tool Lemma 3.3, we prove in Lemma 4.2 that the size of  $S^+$  (i.e., the number of vertices whose degree exceeds  $q + 1$ ) can be bounded from above by  $O(r^2)$ . This will eventually lead to a contradiction by assigning appropriate weights to the edges between the sets  $S_{q+1}$  and  $S$ .

## 4 Proof of Theorem 1.3

We first show that restricted to the range of integers considered in Theorem 1.3, one may always assume that there exists some *extremal graph* with large minimum degree.<sup>4</sup>

**Lemma 4.1.** *Let  $q, r$  be integers satisfying  $1 \leq r \leq 0.3q$ . Assume that  $\text{ex}(q^2 + q + 1 - r, C_4) \geq \frac{1}{2}q(q + 1)^2 - \alpha rq$ , where  $0.2 \leq \alpha \leq 1$ . Then there exists an integer  $r_0 \in [r, 3r]$  such that*

$$\text{ex}(q^2 + q + 1 - r_0, C_4) \geq \frac{1}{2}q(q + 1)^2 - \alpha r_0 q$$

and

$$\text{ex}(q^2 + q + 1 - r_0, C_4) - \text{ex}(q^2 + q - r_0, C_4) \geq 0.2q.$$

In particular, the latter inequality shows that the minimum degree of any extremal graph on  $q^2 + q + 1 - r_0$  vertices is at least  $0.2q$ .

*Proof.* Suppose for a contradiction that there does not exist such an integer  $r_0 \in [r, 3r]$ . So we have  $\text{ex}(q^2 + q + 1 - r, C_4) - \text{ex}(q^2 + q - r, C_4) < 0.2q$ . Let  $s$  be the largest integer in  $[r + 1, 3r]$  such that

$$\text{ex}(q^2 + q + 1 - x, C_4) - \text{ex}(q^2 + q - x, C_4) < 0.2q \text{ holds for all integers } x \in [r, s - 1]. \quad (5)$$

Then either (i)  $s = 3r$  or (ii)  $\text{ex}(q^2 + q + 1 - s, C_4) - \text{ex}(q^2 + q - s, C_4) \geq 0.2q$ . Summing up (5) for all integers  $x \in [r, s - 1]$ , we obtain

$$\text{ex}(q^2 + q + 1 - s, C_4) > \text{ex}(q^2 + q + 1 - r, C_4) - 0.2(s - r)q \geq \frac{1}{2}q(q + 1)^2 - \alpha rq - 0.2(s - r)q. \quad (6)$$

As  $\alpha \geq 0.2$ , we see  $\text{ex}(q^2 + q + 1 - s, C_4) \geq \frac{1}{2}q(q + 1)^2 - \alpha sq$ . If (ii) occurs, then  $s$  would be a desired integer. So (i) occurs, i.e.,  $s = 3r$ . Since  $q \geq 4$  and  $\alpha \leq 1 \leq r$ , we have  $\frac{3}{4}sq - \frac{1}{2}(q + 1) - \alpha rq - 0.2(s - r)q = q \cdot (1.85r - \alpha r - 1/2 - 1/q) > 0$ . This inequality together with (6) give that

$$\text{ex}(q^2 + q + 1 - s, C_4) > \frac{1}{2}(q + 1)(q^2 + q + 1) - \frac{3}{4}sq. \quad (7)$$

Let  $h(x) = \frac{x}{4}(1 + \sqrt{4x - 3})$ . By (1), we see that  $\text{ex}(n, C_4) \leq h(n)$  holds for all  $n$ . Note that  $h(q^2 + q + 1) = \frac{1}{2}(q + 1)(q^2 + q + 1)$  and  $h'(x) = \frac{1 + \sqrt{4x - 3}}{4} + \frac{x}{2\sqrt{4x - 3}} > \frac{3}{4}\sqrt{x}$  for  $x \geq 1$ . So

<sup>3</sup>In literature where certain circumstances apply, the deficiency of a vertex  $v$  is defined by  $\max\{q + 1 - d(v), 0\}$ .

<sup>4</sup>Throughout the rest of the paper, a graph  $G$  is called an *extremal graph* if it is  $C_4$ -free and has the maximum number  $\text{ex}(|V(G)|, C_4)$  of edges.

$h(n+1) - h(n) \geq h'(n) > \frac{3}{4}\sqrt{n} \geq \frac{3}{4}q$  holds for every integer  $n \geq q^2 + q + 1 - s \geq q^2$ , where  $s = 3r \leq 0.9q$ . Adding up the above inequality for all integers  $n$  in  $[q^2 + q + 1 - s, q^2 + q]$ , we have

$$h(q^2 + q + 1 - s) < h(q^2 + q + 1) - \frac{3}{4}q \cdot s = \frac{1}{2}(q+1)(q^2 + q + 1) - \frac{3}{4}sq < \text{ex}(q^2 + q + 1 - s, C_4),$$

where the last inequality is from (7). Clearly, this contradicts the bound  $\text{ex}(n, C_4) \leq h(n)$ . This final contradiction shows that the desired  $r_0 \in [r, 3r]$  indeed exists, completing the proof.  $\square$

Recall Definition 3.1. In the next lemma we show that in  $C_4$ -free graphs with large minimum degree and sufficiently many edges, the size of  $S^+$  can be bounded from above.

**Lemma 4.2.** *Let  $q$  be a sufficiently large integer and  $r$  be an integer satisfying  $0 \leq r \leq 0.033q$ . Let  $G$  be a  $C_4$ -free graph on  $q^2 + q + 2 - r$  vertices, with at least  $\frac{1}{2}q(q+1)^2 - rq$  edges and of minimum degree at least  $0.2q$ . Then  $|S^+| \leq -f(S^+) \leq 4r^2 + 16r + 18$ .*

*Proof.* First we claim that for every vertex  $v$ ,  $|N(v) \cap S^+| \leq 3r + 8$ . Suppose not. Let  $k = 3r + 9$ . Then there exist vertices  $v$  and  $v_i$ 's for  $i \in [k]$  such that every  $v_i \in N(v) \cap S^+$ . Let  $d_i$  be the degree of  $v_i$  and let  $s = \sum_{i \in [k]} d_i$ . Since  $s - k = \sum_{i \in [k]} |N(v_i) \setminus \{v\}| \leq n - 1$ , we have  $k(q+2) \leq s \leq n - 1 + k \leq q^2 + 2q$ . Let  $n = q^2 + q + 2 - r$  and  $m_0 = \frac{1}{2}q(q+1)^2 - rq$ . By the proof of Appendix A, we get

$$F(q, r, s) := 2(n - k) \binom{n - s + k - 1}{2} - 2(n - k)^2 \binom{\frac{2m_0 - s + (k-1)0.2q - nk + k}{n-k}}{2} < 0. \quad (8)$$

Since  $e(G) \geq m_0$  and  $d(v) \geq 0.2q$ , using Lemma 3.3, we can derive that

$$0 > \frac{F(q, r, s)}{2(n - k)} \geq \binom{n - s + k - 1}{2} - (n - k) \binom{\frac{2e(G) - s + (k-1)d(v) - nk + k}{n-k}}{2} \geq 0,$$

a contradiction. This proves the claim that  $|N(v) \cap S^+| \leq 3r + 8$  holds for every vertex  $v$ .

Next we consider the deficiency. Let  $u \in S^+$ . Clearly  $f(u) \leq -1$ . By Lemma 3.2, we have

$$f(N(u)) \geq qd(u) - n + 1 = -qf(u) + r - 1 \geq (q - 1)(-f(u)). \quad (9)$$

Now define a weight function  $w$  on the edges  $uv$  with  $u \in S^+$  and  $v \in S$  by assigning  $w(uv) = f(v)$ . Let us count the total weight  $W$  of these edges. Every vertex  $u \in S^+$  contributes at least  $f(N(u) \cap S) \geq f(N(u))$ , so we can use (9) to obtain that

$$W \geq \sum_{u \in S^+} f(N(u)) \geq (q - 1) \sum_{u \in S^+} (-f(u)) = (q - 1)(-f(S^+)).$$

On the other hand, by the above claim, each vertex  $v \in S$  has at most  $3r + 8$  neighbors in  $S^+$ , so it contributes at most  $3r + 8$  times of its deficiency. Putting them together, we have

$$(q - 1)(-f(S^+)) \leq W \leq \sum_{v \in S} (3r + 8)f(v) = (3r + 8)f(S).$$

By the definition of  $f$ , we get  $f(S) + f(S^+) = f(V(G)) = n(q+1) - 2e(G) \leq (q^2 + q + 2 - r)(q+1) - (q(q+1)^2 - 2rq) = (r+2)q - r + 2$ . Therefore, we can derive

$$(q - 1)(-f(S^+)) \leq (3r + 8)f(S) \leq (3r + 8)((r+2)q - r + 2 - f(S^+)).$$

Further rearranging this inequality gives that

$$-f(S^+) \leq \frac{(3r+8)((r+2)q-r+2)}{q-3r-9} \leq \frac{(3r+8)(r+2)(q+1)}{0.9(q+1)} \leq 4r^2 + 16r + 18,$$

where the last two inequalities hold because  $0 \leq r \leq 0.033q$  and  $q$  is large. This finishes the proof.  $\square$

Now we are ready to prove Theorem 1.3.

**Proof of Theorem 1.3.** Let  $q, r$  be sufficiently large integers satisfying  $r \leq 0.01q$ . Suppose for a contradiction that  $\text{ex}(q^2 + q + 1 - r, C_4) \geq \frac{1}{2}q(q+1)^2 - \alpha r q$ , where  $\alpha = 0.92$ . By Lemma 4.1, there exists some integer  $r_0$  with  $r \leq r_0 \leq 3r \leq 0.03q$  such that any extremal graph  $G$  on  $q^2 + q + 1 - r_0$  vertices has at least  $\frac{1}{2}q(q+1)^2 - \alpha r_0 q$  edges and minimum degree at least  $0.2q$ .

Let  $n = q^2 + q + 1 - r_0$ . By the definition of deficiency,

$$\begin{aligned} f(V(G)) &= (q+1)n - 2e(G) \leq (q+1)(q^2 + q + 1 - r_0) - (q(q+1)^2 - 2\alpha r_0 q) \\ &= (2\alpha - 1)r_0 q - r_0 + q + 1 \leq (2\alpha - 1)r_0 q + q \end{aligned}$$

Applying Lemma 4.2 to  $G$ , we have  $|S^+| \leq -f(S^+) \leq 4(r_0 + 1)^2 + 16(r_0 + 1) + 18 = 4r_0^2 + O(q)$ . So

$$|S| \leq f(S) = f(V(G)) - f(S^+) \leq (2\alpha - 1)r_0 q + 4r_0^2 + O(q). \quad (10)$$

We also have

$$|S_{q+1}| = (q^2 + q + 1 - r_0) - |S^+| - |S| \geq \left(1 - (2\alpha - 1)\frac{r_0}{q} - 8\frac{r_0^2}{q^2}\right) q^2 + O(q). \quad (11)$$

For any vertex  $u \in S_{q+1}$ , by Lemma 3.2 we can derive

$$f(N(u)) \geq qd(u) - n + 1 \geq q(q+1) - (q^2 + q + 1 - r_0) + 1 = r_0.$$

Now define a weight function  $w$  on the edges  $uv$  with  $u \in S_{q+1}$  and  $v \in S$  by assigning  $w(uv) = f(v)$ . Let  $W$  be the total weight of these edges. Every vertex  $u \in S_{q+1}$  contributes at least  $f(N(u) \cap S) \geq f(N(u))$  to  $W$ , while each vertex  $v \in S$  contributes at most  $f(v)d(v) \leq f(v)q$ . Thus

$$|S_{q+1}|r_0 \leq \sum_{u \in S_{q+1}} f(N(u)) \leq W \leq \sum_{v \in S} f(v)q \leq f(S)q. \quad (12)$$

Using (10) and (11), we have that

$$\begin{aligned} f(S)q - |S_{q+1}|r_0 &\leq ((2\alpha - 1)r_0 q + 4r_0^2)q - \left(1 - (2\alpha - 1)\frac{r_0}{q} - 8\frac{r_0^2}{q^2}\right) r_0 q^2 + O(q^2) \\ &= \left(8\frac{r_0^2}{q^2} + (2\alpha + 3)\frac{r_0}{q} + 2\alpha - 2\right) r_0 q^2 + o(r_0 q^2), \end{aligned}$$

where  $r_0 \geq r$  is sufficiently large. Recall that  $\alpha = 0.92$  and  $r_0 \leq 0.03q$ . We see  $F(\frac{r_0}{q}) = 8\frac{r_0^2}{q^2} + (2\alpha + 3)\frac{r_0}{q} + 2\alpha - 2$  is a quadratic function on  $\frac{r_0}{q}$  with a negative axis of symmetry. So  $F(\frac{r_0}{q}) \leq F(0.03) = -0.0076 < 0$ . Therefore  $f(S)q - |S_{q+1}|r_0 \leq (F(\frac{r_0}{q}) + o(1)) \cdot r_0 q^2 < 0$ , which contradicts to (12). This completes the proof of Theorem 1.3.  $\square$

We would like to remark that the above proof can be modified to show: for any  $\epsilon > 0$ ,

$$\text{ex}(q^2 + q + 1 - r, C_4) \leq \frac{1}{2}q(q+1)^2 - (1 - \epsilon)rq \quad (13)$$

holds whenever  $r/q = O(\epsilon)$  and  $r = \Omega(1/\epsilon)$ .

To conclude this section, we now prove Corollary 1.4 using the above inequality.

**Proof of Corollary 1.4.** Let  $\epsilon > 0$  be any real. Let  $q$  be a prime power and  $r$  be an integer such that  $r = O(\epsilon q)$  and  $r = \Omega(1/\epsilon)$ . By results of [1, 5, 6, 7], there exists some extremal graph  $G$  on  $q^2 + q + 1$  vertices with  $\frac{1}{2}q(q+1)^2$  edges, which has exactly  $q+1$  vertices of degree  $q$  and all other vertices of degree  $q+1$ . Deleting any  $r$  vertices of degree  $q$  in  $G$ , one can obtain a  $C_4$ -free graph on  $q^2 + q + 1 - r$  vertices with at least  $\frac{1}{2}q(q+1)^2 - rq$  edges. This together with (13) show that

$$\frac{1}{2}q(q+1)^2 - rq \leq \text{ex}(q^2 + q + 1 - r, C_4) \leq \frac{1}{2}q(q+1)^2 - (1 - \epsilon)rq,$$

proving the corollary.  $\square$

## 5 Proof of Theorem 1.5

Let  $q$  be a sufficiently large integer. We first prove that for any integer  $r \in [1, 0.3q]$ , it holds that

$$\text{ex}(q^2 + q + 1 + r, C_4) \leq \frac{1}{2}(q^2 + q + 1 + r)(q + 1). \quad (14)$$

Suppose for a contradiction that there exists some integer  $r \in [1, 0.3q]$  satisfying  $\text{ex}(q^2 + q + 1 + r, C_4) > \frac{1}{2}(q^2 + q + 1 + r)(q + 1)$ . Let  $G$  be an extremal graph on  $q^2 + q + 1 + r$  vertices. We will complete the proof of (14) by deriving a contradiction that  $e(G) = \frac{1}{2}(q^2 + q + 1 + r)(q + 1)$ .

To see this, we first claim that every vertex  $v$  in  $G$  with degree  $d(v) \geq 0.7q$  satisfies  $|N(v) \cap S^+| < 0.55q$ . Suppose not. Let  $k = 0.55q$ . Then there exist vertices  $v$  and  $v_i$ 's for  $i \in [k]$  such that every  $v_i \in N(v) \cap S^+$ . Let  $n = q^2 + q + 1 + r$  and  $s = \sum_{i \in [k]} d_i$ . Since  $\sum_{i \in [k]} (d_i - 1) = \sum_{i \in [k]} |N(v_i) \setminus \{v\}| \leq n - 1$ , we know  $k(q + 2) \leq s \leq n - 1 + k \leq q^2 + 2q$ . Let  $m_0 = \frac{1}{2}n(q + 1)$  so that  $e(G) \geq m_0$ . By the calculations in Appendix B, we get

$$G(q, r, s) := 2(n - k) \binom{n - s + k - 1}{2} - 2(n - k)^2 \binom{\frac{2m_0 - s + (k-1)0.7q - nk + k}{n - k}}{2} < 0. \quad (15)$$

Note that  $d(v) \geq 0.7q$  and  $e(G) \geq m_0$ . We can use Lemma 3.3 to derive that

$$0 > \frac{G(q, r, s)}{2(n - k)} \geq \binom{n - s + k - 1}{2} - (n - k) \binom{\frac{2e(G) - s + (k-1)d(v) - nk + k}{n - k}}{2} \geq 0,$$

a contradiction. So indeed,  $|N(v) \cap S^+| < 0.55q$  holds for any vertex  $v$  with  $d(v) \geq 0.7q$  in  $G$ .

Now consider the deficiency of vertices. Any  $u \in S^+$  has  $f(u) \leq -1$ . By Lemma 3.2, we have

$$\begin{aligned} f(N(u)) &\geq qd(u) - n + 1 = q(q + 1 - f(u)) - (q^2 + q + 1 + r) + 1 \\ &= -f(u)q - r \geq (-f(u)) \cdot (q - r). \end{aligned} \quad (16)$$



Define a weight function  $w$  on the edges  $uv$  with  $u \in S^+$  and  $v \in S$  by assigning  $w(uv) = f(v)$ . Let  $W$  be the total weight of these edges. As every vertex  $u \in S^+$  contributes at least  $f(N(u) \cap S) \geq f(N(u))$ , from (16) we can get that

$$W \geq \sum_{u \in S^+} f(N(u)) \geq (q-r) \sum_{u \in S^+} (-f(u)) = (q-r)(-f(S^+)).$$

On the other hand, as we know  $|N(v) \cap S^+| < 0.55q$  for vertices  $v$  with  $d(v) \geq 0.7q$ , we see that in fact every vertex  $v \in S$  has at most  $0.7q - 1$  neighbors in  $S^+$ . So it contributes at most  $0.7q - 1$  times of its deficiency. Putting the above together, we get

$$(q-r)(-f(S^+)) \leq W \leq \sum_{v \in S} ((0.7q-1)f(v)) = (0.7q-1)f(S).$$

As  $1 \leq r \leq 0.3q$ , this implies that

$$-f(S^+) \leq f(S), \tag{17}$$

where the equality holds if and only if  $-f(S^+) = f(S) = 0$ . By the definition of  $f$ , we have  $f(S) + f(S^+) = f(V(G)) = n(q+1) - 2e(G) \leq 0$ . That is,  $f(S) \leq -f(S^+)$ . In view of (17), we can then derive  $-f(S^+) = f(S) = 0$ , which says that  $S = S^+ = \emptyset$ . This also says that all vertices in  $G$  have degree  $q+1$ , so  $e(G) = \frac{1}{2}n(q+1) = \frac{1}{2}(q^2 + q + 1 + r)(q+1)$ . However, this contradicts our assumption on the choice of  $G$ , thus completing the proof of (14).

It remains to consider when  $0.3q < r \leq 0.6q$ . In fact for any  $n \in I_q^+$ , by (1) that  $\text{ex}(n, C_4) \leq \frac{n}{4}(1 + \sqrt{4n-3})$ , we see that the minimum degree of any  $n$ -vertex extremal graph is at most  $q+1$ . So it follows that  $\text{ex}(n, C_4) \leq \text{ex}(n-1, C_4) + (q+1)$  for any  $n \in I_q^+$ . Then for  $0.3q < r \leq 0.6q$ , we can derive that

$$\text{ex}(q^2 + q + 1 + r, C_4) \leq \text{ex}(q^2 + q + 1 + 0.3q, C_4) + (r - 0.3q)(q+1) \leq \frac{1}{2}(q^2 + q + 1 + 2r - 0.3)(q+1),$$

where the last inequality follows by (14). This finishes the proof of Theorem 1.5.  $\square$

## 6 Concluding remarks

This paper mainly concerns upper bounds of the extremal number  $\text{ex}(n, C_4)$ . We prove that the upper bound estimations in Theorems 1.3 and 1.5 hold for a positive proportion of integers in  $I_q^-$  and  $I_{q-1}^+$ , respectively. We actually speculate that the upper bound of Theorem 1.3 and the restricted form of Theorem 1.5 for  $r \leq 0.3q$  (i.e., the line 2 and the line 3 in Figure 1) hold for all but  $o(q)$  integers in  $I_q^- \cup I_{q-1}^+$ . This would be one step closer to the precise value of  $\text{ex}(n, C_4)$  and in particular, would imply that the real  $\epsilon$  in Theorem 1.2 can be chosen to be any positive real less than  $1/4$  for a positive density of integers  $n$ .

The proof of Theorem 1.3 indicates that it would be very helpful for estimating the extremal number if the high minimum degree condition can be provided. However, we tend to believe that this approach does not work in general (we plan to explore this and related topics in a coming paper).

**Acknowledgements.** The authors would like to thank Zoltán Füredi for helpful discussions and for bringing the reference [10] to our attention. The authors are also grateful to Boris Bukh for many kind suggestions.

## References

- [1] W. G. Brown, On graphs that do not contain a Thomsen graph, *Canad. Math. Bull.* **9** (1966), 281–289.
- [2] P. Erdős, On sequences of integers no one of which divides the product of two others and some related problems, *Izvestiya Naustno-Issl. Inst. Mat. i Meh. Tomsk* **2** (1938), 74–82. (Zbl. **20**, p.5)
- [3] P. Erdős, Problems and results in combinatorial analysis, *Colloquio Internazionale sulle Teorie Combinatorie* (Rome, 1973), Tomo II, Atti dei Convegni Lincei, No. 17, pp. 3–17, Accad. Naz. Lincei, Rome, 1976.
- [4] P. Erdős, Some extremal problems on families of graphs and related problems, *Combinatorial Mathematics* (Proc. Internat. Conf. Combinatorial Theory, Australian Nat. Univ., Canberra, 1977), Lecture Notes in Math., 686. Springer, Berlin (1978) 13–21.
- [5] P. Erdős, A. Rényi and V. T. Sós, On a problem of graph theory, *Studia Sci. Math. Hungar.* **1** (1966), 215–235.
- [6] Z. Füredi, Graphs without quadrilaterals, *J. Combin. Theory Ser. B* **34** (1983), 187–190.
- [7] Z. Füredi, On the number of edges of quadrilateral-free graphs, *J. Combin. Theory Ser. B* **68** (1996), 1–6.
- [8] Z. Füredi and M. Simonovits, The history of degenerate (bipartite) extremal graph problems, in *Erdős centennial*, volume 25 of *Bolyai Soc. Math. Stud.*, pages 169–264. János Bolyai Math. Soc., Budapest, 2013.
- [9] F. A. Firke, P. M. Kosek, E. D. Nash and J. Williford, Extremal graphs without 4-cycles, *J. Combin. Theory Ser. B* **103** (2013), 327–336.
- [10] G. Damásdi, T. Héger and T. Szőnyi, The Zarankiewicz problem, cages, and geometries, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **56** (2013), 3–37.
- [11] J. He, J. Ma and T. Yang, Some extremal results on 4-cycles, *J. Combin. Theory Ser. B* **149** (2021), 92–108.
- [12] T. Kővári, V. Sós and P. Turán, On a problem of K. Zarankiewicz, *Colloquium Math.* **3** (1954), 50–57.
- [13] I. Reiman, Über ein Problem von K. Zarankiewicz, *Acta Math. Acad. Sci. Hungar.* **9** (1958), 269–278.

## Appendices

### A. Justification of the inequality (8)

Here we present a detailed proof for the inequality (8). Note that  $n = q^2 + q + 2 - r$ ,  $k = 9 + 3r$ ,  $m_0 = \frac{1}{2}q(q+1)^2 - rq$ ,  $k(q+2) \leq s \leq q^2 + 2q$  and  $0 \leq r \leq 0.03q$ . Also we have

$$\begin{aligned} F(q, r, s) &= (n - s + k - 1)(n - s + k - 2)(n - k) \\ &\quad - (2m_0 - s + (k - 1)0.2q - nk + k)(2m_0 - s + (k - 1)0.2q - nk - n + 2k). \end{aligned}$$

It suffices to show that under the conditions, namely,  $q$  is sufficiently large,  $k(q+2) \leq s \leq q^2 + 2q$  and  $0 \leq r \leq 0.03q$ , the maximum value  $F_{\max}$  of  $F(q, r, s)$  is less than 0. For fixed  $q$  and  $r$ ,  $F$  is a quadratic function with variable  $s$ . The coefficients of  $s^2$  and  $s$  are  $q^2 + q - 4r - 8$  and  $-2q^4 - 2q^3 + q^2(-2r - 22) + q(-4.8r - 18.8) + 22r^2 + 120r + 122$ , respectively. So the axis of symmetry is

$$\frac{-2q^4 - 2q^3 - 2q^2r + o(q^3)}{-2(q^2 + q - 4r - 8)} = q^2 + o(q^2).$$

Since  $k(q+2) \leq s \leq q^2 + 2q$ , we have

$$\begin{aligned} F_{\max} &= F(q, r, k(q+2)) = q^4(-0.2r - 0.2) + q^3(6.6r^2 + 50.6r + 104) \\ &\quad + q^2(-18r^3 - 180.76r^2 - 633.32r - 729.56) + q(-51.6r^3 - 487r^2 - 1567.2r - 1630.8) \\ &\quad - 9r^4 - 76r^3 - 243r^2 - 656r - 1044 = q^4(-0.2r - 0.2) + 6.6q^3r^2 - 18q^2r^3 + o(q^4(r+1)) \\ &= r(-0.2q^4 + 6.6q^3r - 18q^2r^2) - 0.2q^4 + o(q^4(r+1)). \end{aligned}$$

In the range  $0 \leq r \leq 0.03q$ , we have  $-0.2q^4 + 6.6q^3r - 18q^2r^2 \leq -0.01q^4$ , implying that  $F_{\max} < 0$ .  $\square$

## B. Justification of the inequality (15)

Recall that here we have  $n = q^2 + q + 1 + r$ ,  $k = 0.55q$  and  $m_0 = \frac{1}{2}n(q+1)$ . Moreover,  $1 \leq r \leq 0.3q$  and  $k(q+2) \leq s \leq q^2 + 2q$ . The function  $G(q, r, s)$  can be rewrote as the following

$$\begin{aligned} G(q, r, s) &= (n - s + k - 1)(n - s + k - 2)(n - k) \\ &\quad - (2m_0 - s + (k - 1)0.7q - nk + k)(2m_0 - s + (k - 1)0.7q - nk - n + 2k). \end{aligned}$$

Similar to Appendix A, it suffices to show that the maximum value  $G_{\max}$  of  $G(q, r, s)$  is less than zero, subject to the restrictions that  $1 \leq r \leq 0.3q$  and  $k(q+2) \leq s \leq q^2 + 2q$ .

For fixed  $q$  and  $r$ ,  $G$  is a quadratic function with the variable  $s$ . The coefficients of  $s^2$  and  $s$  are  $q^2 + 0.45q + r$  and  $-2q^4 - 3.1q^3 - (4r - 0.275)q^2 - (3.1r + 0.5)q - 2r^2 + 2$ , respectively. So the axis of symmetry is

$$\frac{2q^4 + 3.1q^3 + 4q^2r + o(q^3)}{2(q^2 + 0.45q + r)} = q^2 + o(q^2).$$

Since  $k(q+2) \leq s \leq q^2 + 2q$ , we have

$$\begin{aligned} G_{\max} &= G(q, r, k(q+2)) = -0.210375q^5 + (0.6975r - 0.206475)q^4 + (0.5535r - 0.342125)q^3 \\ &\quad + (1.6975r^2 - 0.205r - 0.685)q^2 + (0.9r^2 - 0.2r - 0.2)q + r^3 - r \leq -0.21q^5 + 0.698rq^4 + o(q^5). \end{aligned}$$

As  $1 \leq r \leq 0.3q$ , we get  $-0.21q^5 + 0.698rq^4 \leq -0.0006q^5$ . Therefore, we have  $G_{\max} < 0$  and thus complete the proof.  $\square$