Some extremal results on 4-cycles

Jialin He Jie Ma Tianchi Yang

Abstract

We present two extremal results on 4-cycles. Let q be a large even integer. First we prove that every (q^2+q+1)-vertex C_4-free graph with more than $\frac{1}{2}q(q+1)^2 - 0.2q$ edges must be a spanning subgraph of a unique polarity graph. This implies a stability refinement of a special case of the seminal work of Füredi on the extremal number of C_4. Second we prove that every (q^2+q+1)-vertex graph with $\frac{1}{2}q(q+1)^2+1$ edges contains at least $q-1$ copies of C_4, where we also characterize the extremal graphs. This confirms infinitely many cases of a longstanding conjecture of Erdős and Simonovits on the number of C_4. The proof of the first result combines some earlier and novel ideas, while the proof of the second result builds on the first stability result.

1 Introduction

As one of the origins of extremal graph theory, Erdős [5] proposed the study of the maximum number $\mathrm{ex}(n, C_4)$ of edges in an n-vertex graph which does not contain any cycle of length four (such a graph is called *C_4-free*). An early result (see Kővári-Sós-Turán [14] and Reiman [16]) gives the general upper bound $\mathrm{ex}(n, C_4) \le \frac{n}{4}(1+\sqrt{4n-3})$. Using polarities from projective planes,[1] Brown [2] and Erdős-Rényi-Sós [6] independently and simultaneously proved that

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

In a striking breakthrough, Füredi [8, 10] confirmed a conjecture of Erdős by showing that this inequality holds as an equality, where he proved for $q = 2^k$ in [8] and for $q \ge 14$ in [10].

Theorem 1.1 (Füredi, [8, 10])**.** *If $q \notin \{1, 7, 9, 11, 13\}$, then $\mathrm{ex}(q^2+q+1, C_4) \le \frac{1}{2}q(q+1)^2$. Hence for all prime powers $q \ge 14$, $\mathrm{ex}(q^2+q+1, C_4) = \frac{1}{2}q(q+1)^2$.*

Füredi also proved that extremal graphs for $n = q^2+q+1$ where $q \ge q_0$ must be orthogonal polarity graphs (unpublished, see [11]). More recently, $\mathrm{ex}(q^2+q, C_4)$ was determined in [12] for all $q = 2^k$.

The first contribution of this paper is a stability result on 4-cycles.

Theorem 1.2. *Let $q \ge 10^9$ be an even integer and G be a C_4-free graph on q^2+q+1 vertices with $e(G) \ge \frac{1}{2}q(q+1)^2 - 0.2q + 1$. Then G is a subgraph of a unique polarity graph of order q.*

The proof of this result combines several arguments from Füredi's work as well as some novel ideas (to find 1-intersecting hypergraphs and polarities). As one application, this can be used to improve the upper bound of $\mathrm{ex}(n, C_4)$ in Theorem 1.1 for some specific integers n as follows: If q is a large even integer such that a projective plane of order q does not exist, then $\mathrm{ex}(q^2+q+1, C_4) \le \frac{1}{2}q(q+1)^2 - 0.2q$. We remark that by the celebrated Bruck-Ryser theorem [3] there are infinitely many such integers q.[2]

A longstanding conjecture of Erdős and Simonovits [7] asserts that any n-vertex graph with $\mathrm{ex}(n, C_4)+1$ edges contains at least $\big(1+o(1)\big)\sqrt{n}$ many C_4's. If true this will be sharp for infinitely many integers n (examples are given by orthogonal polarity graphs). The following is another result of this paper.

University of Science and Technology of China. Partially supported by NSFC grant 11622110, National Key Research and Development Project SQ2020YFA070080, and Anhui Initiative in Quantum Information Technologies grant AHY150200. Emails: hjxhjl@mail.ustc.edu.cn, jiema@ustc.edu.cn, ytc@mail.ustc.edu.cn.

[1] We will postpone the formal definitions in Section 2.

[2] The Bruck-Ryser theorem states that if $q \equiv 1$ or $2 \mod 4$ is an integer which cannot be expressed as a sum of two square numbers, then there exist no projective planes of order q.

Theorem 1.3. *Let $q \geq 10^{12}$ be even and let G be a graph on q^2+q+1 vertices with $\frac{1}{2}q(q+1)^2+1$ edges. Then either G contains at least $2q-3$ copies of C_4, or G is obtained from an orthogonal polarity graph of order q by adding a new edge. In the latter case, G contains $q-1, q$ or $q+1$ copies of C_4.*

One key ingredient in this proof is the use of the stability Theorem 1.2. Combining with Theorem 1.1, this gives the following exact result and thus confirms the conjecture of Erdős and Simonovits for infinitely many cases: For $q = 2^k$ where $k \geq 40$, any graph with q^2+q+1 vertices and $\mathrm{ex}(q^2+q+1, C_4)+1$ edges contains at least $q-1$ many C_4's. It also can be shown that the equality (exactly $q-1$) holds if and only if the graph is obtained from an orthogonal polarity graph of order q by adding a new edge between any two vertices of degree q.

The organization of this paper is as follows. Section 2 consists of preliminaries, where we give notations and collect some lemmas. In Sections 3 and 4, we prove Theorems 1.2 and 1.3, respectively.

2 Preliminaries

Let $\mathcal{H}$ be a hypergraph. The *degree* $d_{\mathcal{H}}(x)$ of a vertex x denotes the number of edges of $\mathcal{H}$ containing x. We say $\mathcal{H}$ is *k-regular* if all vertices have degree k and *k-uniform* if all edges have k vertices. We say $\mathcal{H}$ is *1-intersecting* if any two edges of $\mathcal{H}$ have exactly one common vertex. The *incidence matrix* of $\mathcal{H}$ is an $|E(\mathcal{H})| \times |V(\mathcal{H})|$ matrix $\mathcal{M}$, where $\mathcal{M}(e,x) = 1$ if $x \in e \in E(\mathcal{H})$ and 0 otherwise.

A *finite projective plane of order q*, is a $(q+1)$-uniform $(q+1)$-regular 1-intersecting hypergraph $\mathcal{H} = (P, \mathcal{L})$ with $|P| = q^2+q+1$, where P consists of *points* and $\mathcal{L}$ consists of *lines*. It follows that $|\mathcal{L}| = q^2+q+1$ and any two points are contained in a unique line. We need the following results.

Theorem 2.1 ([15])**.** *Let $\mathcal{H}$ be a 1-intersecting $(q+1)$-graph with q^2+q+1 vertices and more than $q^2 - \frac{\sqrt{5}-1}{2}q + 17\sqrt{q/5}$ edges for $q \geq 3900$. Then $\mathcal{H}$ can be embedded into a projective plane of order q.*

Theorem 2.2 ([4])**.** *Let $\mathcal{H}$ be a 1-intersecting $(q+1)$-graph with q^2+q+1 vertices and more than q^2-q+1 edges. If $\mathcal{H}$ can be embedded into a projective plane of order q, then this projective plane and the embedding both are unique.*

A *polarity* π of a projective plane $\mathcal{H} = (P, \mathcal{L})$ of order q is a bijection $\pi : P \cup \mathcal{L} \to P \cup \mathcal{L}$ such that $\pi(P) = \mathcal{L}$, $\pi(\mathcal{L}) = P$, π^2 is the identity function, and for any $x \in L$ where $L \in \mathcal{L}$, one has $\pi(L) \in \pi(x)$. The *polarity graph* $G(\pi)$ (of order q) is a simple graph on the vertex set P such that $xy \in E(G(\pi))$ if and only if $x \in \pi(y)$. Let $a(\pi)$ denote the number of *absolute points*, i.e., points $x \in P$ satisfying $x \in \pi(x)$. Baer [1] proved that $a(\pi) \geq q+1$ and it is also known (see [9]) that $a(\pi) = q+1+m_\pi\sqrt{q}$ for some integer $m_\pi \geq 0$. A polarity π and its polarity graph $G(\pi)$ are called *orthogonal* if $a(\pi) = q+1$. For any prime power q, orthogonal polarity graphs of order q exist. The following lemmas will be frequently used in the forthcoming proofs.

Lemma 2.3. *Let π be a polarity of a projective plane of order q. Then the polarity graph $G(\pi)$ is a C_4-free graph on q^2+q+1 vertices with exactly $\frac{1}{2}q(q+1)^2 - \frac{m_\pi}{2}\sqrt{q}$ edges such that $a(\pi)$ vertices have degree q and all others have degree $q+1$.*

Lemma 2.4. *Let G be a polarity graph of order q with $uv \notin E(G)$. Then $G \cup \{uv\}$ contains $q-1, q$ or $q+1$ four-cycles, any two of which share uv as the unique common edge. Moreover, $G \cup \{uv\}$ contains $q-1$ four-cycles if and only if both u, v have degree q in G.*

Proof. Let π be the corresponding polarity of G. We see $q \leq d_G(u), d_G(v) \leq q+1$. For any $u_i \in \pi(u)$, there exists a unique vertex v_i in $\pi(u_i) \cap \pi(v)$. So there are exactly $q+1$ sequences uu_iv_iv satisfying $u_i \in \pi(u)$ and $v_i \in \pi(u_i) \cap \pi(v)$, where $u_i \neq v$ and $v_i \neq u$ for all $i \in [q+1]$.[3] Consider $d_G(u) = d_G(v) = q$. Excluding uuv_iv and uu_ivv, there are exactly $q-1$ sequences uu_iv_iv with $\{u,v\} \cap \{u_i, v_i\} = \emptyset$. We also have $u_i \neq v_i$ (as otherwise, $\{u, u_i\} \subseteq \pi(u) \cap \pi(u_i)$, a contradiction). This shows exactly $q-1$

[3]Throughout this paper, for any positive integer k, we write $[k]$ as the set $\{1, 2, ..., k\}$.

four-cycles in $G \cup \{uv\}$. Now we may assume $N_G(u) = \{u_1, ..., u_{q+1}\}$. We claim that there is at most one i satisfying $v_i \in \{u_i, v\}$. To see this, note that if $v_i \in \{u_i, v\}$, then $u_i \in N_G(u) \cap N_G(v)$; thus two such i, j would force a four-cycle uu_ivu_ju in G, a contradiction. From this claim, we see $G \cup \{uv\}$ has q or $q+1$ four-cycles. To complete the proof, it suffices to show that all such paths uu_iv_iv are edge-disjoint in G. This follows by the fact that each middle edge u_iv_i can only appear once; as otherwise there are two paths uu_iv_iv and uv_iu_iv which would force a four-cycle uu_ivv_iu in G. □

Let G be a graph and $x, y \in V(G)$. We use $N_G[x]$ to denote the union of $\{x\}$ and the neighborhood $N_G(x)$ of x. For $A \subseteq V(G)$, let $N_G(A)$ be the set of vertices $x \notin A$ adjacent to some vertex in A. Let $d_G(x, y) = |N_G(x) \cap N_G(y)|$. We say $\{x, y\}$ is *uncovered* if $d_G(x, y) = 0$ and *covered* otherwise. We often drop the subscripts for all above notations when they are clear from context. Let UP be the set of uncovered pairs of G and let P_2 be the set of all paths of length 2 in G. Let $UP \cap A$ be the set of uncovered pairs $\{x, y\} \subseteq A$ of G and let $P_2 \cap A$ be the set of paths of P_2 with both endpoints in A. By $\#C_4$, we denote the number of copies of C_4 in G.

Proposition 2.5. *Let G be a graph with $A \subseteq V(G)$. Then $2\#C_4 \geq |P_2 \cap A| + |UP \cap A| - \binom{|A|}{2}$.*

Proof. Let B be the set of covered pairs $\{u, v\} \subseteq A$. Then $|B| = \binom{|A|}{2} - |UP \cap A|$. It holds that $2\#C_4 \geq \sum_B \binom{d(u,v)}{2} \geq \sum_B \big(d(u, v) - 1\big)$, which equals $|P_2 \cap A| + |UP \cap A| - \binom{|A|}{2}$. □

The next easy-to-use lemma is often adopted in replace of standard Cauchy-Schwarz inequalities.

Lemma 2.6. *Let $a_1, ..., a_m$ be nonnegative integers satisfying $\sum_{i=1}^m a_i \geq km + r$, where m, k, r are integers with $m, k > 0$ and $r \geq -m$. Then $\sum_{i=1}^m \binom{a_i}{2} \geq m\binom{k}{2} + rk$.*

Proof. Write $\sum_{i=1}^m a_i = tm + x$ for integers t, x with $0 \leq x < m$. We have $\sum_{i=1}^m \binom{a_i}{2} \geq x\binom{t+1}{2} + (m - x)\binom{t}{2} = m\binom{t}{2} + xt$ (see (3.1) from [9]). By letting $\Lambda := \sum_{i=1}^m \binom{a_i}{2} - m\binom{k}{2} - rk$, we get that

$$\Lambda \geq m\binom{t}{2} + xt - m\binom{k}{2} - rk = \frac{1}{2}m(t-k)(t+k-1) + xt - rk. \tag{1}$$

Since $(t+1)m > tm + x \geq km + r \geq (k-1)m$, we obtain $t - k \geq -1$. Now consider the following three cases. If $t - k = -1$, then $-r \geq m - x > 0$ and by (1), $\Lambda \geq -(m-x)(k-1) + (-r)k \geq m - x > 0$. If $t = k$, then $x \geq r$ and by (1) it is easy to see $\Lambda \geq k(x - r) \geq 0$. Lastly, we consider $t \geq k + 1$. Since $t + k - 1 \geq 2k$ and $m(t - k) \geq r - x$, by (1) again, we see $\Lambda \geq (r - x)k + xt - rk = x(t - k) \geq 0$. □

3 Proof of Theorem 1.2

We first reduce Theorem 1.2 to the following restricted version (with maximum degree $\Delta(G) \leq q+1$). Given a graph G, let $S_i = \{v \in V(G) : d_G(v) = i\}$.

Lemma 3.1. *Let $q \geq 10^9$ be even and G be a C_4-free graph on $q^2 + q + 1$ vertices with $\Delta(G) \leq q + 1$ and $e(G) \geq \frac{1}{2}q(q+1)^2 - 0.2q$. Then G is a subgraph of a unique polarity graph of order q.*

Proof of Theorem 1.2 (Assuming Lemma 3.1). Let q and G be from Theorem 1.2. Let $V(G) = \{v_1, ..., v_n\}$ where $n = q^2 + q + 1$ and $\Delta := \Delta(G) = d(v_1)$.

Using arguments in [8], we first show $\Delta \leq q + 2$ and $|S_{q+2}| \leq 1$. Since G is C_4-free, we see $|N(v_i) \setminus N(v_1)| = d(v_i) - d(v_i, v_1) \geq d(v_i) - 1$ for $2 \leq i \leq n$. As $e(G) \geq \frac{1}{2}q(q+1)^2 - 0.2q + 1$, we have $\sum_{i=2}^n |N(v_i) \setminus N(v_1)| \geq 2e(G) - \Delta - (n-1) \geq q^3 + q^2 - 0.4q - \Delta + 2$. Let $A = V \backslash N(v_1)$. By Proposition 2.5 and convexity, we obtain that

$$\binom{q^2+q+1-\Delta}{2} = \binom{|A|}{2} \geq |P_2 \cap A| \geq \sum_{i=2}^n \binom{|N(v_i) \setminus N(v_1)|}{2} \geq (q^2+q)\binom{\frac{q^3+q^2-0.4q-\Delta+2}{q^2+q}}{2}. \tag{2}$$

By the calculations in Appendix A, we see this inequality does not hold for $q+3 \le \Delta \le q^2+q$ when $q \ge 10^9$, a contradiction. So $\Delta(G) \le q+2$ holds.

To show $|S_{q+2}| \le 1$, suppose for the contrary that $d(v_1) = d(v_2) = \Delta = q+2$. Let $B = N(v_1) \cup N(v_2)$. If $N(v_1) \cap N(v_2) = \emptyset$, then $|B| = 2\Delta$. Since G is C_4-free, we have $|N(v_i) \setminus B| = d(v_i) - d(v_i, v_1) - d(v_i, v_2) \ge d(v_i) - 2$ for $2 < i \le n$. Similarly as above, we can derive that

$$\binom{n-2\Delta}{2} \ge \sum_{i=3}^{n} \binom{|N(v_i) \setminus B|}{2} \ge (n-2)\binom{\frac{\sum_{i=3}^{n}(d(v_i)-2)}{n-2}}{2} = (n-2)\binom{\frac{2e(G)-2\Delta-2n+4}{n-2}}{2}. \tag{3}$$

This is a contradiction (see its justification in Appendix A). Hence, we may assume $N(v_1) \cap N(v_2) = \{v_3\}$ and $|B| = 2\Delta - 1$. Let $C = N(v_3) \setminus \{v_1, v_2\}$. Then we have $|N(v_i) \setminus B| \ge d(v_i) - 1$ for $v_i \in C$ and $|N(v_i) \setminus B| \ge d(v_i) - 2$ for $v_i \notin N(v_3)$. By Proposition 2.5, we have

$$\begin{aligned}\binom{n-2\Delta+1}{2} &\ge \sum_{i=3}^{n} \binom{|N(v_i) \setminus B|}{2} \ge \binom{d(v_3)-2}{2} + \sum_{v_i \in C} \binom{d(v_i)-1}{2} + \sum_{v_j \notin N[v_3]} \binom{d(v_j)-2}{2} \\ &\ge \binom{d(v_3)-2}{2} + (n-3)\binom{\frac{2e(G)-2\Delta-2n+4}{n-3}}{2} \ge (n-3)\binom{\frac{2e(G)-2\Delta-2n+4}{n-3}}{2},\end{aligned} \tag{4}$$

again a contradiction (see its justification in Appendix A). Thus we have $|S_{q+2}| \le 1$.

Now we can delete at most one edge from G to get a subgraph G' with $\Delta(G') \le q+1$ and $e(G') \ge e(G) - 1 \ge \frac{1}{2}q(q+1)^2 - 0.2q$. By Lemma 3.1, there exists a unique polarity graph H of order q containing G' as a subgraph. Let e be the possible edge in $E(G) \backslash E(G')$. If e does not exist or $e \in E(H)$, then the conclusion holds for G. So $e \notin E(H)$. By Lemma 2.4, $H \cup \{e\}$ contains at least $q-1$ copies of C_4, all of which contain e and are edge-disjoint otherwise. Since $G = G' \cup \{e\}$ is C_4-free, any of these copies of C_4 has an edge not contained in G, all of which are distinct. By Lemma 2.3, this shows that $e(G') \le e(H) - (q-1) \le \frac{1}{2}q(q+1)^2 - (q-1)$, a contradiction to $e(G') \ge \frac{1}{2}q(q+1)^2 - 0.2q$. Assuming Lemma 3.1, we have completed the proof of Theorem 1.2. ■

In the rest of the section, we prove Lemma 3.1.

Proof of Lemma 3.1. Let q and G be from Lemma 3.1. If $\Delta(G) \le q$, then $e(G) \le \frac{q}{2}(q^2+q+1) < \frac{1}{2}q(q+1)^2 - 0.2q$, a contradiction. So we may assume $\Delta(G) = q+1$ in the rest of the proof.

The *deficiency* of a vertex v is defined by $f(v) = q+1-d(v)$, and the deficiency of a subset $T \subseteq V(G)$ is $f(T) = \sum_{v \in T} f(v)$. We will write $V = V(G)$ for short. Let $S = \cup_{i=0}^{q} S_i$. Let $B = \{x \in V : |N(x) \cap S| \ge 0.1q\}$ and $A = S_{q+1} \backslash B$. Finally, let $\mathcal{R} = \{N(x) : x \in A\}$.

Claim 3.1. $|B| \le 14$ *and* $|\mathcal{R}| \ge q^2 - 0.4q - 14$.

Proof. It is known (see Corollary 5.2 in [9]) that for even q, if G is C_4-free with q^2+q+1 vertices and $\Delta(G) = q+1$, then any $x \in S_{q+1}$ has a neighbor in S and $|S| \ge q+1$. Thus, we have

$$q+1 \le |S| \le \sum_{i=0}^{q} (i+1)|S_{q-i}| = f(V) = (q+1)n - 2e(G) \le 1.4q+1. \tag{5}$$

So $q^2 - 0.4q \le |S_{q+1}| \le q^2$ and $|S| \le 1.4q+1 \le 2q$. For each $T \subseteq S$, we have

$$1.4q+1 \ge f(V) \ge f(T) + (|S| - |T|) \ge f(T) + (q+1-|T|).$$

This implies that $f(T) \le |T| + 0.4q$ for any $T \subseteq S$ and in particular,

$$d(x) \ge 0.6q \quad \text{and} \quad d(x) + d(y) \ge 1.6q \quad \text{for any vertices } x, y. \tag{6}$$

Let t be the number of ordered adjacent pairs (b, v) with $b \in B$ and $v \in S$. Since $|B| \cdot 0.1q \le t \le |S|q \le 2q^2$, we see $|B| \le 20q$. Consider the subgraph G_0 of G induced by the set $B \cup S$, where $|B \cup S| \le$

$22q$. By the classic Reiman's bound, we derive that $\frac{1}{2}|B|\cdot 0.1q \le e(G_0) \le \frac{22q}{4}\cdot 10q^{1/2} = 55q^{3/2}$ and thus $|B| \le 1100\sqrt{q}$. For any $b, b' \in B$, we have $|N_S(b)\cap N_S(b')| \le 1$. Using (5) and inclusion-exclusion principle, $1.4q+1 \ge |S| \ge |\cup_{b\in B} N_S(b)| \ge \sum_{b\in B}|N_S(b)| - \sum_{b,b'\in B}|N_S(b)\cap N_S(b')| \ge |B|\cdot 0.1q - \binom{|B|}{2}$. Solving this quadratic inequality on $|B|$ and using the fact that $|B| \le 1100\sqrt{q}$ for large q, we can infer that $|B| \le 14$. So $|\mathcal{R}| = |A| = |S_{q+1}| - |B| \ge q^2 - 0.4q - 14$, proving this claim. □

Claim 3.2. *$\mathcal{R}$ is a 1-intersecting $(q+1)$-graph.*

Proof. We say a vertex v has **property 1**, if $v \in S_{q+1}$, $|N(v)\cap S_{q+1}| = q$ and $|N(v)\cap S_q| = 1$. Let V_1 be the set of all vertices of property 1. We assert that $|V_1| \ge 0.6q^2 - 1.8q$ which we prove now. For each $uv \in E(G)$ with $u \in S$ and $v \in S_{q+1}$, assign its weight to be $f(u)$. Let W denote the sum of weights of these edges. We note that any vertex in V_1 contributes one to W, while any vertex in $S_{q+1}\backslash V_1$ contributes at least two. By (5), $|V_1| + 2(|S_{q+1}| - |V_1|) \le W \le \sum_{i=0}^{q}(q-i)(i+1)|S_{q-i}| \le q\cdot f(V) \le q(1.4q+1)$. So $|V_1| \ge 2|S_{q+1}| - q(1.4q+1) \ge 0.6q^2 - 1.8q$.

The following property will be key to show $\mathcal{R}$ is 1-intersecting. Suppose $v \in V_1$ has $N(v) = \{v_1, ..., v_{q+1}\}$ and let $N_i = N(v_i)\backslash N[v]$ for $i \in [q+1]$. Consider a vertex $u \in S_{q+1}\setminus N[v]$ and suppose that u is adjacent to some vertex in $S_{q+1}\cap N(v)$, say v_1. Since $v \in V_1$, the subgraph $G[N(v)]$ consists of a matching of size $\frac{q}{2}$ plus an isolated vertex of degree q (see Proposition 5.4 in [9]). Without loss of generality, assume $v_1v_2 \in E(G)$ so $v_1, v_2 \in S_{q+1}$. Then u has exactly one neighbor in $N(v)$, namely v_1, and no neighbors in N_2 (else we get a C_4). The sets $N_1, ..., N_{q+1}$ partition $V\backslash N[v]$ and u cannot have two neighbors in some N_i. However, $d(u) = q+1$ and so this forces u to have exactly one neighbor in each N_i with $i \ne 2$, and $N(u)\cap N(v_2) = \{v_1\}$. We conclude that

$$|N(u)\cap N(v_i)| = 1 \text{ for all } i \in [q+1]. \tag{7}$$

We now summarize this as **property** $(\star)$: For any $v \in V_1$ with $N(v) = \{v_1, ..., v_{q+1}\}$, if $u \in S_{q+1}\setminus N[v]$ is adjacent to some vertex in $S_{q+1}\cap N(v)$, then (7) holds for u.

We then show that the neighborhood of any $x \in A$ contains many vertices of property 1. For $x \in A$, let $S_x = N(x)\cap S$ and $S_x^* = S_x \cup (N(S_x)\cap N(x))$. We will prove that $|(N(x)\backslash S_x^*)\cap V_1| \ge 0.3q+1$. Now since $x \in A$, we have $|S_x| \le 0.1q$ by definition of A and B. Every vertex in S_x has at most one neighbor in $N(x)$, so $|S_x^*\backslash S_x| \le |S_x|$ and thus $|S_x^*| \le 2|S_x|$. Let $N(x) = \{x_1, ..., x_{q+1}\}$ and $N_i = N(x_i)\setminus N[x]$ for $i \in [q+1]$. We first assert that $f(N_i) \ge 1$ for any $x_i \in N(x)\backslash S_x^*$. Indeed, by definition of S_x^*, such $x_i \in S_{q+1}$ and if y is a neighbor of x_i in S, then y must lie outside of $N[x]$ (that is in N_i). On the other hand, as pointed out in the beginning of the proof of Claim 3.1, such x_i must have a neighbor in S and thus in N_i. This shows that $f(N_i) \ge 1$. From this argument, we also see that $x_i \in N(x)\backslash S_x^*$ has $f(N_i\cup\{x_i\}) = 1$ if and only if $x_i \in V_1$. That says, $f(N_i\cup\{x_i\}) \ge 2$ if $x_i \in (N(x)\backslash S_x^*)\backslash V_1$, and $f(N_i\cup\{x_i\}) \ge 1$ if $x_i \in S_x$ or $x_i \in (N(x)\backslash S_x^*)\cap V_1$. Let $m = |(N(x)\backslash S_x^*)\cap V_1|$. Then we can get

$$m + 2(|N(x)| - |S_x^*| - m) + |S_x| \le \sum_{i\in[q+1]} f(N_i\cup\{x_i\}) \le f(V) \le 1.4q+1.$$

Using $|N(x)| = d(x) = q+1$ and $2|S_x^*| - |S_x| \le 3|S_x| \le 0.3q$, we can derive that $m \ge 0.3q+1$.

Now we are ready to prove that $\mathcal{R}$ is 1-intersecting. Suppose for the contrary that there exist some $x, y \in A$ with no common neighbor. If $xy \in E(G)$, by the previous paragraph, there exists $z \in N(x)\cap V_1 - \{y\}$ with $yz \notin E(G)$. Apply property $(\star)$ by viewing z as the vertex v therein. Since $y \in S_{q+1}\backslash N[z]$ is adjacent to $x \in S_{q+1}\cap N(z)$, it shows $|N(y)\cap N(x)| = 1$, a contradiction.

Hence $xy \notin E(G)$. Let $N(x) = \{x_1, ..., x_{q+1}\}$. Let $N_i = N(x_i)\backslash N[x]$ for $i \in [q+1]$ and $Y = V\backslash(N[x]\cup N_1\cup ...\cup N_{q+1})$. So $y \in Y$. Since each x_i has at most one neighbor in $N(x)$, $|Y| \le n - (q+2) - \sum_{i=1}^{q+1}(d(x_i)-2) = \sum_{i=1}^{q+1} f(x_i)$. Let $N_1(x)$ be the set of vertices in $N(x)\backslash S_x^*$ of property 1 and $N_2(x) = N(x)\backslash(N_1(x)\cup S_x^*)$. Note that $|N_1(x)| \ge 0.3q+1$, $f(N_i) = 1$ for each $x_i \in N_1(x)$, and $f(N_j) \ge 2$ for each $x_j \in N_2(x)$. So $|Y| \le \sum_{i=1}^{q+1} f(x_i) = \sum_{x_i\in S_x} f(x_i) \le 1.4q+1-|N_1(x)| - 2|N_2(x)|$. Since $N(x) = N_1(x)\cup N_2(x)\cup S_x^*$, the number of neighbors of y in those N_i's satisfying $x_i \in N_1(x)$ is at least $d(y)-(|Y|-1)-|S_x^*|-|N_2(x)|$, which is at least $(q+2)-(1.4q+1)+|N_1(x)|+|N_2(x)|-|S_x^*| \ge 0.2q$.

Here we use the above estimation on $|Y|$ and the facts that $|N_1(x)| + |N_2(x)| + |S_x^*| = q+1$ and $|S_x^*| \le 2|S_x| \le 0.2q$. Since $|N(y) \cap S| < 0.1q$, among those $0.2q$ neighbors of y, there is a vertex $z \in N(y) \cap S_{q+1}$. Let $z \in N_j$ for some $x_j \in N_1(x) \subseteq V_1$. Apply property $(\star)$ by viewing this x_j as the vertex v. Since $y \in S_{q+1} \backslash N[x_j]$ is adjacent to $z \in N(x_j) \cap S_{q+1}$, we can derive that y and $x \in N(x_j)$ have a common neighbor, a contradiction to the assumption. This proves the claim. $\square$

By the above claim and Theorems 2.1 and 2.2, $\mathcal{R}$ can be embedded into a projective plane $\mathcal{P}$ of order q and this embedding is unique. Using $\mathcal{R}$ and $\mathcal{P}$, we now construct a unique polarity graph of order q, which contains G as a subgraph. Let $\mathcal{R}^c = \mathcal{P} \backslash \mathcal{R}$ with $|\mathcal{R}^c| \le 1.4q + 15$.

We say $v \in V(G)$ is **feasible**, if $N(v) \subseteq L$ for some line $L \in \mathcal{P}$; otherwise v is *non-feasible*. For non-feasible v, we say it is **near-feasible**, if there exist a subset $K_v \subseteq N(v)$ and a line $L \in \mathcal{R}^c$ such that $|K_v| \le 2$ and $N(v) \backslash K_v \subseteq L$. In both definitions, we say v and L are *associated* with each other. For feasible v, we also let $K_v = \emptyset$. By (6), for any vertices u and v which is feasible or near-feasible,

$$|(N(u)\backslash K_u) \cup (N(v)\backslash K_v)| \ge \big(d(u) - 2\big) + \big(d(v) - 2\big) - 1 \ge 1.6q - 5 > q + 1. \tag{8}$$

This implies that each line in $\mathcal{P}$ is associated with at most one feasible or near-feasible vertex. On the other hand, if there are two lines in $\mathcal{P}$ associated with the same feasible or near-feasible vertex v, as $d(v) \ge 0.6q$ from (6), it is easy to see that these two lines will intersect with more than two vertices, a contradiction. So each feasible or near-feasible vertex is associated with a unique line in $\mathcal{P}$.

We show that any $v \in V(G)$ has a neighbor u with $d_{\mathcal{R}}(u) = |N(u) \cap A| \ge q - 2$. Let $N(v) = \{v_1, ..., v_d\}$, where $d = d(v) \ge 0.6q$. Let $N_i = N(v_i) \backslash N[v]$ for $i \in [d]$. Since the sets $N_i \cup \{v_i\}$ are disjoint over $i \in [d]$, we have $1.4q + 1 \ge f(V) \ge \sum_{i \in [d]} f(N_i \cup \{v_i\}) + f(v) = \sum_{i \in [d]} f(N_i \cup \{v_i\}) + (q + 1 - d)$. This implies at least $(d - 0.4q)/2 \ge 15$ distinct $j \in [d]$ with $f(N_j \cup \{v_j\}) \le 1$. As $|B| \le 14$, there is some $j \in [d]$ with $f(N_j \cup \{v_j\}) \le 1$ and $N_j \cap B = \emptyset$. Therefore, $d_{\mathcal{R}}(v_j) = |N(v_j) \cap A| \ge |N_j \cap A| \ge |N_j| - |N_j \cap S| \ge \big(d(v_j) - 2\big) - f(N_j) = \big(q - 1 - f(v_j)\big) - f(N_j) \ge q - 2$, as desired.

Now we consider some properties on non-feasible vertices v. Clearly $v \notin A$. Recall that any two points in a projective plane are contained in a unique line. So for any $x, y \in N(v)$, the pair $\{x, y\}$ is contained in a unique line in the projective plane $\mathcal{P}$ but not in $\mathcal{R}$ (suppose otherwise that $\{x, y\} \subseteq N(a) \in \mathcal{R}$, then $xayvx$ would form a C_4 in G). Let $\mathcal{L}_v$ be the family of all lines $L \in \mathcal{P}$ which contains at least two vertices of $N(v)$. Then $\mathcal{L}_v \subseteq \mathcal{R}^c$ and we also point out that any vertex in $N(v)$ appears in at least two lines of $\mathcal{L}_v$ (suppose for a contradiction that say $x \in N(v)$ appears only in the unique line $L \in \mathcal{L}_v$, then for any $y \in N(v) \backslash \{x\}$ the pair $\{x, y\}$ must be contained in L, which implies that $N(v) \subseteq L \in \mathcal{P}$, contradicting that v is non-feasible).

Next we show all non-feasible vertices are in fact near-feasible. Let v be any non-feasible vertex. Note that there exists $u \in N(v)$ with $d_{\mathcal{R}}(u) \ge q - 2$. Then there are $\alpha \in \{2, 3\}$ lines in $\mathcal{L}_v$ containing u, say L_i for $i \in [\alpha]$. Let $D_i = L_i \cap N(v)$ for $i \in [\alpha]$ such that $|D_i| \ge |D_j| \ge 2$ whenever $i < j$. We see that for any $x \in N(v) \backslash \{u\}$, the pair $\{u, x\}$ is contained in a unique D_i. Therefore, $N(v) = \cup_{i \in [\alpha]} D_i$ and $D_i \cap D_j = \{u\}$ for any $i < j$. Consider $x \in D_i \backslash u$ and $y \in D_j \backslash u$. By the previous paragraph, any such pair $\{x, y\}$ is contained in a line in $\mathcal{R}^c$. Also any $L \in \mathcal{R}^c$ contains at most one such pair (as otherwise, $|L \cap L_k| \ge 2$ for some $k \in [\alpha]$). Hence, we can derive $1.4q + 15 \ge |\mathcal{R}^c| \ge (|D_i| - 1)(|D_j| - 1)$, where $\sum_{i \in [\alpha]} (|D_i| - 1) = d(v) - 1 \ge 0.6q - 1$. This would imply that $|D_i| - 1 \le 2$ for any $2 \le i \le \alpha$ and thus $|D_1| \ge d(v) - 4$. Let $K_v = \big(\cup_{2 \le i \le \alpha} D_i\big) \backslash u$. So $|K_v| \le 4$ and $N(v) \backslash K_v \subseteq L_1$. For any $w \in K_v$, let T_w be the set of all lines in $\mathcal{L}_v$ containing w. If there exists $L \in T_w$ with $|L \cap D_1| \ge 2$, then we see $|L \cap L_1| \ge 2$ and $L \ne L_1$, a contradiction. This shows $|T_w| \ge |D_1| \ge d(v) - 4 \ge 0.6q - 4$ for each $w \in K_v$. If $|K_v| \ge 3$, since $1.4q + 15 \ge |\mathcal{R}^c| \ge |\mathcal{L}_v| \ge |\cup_{w \in K_v} T_w| \ge \sum_{w \in K_v} |T_w| - \sum_{w, w' \in K_v} |T_w \cap T_{w'}|$, then this forces $|T_w \cap T_{w'}| \ge 2$ for some $w, w' \in K_v$, which gives two lines in $\mathcal{P}$ containing $\{w, w'\}$, a contradiction. Hence $|K_v| \le 2$ and v is near-feasible.

Let $K = \cup_{v \in V(G)} K_v$. We claim that either there is at most one near-feasible vertex or $|K| \le 1$. First we point out that $|K| \le 2$. Suppose there are three distinct vertices $w_i \in K_{v_i}$ for $i \in [3]$, where v_1, v_2, v_3 are near-feasible (not necessarily distinct). Let T_{w_i} be the set of all lines in $\mathcal{L}_{v_i}$ containing w_i. By the same arguments in the previous paragraph, we obtain $|T_{w_i}| \ge d(v_i) - 4 \ge 0.6q - 4$ and

$|T_{w_i} \cap T_{w_j}| \le 1$ for $i, j \in [3]$. Then we arrive at a contradiction that $1.4q + 15 \ge |\mathcal{R}^c| \ge |\cup_{i\in[3]} \mathcal{L}_{v_i}| \ge |\cup_{i\in[3]} T_{w_i}| \ge \sum_{i\in[3]} |T_{w_i}| - 3 \ge 1.8q - 15$, thereby proving $|K| \le 2$. Now it suffices to assume that $|K| = 2$ and there are two near-feasible vertices v_1, v_2. Let $K = \{w_1, w_2\}$ such that $w_i \in N(v_i)$ for $i \in [2]$. Similarly as before, if T_{w_i} denotes the set of all lines in $\mathcal{L}_{v_i}$ containing w_i, then we can show $|T_{w_i}| \ge d(v_i)-4$. By (6), $|T_{w_1}|+|T_{w_2}| \ge d(v_1)+d(v_2)-8 \ge 1.6q-8$. Since $|T_{w_1} \cup T_{w_2}| \le |\mathcal{R}^c| \le 1.4q+15$, this forces two lines in $\mathcal{P}$ containing $\{w_1, w_2\}$, again a contradiction. This proves the claim.

Let $V = \{v_1, ..., v_n\}$ for $n = q^2 + q + 1$. For each $i \in [n]$, let $L_i \in \mathcal{P}$ be the unique line associated with v_i. Let $\pi : V \leftrightarrow \mathcal{P}$ be a function which maps $v_i \leftrightarrow L_i$ for every $i \in [n]$. Let $\mathcal{M} = (m_{ij})$ be the incidence matrix of $\mathcal{P}$ with respect to π. We now show $\mathcal{M}$ is symmetric. To show this, we first assert that if $v_i \in A\backslash K$, then $m_{ij} = m_{ji}$ for all $j \in [n]$. For $m_{ij} = 1$, as $v_i \in A$, we have $v_j \in L_i = N(v_i) \in \mathcal{R}$; since $v_i \notin K$, we see $v_i \in N(v_j)\backslash K \subseteq N(v_j)\backslash K_{v_j} \subseteq L_j$, which shows that $m_{ji} = 1 = m_{ij}$. Now we observe that as $v_i \in A$, the i'th column and the i'th row of $\mathcal{M}$ have exactly $q + 1$ many 1-entries, and all these 1-entries are in the symmetric positions. This shows that the i'th column and the i'th row are symmetric, proving the assertion. Note that $|A\backslash K| \ge |A| - 2 > q^2 - q + 3$. Then by Lemma 3.7 in [9], one can conclude that the whole matrix $\mathcal{M}$ is symmetric.

Therefore the function π is a polarity of the projective plane $\mathcal{P}$. Let H be the polarity graph of π. Finally we show G is a subgraph of H. This is equivalent to show that the adjacent matrix $\mathcal{A} = (a_{ij})$ of G is at most $\mathcal{M} = (m_{ij})$; that is $a_{ij} \le m_{ij}$ for each i, j. Suppose that v_n is the only near-feasible vertex. Then we see $a_{ij} \le m_{ij}$ holds for any $i \ne n$. Since $\mathcal{A}, \mathcal{M}$ are symmetric and $a_{nn} = 0$, $\mathcal{A}$ is indeed at most $\mathcal{M}$. Now it suffices to consider the case $|K| = 1$, say $K = \{v_k\}$. We see $a_{ij} \le m_{ij}$ holds for all $j \ne k$. Similarly we can show that G is a subgraph of H. By the uniqueness of the projective plane $\mathcal{P}$, it also can be derived from the above arguments that the polarity graph H is unique. ■

We remark the same proof also works when $e(G) \ge \frac{1}{2}q(q + 1)^2 - (0.25 - o(1))q$ for large even q.

4 Proof of Theorem 1.3

Let q and G be from Theorem 1.3. We assume $\#C_4 \le 2q - 4$ and aim to show that G is obtained from an orthogonal polarity graph of order q by adding an edge.

Let $S_i = \{v \in V : d(v) = i\}$ and $S = \cup_{i=0}^{q} S_i$. For a vertex v, let $c(v)$ be the number of C_4 containing v and its *deficiency* $f(v) = \max\{q + 1 - d(v), 0\}$. The deficiency of $A \subseteq V(G)$ is $f(A) = \sum_{v\in A} f(v)$. A pair $\{u, v\} \subseteq V$ is *opposite* if $d(u, v) \ge 2$. Let $d_0(v)$ be the number of vertices u with $d(u, v) = 0$.

Claim 4.1. *Any vertex v satisfies $c(v) \ge (d(v) - q - 1)q - f(N(v)) + d_0(v)$, and if $d(v) \ge q + 1$ and $N(v) \cap S = \emptyset$ then $c(v) \ge 1$.*

Proof. To see this, by counting the paths of length 2 with an endpoint v, we get

$$\sum_{u\in V\backslash\{v\}} d(v, u) = \sum_{w\in N(v)} (d(w) - 1) \ge \sum_{w\in N(v)} (q - f(w)) = d(v)q - f(N(v)). \tag{9}$$

Thus, $c(v) = \sum_{u\in V\backslash\{v\}} \binom{d(v,u)}{2} \ge \sum_{u\in V\backslash\{v\}}(d(v, u) - 1) + d_0(v) \ge (d(v) - q - 1)q - f(N(v)) + d_0(v)$, where the first inequality holds because each vertex u with $d(u, v) \ge 1$ satisfies $\binom{d(v,u)}{2} \ge d(v, u) - 1$ and the contribution of all vertices u' with $d(u', v) = 0$ cancels out in the expression. For the second assertion, we have $f(N(v)) = 0$ and by the above inequality, we may assume $d(v) = q + 1$ and $d_0(v) = 0$. So every vertex in $V\backslash\{v\}$ has at least one common neighbor with v. Since $d(v) = q + 1$ is odd, $G[N(v)]$ can not consist of a perfect matching. This implies $c(v) \ge 1$. □

Claim 4.2. *We have $\Delta(G) \le q + 3$.*

Proof. Suppose for the contrary that there exists some $v_1 \in V(G)$ with $d(v_1) = q + k$ for some $4 \le k \le q^2$. Let $a_i = |N(v_1) \cap N(v_i)|$ for $2 \le i \le n$. There are $\binom{a_i}{2}$ copies of C_4 with the opposite pair $\{v_1, v_i\}$, so we have $2q - 4 \ge c(v_1) = \sum_{i=2}^{n} \binom{a_i}{2} \ge \sum_{i=2}^{n}(a_i - 1)$, implying that $\sum_{i=2}^{n} a_i \le$

q^2+3q-4. Also we have $\sum_{i=2}^n (d(v_i)-a_i) = 2e(G) - d(v_1) - \sum_{i=2}^n a_i = (q^2+q)(q-1)+X$, where $X = 2q^2+q+2-k-\sum_{i=2}^n a_i \geq -(q^2+q)$. By Proposition 2.5 and Lemma 2.6,

$$2\#C_4 + \binom{n-d(v_1)}{2} \geq |P_2 \cap (V \backslash N(v_1))| = \sum_{i=2}^n \binom{d(v_i)-a_i}{2} \geq (q^2+q)\binom{q-1}{2} + (q-1)X.$$

Further calculations give $2\#C_4 \geq q^3-2q^2+2q-(q-1)\sum_{i=2}^n a_i - 0.5k^2 + k(q^2-q+1.5)-2 \geq 5q-8$, where the last inequality holds as $k \geq 4$ and $\sum_{i=2}^n a_i \leq q^2+3q-4$, a contradiction to $\#C_4 \leq 2q-4$. $\square$

Now we have that

$$f(V) = (q+1)n + |S_{q+2}| + 2|S_{q+3}| - 2e(G) = q-1+|S_{q+2}|+2|S_{q+3}|. \tag{10}$$

Claim 4.3. $|S| \geq q-8$ *and* $|S_{q+2} \cup S_{q+3}| \leq 5$.

Proof. We first show that any v_1, v_2 in $S_{q+2} \cup S_{q+3}$ with $c(v_i) < 0.2q$ for $i \in [2]$ form an opposite pair. Otherwise, we have $d(v_1, v_2) \leq 1$. Let $d(v_1) = q+2+\delta_1$ and $d(v_2) = q+2+\delta_2$ for $\delta_1, \delta_2 \in \{0,1\}$. Let $B = N(v_1) \cup N(v_2)$ and $a_i = |N(v_i) \cap B|$ for all $3 \leq i \leq n$. Then at least $a_i - 2$ copies of C_4 contain $\{v_i, v_1\}$ or $\{v_i, v_2\}$ as their opposite pairs. Consider the case when v_1, v_2 have a common neighbor say v_3. In this case, for any $v_i \in N(v_3) \backslash \{v_1, v_2\}$, at least $a_i - 1$ copies of C_4 contain $\{v_i, v_1\}$ or $\{v_i, v_2\}$ as their opposite pairs. Thus $0.4q > c(v_1)+c(v_2) \geq \sum_{i=3}^n (a_i-2)+d(v_3)-2$, implying that $\sum_{i=3}^n a_i + d(v_3) \leq 2q^2+2.4q$. Hence we have $\sum_{i=4}^n (d(v_i)-a_i) = 2e(G)-d(v_1)-d(v_2)-d(v_3)-\sum_{i=4}^n a_i = (q^2+q-2)(q-1)+X$, where $X = 2q^2+2q-4-\delta_1-\delta_2-\sum_{i=4}^n a_i - d(v_3) \geq -0.4q-6 \geq -(q^2+q-2)$. By Proposition 2.5 and Lemma 2.6, if we write $A = V \backslash B$, then one can derive

$$\begin{aligned} 2 \cdot \#C_4 &\geq |P_2 \cap A| - \binom{|A|}{2} \geq \sum_{i=4}^n \binom{d(v_i)-a_i}{2} - \binom{q^2-q-2-\delta_1-\delta_2}{2} \\ &\geq (q^2+q-2)\binom{q-1}{2} - (0.4q+6)(q-1) - \binom{q^2-q-2}{2} = 0.1q^2-4.1q+1 > 4q-8, \end{aligned}$$

a contradiction. The case when v_1, v_2 have no common neighbor can be treated similarly.

If more than $\sqrt{8q}$ vertices $u \in S_{q+2} \cup S_{q+3}$ have $c(u) < 0.2q$, then from the previous paragraph we see at least $\binom{\sqrt{8q}+1}{2} \geq 4q$ opposite pairs, providing at least $2q$ copies of C_4. So we may assume at most $\sqrt{8q}$ vertices $u \in S_{q+2} \cup S_{q+3}$ with $c(u) < 0.2q$. Since $\sum_{v \in V} c(v) = 4\#C_4 < 8q$, there are at most $8q/0.2q = 40$ vertices $w \in V$ with $c(w) \geq 0.2q$. Hence $|S_{q+2} \cup S_{q+3}| \leq \sqrt{8q}+40 \leq 3\sqrt{q}$.

Observe $|S| \leq f(V) = q-1+|S_{q+2}|+2|S_{q+3}| \leq q+6\sqrt{q}-1$. Then $|S_{q+1}| = n-|S|-|S_{q+2} \cup S_{q+3}| > q^2-9\sqrt{q}$. If $|S| \leq q-9$, then there are at least $|S_{q+1}| - q|S| \geq 8q$ vertices u in S_{q+1} with no neighbors in S; by Claim 4.1, every such u has $c(u) \geq 1$, implying that $\#C_4 \geq 2q$, a contradiction. Therefore we have $q-8 \leq |S| \leq q+6\sqrt{q}-1$. Furthermore, for any $T \subseteq V$, it gives that

$$q+6\sqrt{q}-1 \geq f(V) \geq f(T) + (|S|-|S \cap T|) \geq f(T) - |S \cap T| + (q-8), \tag{11}$$

which implies that

$$f(T) \leq |S \cap T| + 6\sqrt{q} + 7.$$

Now suppose for the contrary that $S_{q+2} \cup S_{q+3}$ contains v_i for $i \in [6]$. For any $1 \leq i < j \leq 6$, $d(v_i, v_j) \leq 2\sqrt{q}$ (as otherwise $\#C_4 \geq 2q$). Fix $j \in [6]$. Using (9) together with the fact $d(v_j) \geq q+2$ and the above bound on $f(T)$ (with $T = N(v_j)$), we get that

$$\begin{aligned} \sum_{i=7}^n \binom{d(v_j, v_i)}{2} &\geq \sum_{i=7}^n (d(v_j, v_i) - 1) \geq \sum_{i \in [n] \backslash \{j\}} d(v_j, v_i) - \left(q^2+q+10\sqrt{q}\right) \\ &\geq d(v_j)q - f(N(v_j)) - \left(q^2+q+10\sqrt{q}\right) \geq q - 16\sqrt{q} - 7 - |N(v_j) \cap S|. \end{aligned}$$

Since $2\#C_4 \geq \sum_{j=1}^{6}\sum_{i=7}^{n}\binom{d(v_j,v_i)}{2}$, this gives $4q \geq 2\#C_4 \geq 6(q-16\sqrt{q}-7)-\sum_{j=1}^{6}|N(v_j)\cap S|$. So $\sum_{j=1}^{6}|N(v_j)\cap S| \geq 2q-96\sqrt{q}-42$. By inclusion-exclusion principle,

$$|S| \geq |\bigcup_{1\leq j\leq 6}(N(v_j)\cap S)| \geq \sum_{1\leq j\leq 6}|N(v_j)\cap S| - \sum_{1\leq i<j\leq 6}|N(v_i)\cap N(v_j)| \geq 2q-126\sqrt{q}-42,$$

where last inequality holds as $d(v_i,v_j) \leq 2\sqrt{q}$ for $1 \leq i < j \leq 6$. But $|S| \leq q+6\sqrt{q}-1$, where q is large. This final contradiction proves Claim 4.3. $\square$

By (10) and Claim 4.3, one can easily deduce that $q-8 \leq |S| \leq f(V) \leq q+9$. This together with (11) shows that for any $T \subseteq V(G)$, $f(T) \leq |S\cap T|+17$. In particular, for any vertex v, $f(v) \leq 18$ and $d(v) \geq q+1-f(v) \geq q-17$.

Claim 4.4. *For any vertex v, either $|N(v)\cap S| \leq 20$ or $|N(v)\cap S| \geq q-28$. Moreover, there is at most one vertex (say z if it exists) in G with $|N(z)\cap S| \geq q-28$.*

Proof. Consider any $v \in V$. Let $N(v)\backslash S = \{v_1,...,v_t\}$ with $t = d(v)-|N(v)\cap S|$. Let $N_i = N(v_i)\backslash\{v\}$ for $i \in [t]$. If N_i, N_j share a common vertex x, then vv_ixv_jv forms a distinct C_4. Since $\#C_4 < 2q$ and $|N_i| \geq d(v_i)-1 \geq q$ for $i \in [t]$, we can derive that $|\bigcup_{i\in[t]}N_i| \geq \sum_{i\in[t]}|N_i| - 2q \geq qt-2q$.

Let $B_i = N_i \cap S$ and C_i be the set of vertices $x \in N_i$ with $c(x) = 0$. If $x \in C_i \cap C_j$, then vv_ixv_jv forms a C_4, a contradiction. So C_i's are disjoint over $i \in [t]$. This shows $\sum_{i\in[t]}|B_i\cap C_i| = |\cup_{i\in[t]}(B_i\cap C_i)| \leq |\cup_{i\in[t]}B_i| \leq |S| \leq q+9$. We now show every $x \in C_i\backslash B_i$ has at least one neighbor in $S\backslash N(v)$. Since $d(x) \geq q+1$ and $c(x) = 0$, by Claim 4.1, we see x has a neighbor say $y \in S$; if $y \in N(v)\cap S$, then vv_ixyv is a C_4, a contradiction to $x \in C_i$. Also it is clear that every vertex in $S\backslash N(v)$ has at most one neighbor in $C_i\backslash B_i$. Hence $|C_i\backslash B_i| \leq |S\backslash N(v)| \leq q+9-|N(v)\cap S|$. Putting these together, we see $|\bigcup_{i\in[t]}C_i| = |\bigcup_{i\in[t]}(C_i\backslash B_i)| + |\bigcup_{i\in[t]}(B_i\cap C_i)| \leq t(q+9-|N(v)\cap S|)+q+9$.

Since each $x \in N_i\backslash C_i$ for $i \in [t]$ has $c(x) \geq 1$, we can derive the following

$$8q \geq \sum_{x\in V}c(x) \geq |\bigcup_{i\in[t]}(N_i\backslash C_i)| \geq |\bigcup_{i\in[t]}N_i| - |\bigcup_{i\in[t]}C_i| \geq (d(v)-|N(v)\cap S|)\cdot(|N(v)\cap S|-9)-3q-9,$$

where $d(v) \geq q-17$ and q is large. Solving this inequality gives that either $|N(v)\cap S| \leq 20$ or $|N(v)\cap S| \geq q-28$. If there exist z_1, z_2 with $|N(z_i)\cap S| \geq q-28$ for $i \in [2]$, then as $|S| \leq q+9$, we have $|N(z_1)\cap N(z_2)| \geq q-65$. This would give at least $\binom{q-65}{2} \gg 2q$ copies of C_4, a contradiction. Thus at most one vertex z can have $|N(z)\cap S| \geq q-28$. $\square$

Let $W = S_{q+2}\cup S_{q+3}\cup\{z\}$. By Claims 4.3 and 4.4, we see $|W| \leq 6$ and any vertex in $V\backslash\{z\}$ has at most 20 neighbors in S. Let $\ell = 900$ so that $0.01\ell q - \binom{\ell}{2}\cdot 2\sqrt{q} > 8q$ holds.

Claim 4.5. *If there are ℓ opposite pairs $\{u_i, v_i\}$ for $i \in [\ell]$ such that $u_i, v_i \in V\backslash W$ and all v_i are distinct, then there is some u_i with $c(u_i) > 0.8q$.*

Proof. Let $c'(v)$ denote the number of vertices $x \in N(v)$ with $c(x) \geq 1$. The key for this claim is to show: if $\{u, v\}$ is an opposite pair with $u \in V\backslash W$ and $v \in V\backslash\{z\}$, then $c(u)+19\cdot c'(v) \geq q-740$.

Note that $|N(u)\cap S| \leq 20$ and $|N(v)\cap S| \leq 20$. Let $d(u) = q+1-a$. Then $0 \leq a \leq 18$, $f(u) = a$ and $f(N(u)) \leq |N(u)\cap S|+17 \leq 37$. Let V_u be the set of vertices $x \in V$ with $d(x,u) = 0$. By Claim 4.1, $|V_u| = d_0(u) \leq c(u)+aq+f(N(u)) \leq c(u)+aq+37$. Let $N(v) = A\cup B$, where A consists of vertices x with $c(x) = 0$ and B consists of vertices y with $c(y) \geq 1$. Further we let $N_1 = N(v)\cap S$, $N_2 = N(v)\cap N[u]$, $N_3 = \{x \in N(v)\backslash(N_1\cup N_2) : |N(x)\cap V_u| \leq a\}$ and $N_4 = N(v)\backslash(N_1\cup N_2\cup N_3)$. By definition, we know $|N_1| \leq 20$, $N_2 \subseteq B$ (as $\{u, v\}$ is an opposite pair) and $N_3 \subseteq S_{q+1}\cup S_{q+2}\cup S_{q+3}$.

We claim that $N_3 \subseteq B$, i.e., each vertex $x \in N_3$ is contained in a 4-cycle. We first see that for any $x \in N_3$, any vertex in $N(x)\backslash V_u$ has at least one neighbor in $N(u)$, while $v \in N(x)\backslash V_u$ has at least two neighbors in $N(u)$. Thus using the definition of $x \in N_3$, there are at least $|N(x)\backslash V_u|+1 = d(x)-|N(x)\cap V_u|+1 \geq (q+1)-a+1$ edges zy with $z \in N(x)$ and $y \in N(u)$. Since $|N(u)| = q+1-a$, there exists some vertex $y \in N(u)$ with at least two neighbors in $N(x)$. As $x \notin N(u)$, we see $y \neq x$ and thus we can easily find a 4-cycle containing x. This shows $N_3 \subseteq B$.

Since $v \notin V_u$ and every two vertices in A have no common neighbor except v, we deduce that the sets $N(x) \cap V_u$ are disjoint over all $x \in N_4 \cap A$, where $|N(x) \cap V_u| \geq a+1$. So $|N_4 \cap A| \leq |V_u|/(a+1) \leq (c(u)+aq+37)/(a+1)$. Combining with the facts that $|N_1| \leq 20$ and $A \subseteq N_1 \cup (N_4 \cap A)$ (as $N_2 \cup N_3 \subseteq B$), we have

$$c'(v) = |B| = d(v) - |A| \geq (q-17) - |N_1| - |N_4 \cap A| \geq (q-37) - \frac{c(u)+aq+37}{a+1} = \frac{q-c(u)-37}{a+1} - 37.$$

As $0 \leq a \leq 18$, we have $c(u) + 19 \cdot c'(v) \geq c(u) + (a+1) \cdot c'(v) \geq q - 37(a+2) \geq q - 740$, as desired.

Suppose $c(u_i) \leq 0.8q$ for all $i \in [\ell]$. For each $\{u_i, v_i\}$, we have $c(u_i) + 19 \cdot c'(v_i) \geq q - 740$, which implies that $c'(v_i) \geq 0.01q$. We know that every two vertices v_i, v_j have at most $2\sqrt{q}$ common neighbors. Using inclusion-exclusion, the number of vertices in $\bigcup_{i \in [\ell]} N(v_i)$ which lie in a copy of C_4 is at least $0.01\ell q - \binom{\ell}{2} \cdot 2\sqrt{q} > 8q$, a contradiction to the assumption $\#C_4 < 2q$. $\square$

We now show that there exists $E^* \subseteq E(G)$ with $|E^*| \leq 105$ such that $G' = G - E^*$ has at most $0.1q$ copies of C_4. Let $A = \{v \in V : c(v) > 0.8q\}$, $B = A \cup W$ and $E^* = E(G[B])$. So $|B| < 8q/0.8q + 6 = 16$ and $|E^*| \leq 105$. We will show that such E^* is the desired edge set. Let $\mathcal{C}$ be the set of 4-cycles in $G' = G - E^*$ and assume $|\mathcal{C}| > 0.1q$. Suppose first that there exists some $x \in B$ contained in more than $0.001q$ copies of C_4 in $\mathcal{C}$. Each of these 4-cycles offers an opposite pair (u_i, v_i) with $u_i, v_i \in V \backslash B$ and thus these opposite pairs span at least $\sqrt{0.001q} > \ell$ vertices in $V \backslash B$. Then we can choose ℓ opposite pairs among them say $\{u_i, v_i\}$ for $i \in [\ell]$ such that all v_i are distinct. By Claim 4.5, there is a vertex $u \in V \backslash B$ with $c(u) > 0.8q$ which contradicts the definition of B. Hence we may assume that every $x \in B$ is contained in at most $0.001q$ copies of C_4 in $\mathcal{C}$. Since $|B| \leq 15$, there are at least $0.085q$ copies of C_4 in $\mathcal{C}$ disjoint with B. These C_4's span at least $\sqrt[4]{8 \times 0.085q} > \ell$ vertices in $V \backslash B$. Using Claim 4.5, there exists some $u' \in V \backslash B$ with $c(u') > 0.8q$, again a contradiction. So $|\mathcal{C}| \leq 0.1q$.

Note that G' has at most $0.1q$ copies of C_4 with $e(G') \geq e(G) - 105$. We further define a graph G'' to be obtained from G' by deleting one edge from each 4-cycle of G'. Thus we have $e(G'') \geq e(G') - 0.1q \geq q(q+1)^2/2 - 0.1q - 104$. Since G'' is C_4-free, by Theorem 1.2, there exists a unique polarity graph H of order q such that $G'' \subseteq H$.

We claim that $G' \subseteq H$. Suppose not. Then there exists an edge $e \in E(G') \backslash E(G'')$ such that $e \notin E(H)$. By Lemma 2.4, $H+e$ contains at least $q-1$ copies of C_4, any two of which are edge-disjoint except sharing e. Note that $G''+e \subseteq H+e$ and as a subgraph of G', $G''+e$ contains at most $0.1q$ copies of C_4. While preserving the edge e, one needs to delete at least $0.9q-1$ edges from $H+e$ to derive $G''+e$. Also as $e(H) \leq q(q+1)^2/2$, we see $e(G'') = e(G''+e) - 1 \leq e(H+e) - (0.9q-1) - 1 \leq q(q+1)^2/2 - 0.9q + 1$, which contradicts the above lower bound on $e(G'')$. This proves $G' \subseteq H$.

Finally, we are ready to complete Theorem 1.3. Suppose there are three edges e, e', e'' in $E^* \backslash E(H)$. By Lemma 2.4, $H + \{e, e', e''\}$ has $3(q-1)$ distinct copies of C_4, $q-1$ copies of which are in $H + \{e\}$, $q-1$ copies of which in $H + \{e'\}$ and $q-1$ copies of which in $H + \{e''\}$. We also see that each edge in H can appear in at most three of these cycles. Since $G' \subseteq H$ and $e(H) - e(G') \leq (e(G)+1) - e(G') \leq |E^*| - 1 \leq 104$, we know $G' + \{e, e', e''\}$ can be obtained from $H + \{e, e', e''\}$ by deleting at most 104 edges. These together show that $G \supseteq G' + \{e, e', e''\}$ has at least $3(q-1) - 312 > 2q$ copies of C_4, a contradiction. Now we may assume $|E^* \backslash E(H)| \leq 2$. Note that $|E(G \cap H)| + |E^* \backslash E(H)| = |E(G)| \geq |E(H)| + 1$. This shows that $1 \leq |E^* \backslash E(H)| \leq 2$ and by Lemma 2.3, $e(H) = q(q+1)^2/2$ (i.e., H is orthogonal). If $|E^* \backslash E(H)| = 1$, then the above inequality implies that G is a graph obtained from H by adding one new edge, as desired. So $|E^* \backslash E(H)| = 2$ and $e(G \cap H) = e(H) - 1$. That is, G is obtained from H by deleting an edge e'' and adding two new edges e, e'. Now a refinement of Lemma 2.4 shows that G contains at least $2q - 3$ copies of C_4. This proves Theorem 1.3. ■

The proof can be sharpened to tell a bit more. For instance, one can characterize all such graphs with exactly $q-1, q, q+1$ or $2q-3$ many 4-cycles, respectively; see [13].

References

[1] R. Baer, Polarities in finite projective planes, *Bull. Amer. Math. Soc.* **52** (1946), 77–93.

[2] W. G. Brown, On graphs that do not contain a Thomsen graph, *Canad. Math. Bull.* **9** (1966), 281–285.

[3] R. H. Bruck and H. J. Ryser, The non-existence of certain finite projective planes, *Canadian J. Math.* **1** (1949), 88–93.

[4] S. Dow, An improved bound for extending partial projective planes, *Discrete Math.* **45** (1983), 199–207.

[5] 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)

[6] P. Erdős, A. Rényi and V. T. Sós, On a problem of graph theory, *Studia Sci. Math. Hungar.* **1** (1966), 215–235.

[7] P. Erdős and M. Simonovits, Cube-supersaturated graphs and related problems, *Progress in Graph Theory (Waterloo, Ont., 1982)*, pp. 203–218, Academic Press, Toronto, Ont., 1984.

[8] Z. Füredi, Graphs without quadrilaterals, *J. Combin. Theory Ser. B* **34** (1983), 187–190.

[9] Z. Füredi, Quadrilateral-free graphs with maximum number of edges, preprint 1988, `http://www.math.uiuc.edu/~z-furedi/PUBS/furedi_C4from1988.pdf`

[10] Z. Füredi, On the number of edges of quadrilateral-free graphs, *J. Combin. Theory Ser. B* **68** (1996), 1–6.

[11] Z. Füredi, Extremal quadrilater-free graphs, Lecture at the 1st Canadian Discrete and Algorithmic Mathematics Conference (CanaDAM 2007), Banff, May 28–31, 2007. `https://canadam.math.ca/2007/MS/MS_Extremal_Haxell.pdf`

[12] 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.

[13] J. He, J. Ma and T. Yang, Some exact results on 4-cycles: stability and supersaturation, arXiv: 1912.00986v4

[14] T. Kővári, V. Sós and P. Turán, On a problem of K. Zarankiewicz, *Colloquium Math.* **3** (1954), 50–57.

[15] K. Metsch, Linear Spaces with Few Lines, *Springer Lecture Notes in Math.*, 1991.

[16] I. Reiman, Über ein Problem von K. Zarankiewicz, *Acta Math. Acad. Sci. Hungar.* **9** (1958), 269–273.

Appendices

A Justification of the inequalities in the proof of Theorem 1.2.

Here we provide detailed calculations to show why the inequalities (2), (3) and (4) in the proof of Theorem 1.2 do not hold. We are given $q \geq 10^9$ in Theorem 1.2. First, we consider (2) by letting

$$\begin{aligned} g(\Delta) &:= 2(q^2+q)\left(\binom{q^2+q+1-\Delta}{2} - (q^2+q)\binom{\frac{q^3+q^2-0.4q-\Delta+2}{q^2+q}}{2}\right) \\ &= (q^2+q-1)\Delta^2 - (2q^4+2q^3+2q^2+2.8q-4)\Delta + (2q^5+5.8q^4+0.4q^3-1.56q^2+3.6q-4). \end{aligned}$$

So $g(x)$ is a quadratic function with the axis of symmetry $x = \frac{q^4+q^3+q^2+1.4q-2}{q^2+q-1} > q+3$. Since $q \geq 10^9$, we have $g(q+3) = -1.2q^4 - 0.6q^3 + 3.64q^2 + 2.2q - 1 < 0$ and $g(q^2+q) = -q^6 + q^5 + 3.8q^4 - 5.4q^3 - 1.36q^2 + 7.6q - 4 < 0$. These imply that $g(\Delta) < 0$ for all $q+3 \leq \Delta \leq q^2+q$. Thus (2) does not hold.

Next we consider (3). Using $n = q^2+q+1$, $\Delta = q+2$ and $e(G) \geq \frac{1}{2}q(q+1)^2 - 0.2q + 1$, it is straightforward to see that the expression

$$\binom{n-2\Delta}{2} - (n-2)\binom{\frac{2e(G)-2\Delta-2n+4}{n-2}}{2} \leq \frac{1}{2(q^2+q-1)}(-1.2q^4 - 1.4q^3 + 10.04q^2 + 8.4q - 12)$$

is negative when $q \geq 10^9$. So (3) does not hold.

Finally we consider (4). Similarly using the above bounds on n, Δ and $e(G)$, we can derive that

$$\binom{n-2\Delta+1}{2} - (n-3)\binom{\frac{2e(G)-2\Delta-2n+4}{n-3}}{2} \leq \frac{1}{2(q^2+q-2)}(-0.2q^4 - 0.4q^3 + 4.04q^2 + 2.8q - 12)$$

is also negative when $q \geq 10^9$. So (4) does not hold. This completes the computational justification needed in the proof of Theorem 1.2. ■