Stability and supersaturation of 4-cycles

Jialin He[*]   Jie Ma[†]   Tianchi Yang[‡]

School of Mathematical Sciences
University of Science and Technology of China
Hefei, Anhui 230026, China.

Abstract

In this paper, we investigate extremal problems on the subject of the 4-cycle C_4, which has played a heuristic important role in the development of extremal graph theory. As a milestone Füredi proved that the extremal number $\mathrm{ex}(q^2 + q + 1, C_4) \leq \frac{1}{2}q(q+1)^2$ holds for $q \geq 14$. This matches with the fundamental construction of Erdős-Rényi-Sós and Brown from finite geometry for prime powers q, thus providing one of the very rare exact results in the field.

Our first result is a stability of Füredi's theorem. We prove that for all even q, every (q^2+q+1)-vertex C_4-free graph with at least $\frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)$ edges must be a spanning subgraph of a unique polarity graph, which is constructed from a finite projective plane. Among others, this gives an immediate improvement on the upper bound of $\mathrm{ex}(n, C_4)$ for infinite many integers n.

A longstanding conjecture of Erdős and Simonovits states that every n-vertex graph with $\mathrm{ex}(n, C_4) + 1$ edges contains at least $(1 + o(1))\sqrt{n}$ copies of C_4. Building on the above stability, we obtain an exact result in this direction and thus confirm Erdős-Simonovits conjecture for an infinite sequence of integers n. In fact our result implies a stronger assertion that generally speaking, whenever $n \gg \ell$ in these circumstances, we can characterize the graphs achieving the ℓth least number of C_4's. This can be extended to more general settings, which provide enhancements on the supersaturation problem of C_4. We also discuss related problems and formalize a conjecture on $\mathrm{ex}(n, C_4)$, whose affirmation would disprove Erdős-Simonovits conjecture for general n.

1 Introduction

Given a graph F, we say a graph is F-*free* if it does not contain F as a subgraph. The *Turán number* $\mathrm{ex}(n, F)$ of F is the maximum number of edges in an n-vertex F-free graph. Turán type and related extremal problems are the central subjects of extremal graph theory. In this paper, we focus on extremal problems on one of the basic and perhaps most influential objects in this area – the cycle C_4 of length four. (For indistinct notations appeared below, we shall refer readers to Section 2.)

Proposed by Erdős [8] more than 80 years ago, the study of $\mathrm{ex}(n, C_4)$ has a rich history. In [36] Reiman showed a general upper bound that $\mathrm{ex}(n, C_4) \leq \frac{n}{4}(1 + \sqrt{4n - 3})$. However it is known that the equality never holds by the Friendship Theorem of Erdős, Rényi and Sós [14]. One can also deduce from the proof of Reiman that if the number of edges in an n-vertex C_4-free graph is close to $\frac{1}{2}n^{3/2}$, then almost all vertices have roughly $\sqrt{n}$ neighbors and almost all pairs of vertices have one common neighbor. This suggests that perhaps in principle, the neighborhoods of vertices can be regarded as lines of certain projective plane. Indeed, using orthogonal polarity graphs constructed from finite projective planes, Erdős-Rényi-Sós [14] and Brown [4] proved a lower bound that

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

---

[*]Email: hjxhjl@mail.ustc.edu.cn.

[†]Email: jiema@ustc.edu.cn. Research supported in part by NSFC grant 11622110.

[‡]Email: ytc@mail.ustc.edu.cn.

These two results together imply an asymptotic formula that $\mathrm{ex}(n, C_4) = \left(\frac{1}{2} + o(1)\right)n^{3/2}$.

Determining the exact value of $\mathrm{ex}(n, C_4)$ in general seems to be extremely difficult and far beyond reach. On the other hand, Erdős conjectured (e.g. [11]) that the orthogonal polarity graph is optimal; that is, the inequality in (1) should be replaced with an equality for all prime powers q. Füredi [18] first confirmed this for $q = 2^k$ in 1983, by showing $\mathrm{ex}(q^2 + q + 1, C_4) \leq \frac{1}{2}q(q+1)^2$ holds for all even q. In 1996, Füredi [20] proved that the same upper bound holds for all $q \geq 14$. We summarize his results as following.

Theorem 1.1 (Füredi, [18, 20])**.** *If $q \notin \{1, 7, 9, 11, 13\}$, then $\mathrm{ex}(q^2 + q + 1, C_4) \leq \frac{1}{2}q(q+1)^2$. Hence for all prime powers $q \geq 14$, $\mathrm{ex}(q^2 + q + 1, C_4) = \frac{1}{2}q(q+1)^2$.*

Füredi also proved that extremal graphs for $q \geq q_0$ must be orthogonal polarity graphs of order q (unpublished, see [21]). More recently, Firke, Kosek, Nash and Williford [23] proved $\mathrm{ex}(q^2 + q, C_4) \leq \frac{1}{2}q(q+1)^2 - q$ for all even integers q, which implies that $\mathrm{ex}(q^2 + q, C_4) = \frac{1}{2}q(q+1)^2 - q$ for all $q = 2^k$. We remark that these results on C_4 belong to the category of rather rare exact results (if not only ones) for the notorious degenerate extremal graph problems (for a comprehensive survey of which, see Füredi-Simonovits [22]).

The first main result in this paper is a stability type theorem, which gives a structural description for C_4-free graphs whose number of edges is close to the existing extremal graph. Polarity graphs are natural candidates for these near-extremal graphs of $\mathrm{ex}(n, C_4)$. The following result asserts that under some circumstances, near-extremal graphs must be polarity graphs or their subgraphs.

Theorem 1.2. *Let q be even and G be a C_4-free graph on $q^2 + q + 1$ vertices with at least $\frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)$ edges. Then there exists a unique polarity graph of order q, which contains G as a subgraph.*

This stability holds in a strong sense, namely, it only *adds* edges when turning graphs into desired configurations. On the other hand, there are infinitely many examples showing that the same conclusion can not hold if the number of edges is lowered to $\frac{1}{2}q(q+1)^2 - q + 1$. Let $q = 2^k$. Then there exists an orthogonal polarity graph H of order q with $\frac{1}{2}q(q+1)^2$ edges. Choose non-adjacent $u, v \in V(H)$ with $d_H(u) = q+1$ and $d_H(v) = q$. Let G be obtained from H by deleting all edges incident to v and then adding a new edge uv. Clearly G is C_4-free and has $\frac{1}{2}q(q+1)^2 - q + 1$ edges. However, G can not be contained by any polarity graph H' of order q, as $d_G(u) = q+2$ and $\Delta(H') = q+1$.[1] For more discussion on this stability result, we direct readers to the concluding remarks.

As a first application of Theorem 1.2, we can derive the following result on $\mathrm{ex}(n, C_4)$. Let $\lambda(q)$ be the maximum number of edges in a polarity graph of order q (if no such graphs exist, set $\lambda(q) = 0$).

Corollary 1.3. *Let q be even. If $\lambda(q) \geq \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)$, then $\mathrm{ex}(q^2 + q + 1, C_4) = \lambda(q)$, where the equality holds only for polarity graphs of order q with $\lambda(q)$ edges; otherwise, $\mathrm{ex}(q^2 + q + 1, C_4) < \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)$. In particular, $\mathrm{ex}(q^2 + q + 1, C_4) \leq \max\left\{\lambda(q), \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)\right\}$.*

This corollary provides an improvement of Füredi's theorem for even integers q; to be precise, it improves the upper bound by $\frac{1}{2}q - o(q)$ for at least infinite many even q (we shall explain why later). This and some other inference on $\mathrm{ex}(n, C_4)$ will be under further discussion in Section 7.

A closely related extremal problem to Turán numbers is "the problem of supersaturated graphs" (quoted from [16]). It studies the following function: for a given graph F and for integers $n, t \geq 1$,

$$h_F(n, t) = \min\{\#F(G) : |V(G)| = n, |E(G)| = \mathrm{ex}(n, F) + t\},$$

where $\#F(G)$ denotes the number of distinct copies of F in a graph G. A notable example is the study of the triangle K_3. In 1941 Rademacher proved $h_{K_3}(n, 1) = \lfloor n/2 \rfloor$. Erdős [9, 10] extended this by showing that there exists some constant $c > 0$ such that $h_{K_3}(n, t) = t\lfloor n/2 \rfloor$ for all $t < cn$. The best possible constant c was proved by Lovász and Simonovits [26] to be $c = 1/2$, confirming a conjecture

---

[1] In fact the number of edges can not be lowered to $\frac{1}{2}q(q+1)^2 - q + 2$ (this will be more involved to justify).

of Erdős. Since then, there has been extensive research for the problem of supersaturated graphs. To just name a few, see [15, 16, 17, 25, 30, 32, 33, 34, 35, 37] and their references.

Returning to our focus, throughout this paper we write

$$h(n, t) = h_{C_4}(n, t) \text{ and } h(n) = h(n, 1).$$

Analogously as Rademacher's result on the triangle, Erdős and Simonovits conjectured that any n-vertex graph with $\text{ex}(n, C_4) + 1$ edges should contain many copies of C_4. This is one of the favourite problems of Erdős [13] and appears in many literatures. A weak version (see [6], Conjecture 42) asserted that $h(n) \geq 2$ for large n and another form (e.g. in [12, 13]) stated that $h(n) \geq c\sqrt{n}$ for some constant $c > 0$. The strongest version of this conjecture is the following.

Conjecture 1.4. *(Erdős and Simonovits [16]) For integers n, $h(n) \geq (1 + o(1))\sqrt{n}$.*

As indicated in [16], if true, this bound will be sharp (for infinite integers n). We remark that a direct application of Theorem 1.2 can show that $h(q^2 + q + 1) \geq (\frac{1}{2} - o(1))q$ for $q = 2^k$.

The following supersaturation is another main result of this paper.

Theorem 1.5. *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.*

We also can determine these graphs with exactly $2q - 3$ copies of C_4. As a corollary, we can confirm Conjecture 1.4 for an infinite sequence of n with an exact result.

Corollary 1.6. *Let $q = 2^k$ for some $k \geq 40$. Then $h(q^2 + q + 1) = q - 1$, where a graph achieves this equality if and only if it is obtained from an orthogonal polarity graph of order q by adding a new edge between (any) two vertices of degree q.*

Let $\ell \geq 1$ be an integer. Enhancing Theorem 1.5, we can further characterize all graphs for which achieve the ℓth least number of copies of C_4.

Theorem 1.7. *Let $q \gg \ell$ and q be even. Let G be a graph on $q^2 + q + 1$ vertices with $\frac{1}{2}q(q + 1)^2 + 1$ edges. Then either G has at least $(\ell + 1)q - (\ell + 1)^2$ copies of C_4, or there exist some $s \in \{1, 2, ..., \ell\}$ and an orthogonal polarity graph H of order q such that G can be obtained from H by deleting or adding $2s - 1$ edges. In the latter case, the number of copies of C_4 in G is between $sq - s^2$ and $sq + s^2$.*

This also indicates that the numbers of copies of C_4 among all such graphs are distributed sporadically (concentrated around sq for small integers s).

For the general supersaturation problem of C_4, the function $h(n, t)$ is known to be $\Theta(t^4/n^4)$ when $t = \Omega(n^{3/2})$ for all n (e.g. see [16]). We show in the following result that for an infinite sequence of integers n, one can say rather accurately about the function and in particular, one can determine the order of its magnitude for *every* positive integer t.

Theorem 1.8. *The following statements hold for large $q = 2^k$. **(A)** For every $1 \leq t < q^{1/8}/30$, $h(q^2 + q + 1, t) = t(q - 1)$, where the equality holds for graphs G if and only if G is obtained from an orthogonal polarity graph of order q by adding a matching of size t among vertices of degree q.*
***(B)** For every $t \geq 1$, $h(q^2 + q + 1, t) \geq (\frac{1}{2} + o(1))tq$ and $h(q^2 + q + 1, t) = \Theta(tq + t^4/q^8)$.*

This follows by Theorems 9.1 and 9.2 which are stated in some more general settings. We refer readers to Section 9 for their precise statements. As for general n, one also can determine the order of the magnitude of $h(n, t)$ when $t = \Omega(n^{3/2-\epsilon})$ for some $\epsilon \geq 0.2375$.

Proposition 1.9. *Let n be sufficiently large. If $t \geq 3n^{1.2625}$, then $h(n, t) = \Theta(t\sqrt{n} + t^4/n^4)$.*

The organization of this paper is as follows. Section 2 consists of preliminaries, where we give notations and collect some results. In Section 3, we outline the proof of Theorem 1.2, which is divided and completed in Sections 4, 5 and 6. In Section 7, we prove Corollary 1.3 and discuss other consequences on $\mathrm{ex}(n, C_4)$. In Section 8, we consider Conjecture 1.4 and prove Theorem 1.5, Corollary 1.6 and Theorem 1.7. In Section 9, we prove Theorem 1.8 and Proposition 1.9 for the supersaturation of C_4. In Section 10, we discuss several problems in relation to the results here, e.g., Conjecture 10.2 whose affirmation would disprove Conjecture 1.4 of Erdős-Simonovits. We would like to remark that though our results often are stated with parity condition, many arguments in the proofs in fact work without any parity constraints.

2 Preliminaries

2.1 General notations

We follow the notations of Füredi (e.g. [19]). A *hypergraph* $\mathcal{H}$ is an ordered pair $(V, \mathcal{E})$, where V is a finite set consisting of *vertices* and $\mathcal{E}$ is a collection of subsets (called *edges*) of V. We use $e(\mathcal{H})$ to denote the number of edges in $\mathcal{H}$. For $x \in V$, the *degree* $d_{\mathcal{H}}(x)$ of x denotes the number of edges of $\mathcal{H}$ containing x. The *maximum degree* of $\mathcal{H}$ is denoted by $\Delta(\mathcal{H}) = \max\{d_{\mathcal{H}}(x) : x \in V\}$. We say $\mathcal{H}$ is *k-regular* if all vertices have degree k and *k-uniform* if all edges have k vertices. A *k*-uniform hypergraph is also called a *k-graph* (and a *graph* if $k = 2$). We say $\mathcal{H}$ is *1-intersecting* if any two distinct edges of $\mathcal{H}$ have exactly one common vertex. The *incidence matrix* of a hypergraph $\mathcal{H} = (V, \mathcal{E})$ is an $|\mathcal{E}| \times |V|$ matrix $\mathcal{M}(\mathcal{H})$ such that $\mathcal{M}(E, x) = 1$ if $x \in E \in \mathcal{E}$ and 0 otherwise.

Let $G = (V, \mathcal{E})$ be a graph. Let $x \in V$ and $A \subseteq V$. The *neighborhood* $N_G(x)$ of x is the set of vertices $y \in V$ with $xy \in \mathcal{E}$, while the *closed neighborhood* $N_G[x]$ is defined by $N_G(x) \cup \{x\}$. Let $N_A(x) = N_G(x) \cap A$. Define $N_G(A)$ to be the set of vertices $u \in V \backslash A$ adjacent to some vertex in A and $G[A]$ to be the subgraph of G induced on A. For a path P, its *length* $|P|$ denotes the number of edges it contains. We say P is a *k-path* if $|P| = k$. For disjoint sets $A, B \subseteq V$, $e(A, B)$ denotes the number of edges ab in G with $a \in A$ and $b \in B$. A set of edges is called *independent* if their endpoints are pairwise-disjoint. For $u, v \in V$, we let $d_G(u, v) = |N_G(u) \cap N_G(v)|$. We call $\{u, v\}$ an *uncovered* pair if $d_G(u, v) = 0$ and a *covered* pair otherwise. Let UP be the set of uncovered pairs of G and let P_2 be the set of all 2-paths in G. The *adjacency matrix* $\mathcal{A}(G)$ of G is a $|V| \times |V|$ symmetric matrix such that $\mathcal{A}(x, y) = 1$ if $xy \in \mathcal{E}$ and 0 otherwise.

Throughout this paper, the notation $\binom{x}{2}$ means the function $x(x-1)/2$ for all reals x. For any positive integer k, we write $[k]$ as the set $\{1, 2, ..., k\}$. For all above notations, we often drop the subscripts when they are clear from context.

2.2 Projective planes

A *finite projective plane of order* q, denoted by $PG(2, q)$, 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 also follows that $|\mathcal{L}| = q^2 + q + 1$ and any two points are contained in a unique line. The existence of $PG(2, q)$ is well known for all prime powers q. On the other hand, a major conjecture in this field asserts that the order q of $PG(2, q)$ must be a prime power (known for $q \leq 11$ and still open for $q = 12$).

A substantial body of our proofs will be involved with projective planes and 1-intersecting hypergraphs. In preparation we now collect some related results. The first two will play important roles for the constructive nature in our stability result (Theorem 1.2).

Theorem 2.1 ([28]). *Let $q \geq 3900$ and $\mathcal{H}$ be a 1-intersecting $(q+1)$-hypergraph with q^2+q+1 vertices and more than $q^2 - \frac{\sqrt{5}-1}{2}q + 17\sqrt{q/5}$ edges. Then $\mathcal{H}$ can be embedded into a projective plane of order q.*

Theorem 2.2 ([7]). *Let $\mathcal{H}$ be a 1-intersecting $(q+1)$-hypergraph 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.*

The following celebrated Bruck-Ryser theorem [5] gives a sufficient condition for the non-existence of projective planes of given order.

Theorem 2.3 ([5]). *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.*

We also need a useful lemma proved by Füredi (i.e., Lemma 3.7 in [19]).

Lemma 2.4 ([19]). *Let $M = (m_{ij})$ be the incidence matrix of a projective plane of order q. Suppose that $m_{ij} = m_{ji}$ whenever $1 \leq i \leq q^2 - q + 3$ or $1 \leq j \leq q^2 - q + 3$. Then the whole matrix M is symmetric.*

The coming lemma has been used in literatures (e.g. [7]), which will serve as a handy tool for finding a large 1-intersecting hypergraph in the proof of Theorem 1.2. For completion, we give a proof.

Lemma 2.5 (e.g. [7]). *Let $\mathcal{H}$ be a 1-intersecting $(q+1)$-hypergraph on vertex set V with $|V| = q^2+q+1$. Suppose that $\mathcal{F}$ is a $(q+1)$-uniform hypergraph on the same vertex set V such that $\mathcal{F} \cap \mathcal{H} = \emptyset$ and for any edge $f \in \mathcal{F}$, there exist q edges $h_1, ..., h_q \in \mathcal{H}$ satisfying that $f \cup h_1 \cup ... \cup h_q = V$ and $|f \cap h_1 \cap ... \cap h_q| = 1$. Then $\mathcal{H} \cup \mathcal{F}$ is also 1-intersecting.*

Proof. We first point out that to show this, it suffices to prove that for any $f \in \mathcal{F}$, $\mathcal{H} \cup \{f\}$ is 1-intersecting. This is because if we initially set $\mathcal{G} = \mathcal{H}$ and repeatedly operate by applying this for one edge $f \in \mathcal{F}$ at a time and updating $\mathcal{G}$ by $\mathcal{G} \cup \{f\}$, then in the end we would conclude that $\mathcal{G} = \mathcal{H} \cup \mathcal{F}$ is 1-intersecting. Note that this indeed is valid as the conditions in the statement also hold for $\mathcal{G}$ (instead of $\mathcal{H}$) at any given time.

For the above desired statement, suppose on the contrary that there exist $f \in \mathcal{F}$ and $h \in \mathcal{H}$ such that $|h \cap f| = 0$ or $|h \cap f| \geq 2$. We know that there are $h_1, ..., h_q \in \mathcal{H}$ and $u \in V$ such that $f \cup h_1 \cup ... \cup h_q = V$ and $f \cap h_1 \cap ... \cap h_q = \{u\}$. By the size of V, we also see that $f \setminus \{u\}, h_1 \setminus \{u\}, ..., h_q \setminus \{u\}$ must form a partition of $V \setminus \{u\}$. It is then clear that $h \notin \{f, h_1, \cdots, h_q\}$. If $|h \cap f| = 0$, then there must exist some $i \in [q]$ such that $|h \cap h_i| \geq 2$, a contradiction to that $\mathcal{H}$ is 1-intersecting. Hence we may assume $|h \cap f| \geq 2$. Suppose $u \in h$. Then $u \in h \cap h_i$ for all $i \in [q]$; since $\mathcal{H}$ is 1-interesting, we conclude that $h \cap (h_1 \cup ... \cup h_q) = \{u\}$ and thus $h = f$, a contradiction. Now suppose $u \notin h$. Then $|h \cap (f \setminus \{u\})| \geq 2$ and thus there exists some $j \in [q]$ such that $|h \cap (h_j \setminus \{u\})| = 0$, which also shows that $|h \cap h_j| = 0$, a contradiction. We have completed the proof now. □

2.3 Polarity graphs

A *polarity* π of a projective plane $\mathcal{H} = (P, \mathcal{L})$ is a bijection $\pi : P \cup \mathcal{L} \to P \cup \mathcal{L}$ such that

- π^2 is the identity function with $\pi : P \leftrightarrow \mathcal{L}$, and

- for any pair $(x, L) \in P \times \mathcal{L}$, if $x \in L$ then $\pi(L) \in \pi(x)$.

For a projective plane $\mathcal{H} = (P, \mathcal{L})$ of order q, where $P = \{x_i\}$ and $\mathcal{L} = \{L_i\}$, consider a function $\phi : P \leftrightarrow \mathcal{L}$ which maps $x_i \leftrightarrow L_{\sigma(i)}$ for some permutation σ on $[q^2 + q + 1]$. Let $\mathcal{M}(\phi)$ be the incidence matrix of $\mathcal{H}$, where the rows are listed in the order of x_i's and the columns are listed in the order of $L_{\sigma(i)}$'s as i increases. It is worth pointing out that

$$\text{the function } \phi \text{ is a polarity} \iff \text{the incidence matrix } \mathcal{M}(\phi) \text{ is symmetric.} \tag{2}$$

Now let π be a polarity of a projective plane $\mathcal{H} = (P, \mathcal{L})$ of order q. 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 $x \in \pi(y)$. A

point $x \in P$ is called *absolute* (with respect to π) if $x \in \pi(x)$. Let $a(\pi)$ denote the number of absolute points. In [2], Baer proved that

$$\text{there exists some integer } m_\pi \geq 0 \text{ such that } a(\pi) = q + 1 + m_\pi \cdot \sqrt{q}. \qquad (3)$$

A polarity π and its polarity graph $G(\pi)$ are called *orthogonal*, if $a(\pi) = q + 1$ (i.e., $m_\pi = 0$). It is known that for any prime power q, there always exists an orthogonal polarity graph of order q.

Combining the above facts, it is easy to derive the following for polarity graphs.

Proposition 2.6. *Let π be a polarity 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 every vertex has degree q or $q+1$.*

The following lemma on polarity graphs is well-known (see Baer [1] for a proof).

Lemma 2.7. *Any two vertices of degree q in a polarity graph of order q are nonadjacent.*

The next lemma will be frequently used in the forthcoming proofs.

Lemma 2.8. *Let G be a polarity graph of order q with $uv \notin E(G)$. Then $G \cup \{uv\}$ contains either $q-1, q$ 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 G be the polarity graph of π, where π is a polarity of some $\mathcal{H} = PG(2, q)$. Let $uv \notin E(G)$. Then $q \leq d_G(u), d_G(v) \leq q + 1$. For any $u_i \in \pi(u)$, since $\mathcal{H}$ is 1-intersecting, there exists a unique vertex v_i in $\pi(u_i) \cap \pi(v)$. Hence 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 $i \in [q + 1]$.

First assume that $d_G(u) = d_G(v) = q$. Among the above $q+1$ sequences, there are special sequences uuv_iv and uu_jvv (which are obviously different). So there are exactly $q - 1$ sequences uu_iv_iv satisfy $u \neq u_i$ and $v \neq v_i$. We claim that $u_i \neq v_i$. Suppose on the contrary that $u_i = v_i$. Then $u_i \in \pi(u_i)$, contradicting Lemma 2.7 as $uu_i \in E(G)$ and both u, u_i have degree q in G. Hence there are exactly $q - 1$ paths uu_iv_iv of length three between u and v in G, which give $q - 1$ four-cycles in $G \cup \{uv\}$.

Now without loss of generality, let $N_G(u) = \{u_1, ..., u_{q+1}\}$. We claim that there is at most one index $i \in [q + 1]$ satisfying $v_i \in \{u_i, v\}$. To see this, note that if $v_i \in \{u_i, v\}$, then $u_i \in N_G(u) \cap N_G(v)$. Therefore two such $i, j \in [q + 1]$ would force a four-cycle uu_ivu_ju in G, a contradiction. From this claim, we see the number of paths of length three between u and v in G is either q or $q + 1$. So there are q or $q + 1$ four-cycles in $G \cup \{uv\}$.

To complete the proof, it suffices to show that all the above paths uu_iv_iv of length three in G are edge-disjoint. This follows by the fact that each middle edge u_iv_i can only appear once. Indeed, if there are two paths uu_iv_iv and uv_iu_iv in G, then this would force a four-cycle uu_ivv_iu in G. $\square$

We also need a property on orthogonal polarity graphs of even order q from [19].

Proposition 2.9. *Let q be even and G be an orthogonal polarity graph of order q. Then there exists a (unique) vertex w of degree $q + 1$ such that $N(w)$ consists of all vertices of degree q in G.*

2.4 C_4-free graphs

We now give out notations arising from C_4-free graphs, and along the way we also establish some statements for future use.

Throughout this subsection, let $G = (V, \mathcal{E})$ be a C_4-free graph on $n = q^2 + q + 1$ vertices. For a vertex v, let $d_0(v) = |\{u \in V : \{u, v\} \in UP\}|$.

Proposition 2.10. $|UP| = \frac{1}{2}\sum_{v \in V} d_0(v)$ *and* $|P_2| + |UP| = \binom{n}{2}$.

Proof. The first equation follows by the definition. Since G is C_4-free, each 2-path corresponds to a unique covered pair. Thus we have $|P_2| = \binom{n}{2} - |UP|$. $\square$

6

We now introduce an important notation for our proofs. For any $v \in V$, the *deficiency* $f(v)$ of v is defined by

$$f(v) := \max\{q + 1 - d(v), 0\}.$$

The deficiency of a subset $A \subseteq V$ is $f(A) = \sum_{v \in A} f(v)$. Let $S_i = \{v \in V : d(v) = i\}$.

Proposition 2.11. *We have $f(N(v)) \geq q$ for each $v \in S_{q+2}$, and if $\Delta(G) \leq q + 2$ then $f(V) = (q + 1)n - 2e(G) + |S_{q+2}|$.*

Proof. Consider $v \in S_{q+2}$ with $N(v) = \{v_1, ..., v_{q+2}\}$. Since G is C_4-free, $N(v_i)\backslash\{v\}$ are pairwise-disjoint, which implies that $\sum_{v_i \in N(v)}(d(v_i) - 1) \leq n - 1$. Thus we get

$$f(N(v)) \geq \sum_{v_i \in N(v)} (q + 1 - d(v_i)) \geq (q + 2)q - (n - 1) = q.$$

Since $\Delta(G) \leq q + 2$, it is straightforward to see that $f(V) = \sum_{v \in V} \max\{q + 1 - d(v), 0\} = (q + 1)n - 2e(G) + |S_{q+2}|$. $\square$

Lemma 2.12. *Let $S = \{v \in V : d(v) \leq q\}$. If q is even and $\Delta(G) = q + 1$, then any vertex in S_{q+1} has a neighbor in S and moreover, $|S| \geq q + 1$.*

Proof. Suppose on the contrary that there exists some $v \in S_{q+1}$ and all its neighbors have degree $q + 1$. Let m be the number of edges contained in $G[N(v)]$. Clearly these edges form a matching (as otherwise there is a C_4) and since q is even, we have $m \leq \frac{q}{2}$. We count the number M of edges between $N(v)$ and $V\backslash N(v)$. As G is C_4-free, every vertex in $V\backslash N[v]$ has at most one neighbor in $N(v)$. Hence, we have that

$$q^2 + q = n - (q + 2) + (q + 1) \geq M = \sum_{x \in N(v)} d(x) - 2m \geq (q + 1)^2 - q.$$

This is a contradiction, proving the first assertion. By counting the number of edges between S and S_{q+1}, we also see that $n - |S| = |S_{q+1}| \leq \sum_{x \in S} d(x) \leq q|S|$. So $(q + 1)|S| \geq n = q^2 + q + 1$, implying that $|S| \geq q + 1$. $\square$

2.5 Others

We need the following estimation on the distribution of prime numbers given in [3].

Theorem 2.13 ([3]). *For sufficiently large $x > 0$, the interval $[x - x^{0.525}, x]$ contains prime numbers.*

At the end of this section, we give an easy-to-use lemma, which is often adopted in replace of standard Cauchy-Schwarz inequalities.

Lemma 2.14. *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 we have $\sum_{i=1}^{m} \binom{a_i}{2} \geq m\binom{k}{2} + rk$.*

Proof. First we see that if $0 \leq a_i \leq a_j - 2$, then $\binom{a_i}{2} + \binom{a_j}{2} > \binom{a_i+1}{2} + \binom{a_j-1}{2}$. Let $\sum_{i=1}^{m} a_i = tm + x$ for integers t, x with $0 \leq x < m$. By the above fact, we have $\sum_{i=1}^{m} \binom{a_i}{2} \geq x\binom{t+1}{2} + (m-x)\binom{t}{2} = m\binom{t}{2} + xt$. By letting $\Lambda := \sum_{i=1}^{m} \binom{a_i}{2} - m\binom{k}{2} - rk$, we get that

$$\Lambda \geq m\binom{t}{2} + xt - m\binom{k}{2} - rk = \frac{1}{2}m(t - k)(t + k - 1) + xt - rk. \tag{4}$$

Since $(t + 1)m > tm + x = \sum_{i=1}^{m} a_i \geq km + r \geq (k - 1)m$, we obtain $t - k \geq -1$. We now consider the following three cases. If $t - k = -1$, we have $-r \geq m - x > 0$ and by (4), $\Lambda \geq (x - m)(k - 1) - rk \geq m - x > 0$. If $t = k$, we have $x \geq r$ and by (4) it is easy to see that $\Lambda \geq k(x - r) \geq 0$. Lastly, we assume $t \geq k + 1$. Since $t + k - 1 \geq 2k$ and $m(t - k) \geq r - x$, by (4), again we have $\Lambda \geq (r - x)k + xt - rk = x(t - k) \geq 0$. $\square$

3   Proof outline of Theorem 1.2

In this section we discuss the proof of Theorem 1.2. For convenience, we restate Theorem 1.2 in the following thorough version.

Theorem 3.1. *For any $c \in (0,1)$, there exists some q_c such that the following holds for even integers $q \geq q_c$. If G is a C_4-free graph on $q^2 + q + 1$ vertices with at least $\frac{1}{2}q(q+1)^2 - \frac{c}{2}q$ edges, then there exists a unique polarity graph of order q containing G as a subgraph.*

We now give a outline of the proof of Theorem 3.1. In a nutshell, it stems from the work of Füredi [18, 19, 20]. Given a C_4-free graph G on $q^2 + q + 1$ vertices with many edges, our goal is to construct a polarity graph of order q containing G as a subgraph. This is achieved in the following three steps.

Step 1. We show that it suffices to consider for $\Delta(G) = q + 1$.

Step 2. Let $\mathcal{R}$ be the family of all subsets $N_G(x)$ where $x \in V(G)$ has degree $q + 1$ and "almost" all neighbors of x have degree $q + 1$. Then we show that there exists a projective plane $\mathcal{H}$ of order q defined on $V(G)$ with $\mathcal{R} \subseteq \mathcal{H}$.

Step 3. We show that there exists a polarity π of the above projective plane $\mathcal{H}$ such that its polarity graph $G(\pi)$ contains G as a subgraph.

To say more, Step 1 will be handled in Section 4, where we reduce Theorem 3.1 to the following statement (with restriction $\Delta(G) = q + 1$).

Theorem 3.2. *For any $\epsilon \in (0,1)$, there exists some q_ϵ such that the following holds for even integers $q \geq q_\epsilon$. If G is a C_4-free graph on $q^2 + q + 1$ vertices with maximum degree $q + 1$ and at least $\frac{1}{2}q(q+1)^2 - \frac{\epsilon}{2}q$ edges, then there exists a unique polarity graph of order q containing G as a subgraph.*

Our reduction shows that $\epsilon = c + o(1)$ holds for Theorem 3.2 $\Longrightarrow c$ holds for Theorem 3.1.

For Theorem 3.2, we will divide its proofs into Sections 5 and 6. In Section 5, we complete Step 2 by establishing Lemma 5.1, which asserts that there exists a 1-intersecting $(q + 1)$-hypergraph $\mathcal{H}$ containing $\mathcal{R}$ and at least q^2 lines. This indeed is enough to accomplish Step 2 as we can apply Theorem 2.1 to enlarge $\mathcal{H}$ into a projective plane of order q containing $\mathcal{R}$. The proof of Lemma 5.1 is involved, where the main technical difficulty lies in accurate analysis on the intricate relations between neighborhoods of vertices of degree q or $q+1$. Finally, we finish Step 3 in Section 6 and thus the proof of Theorem 3.2. The arguments of this step will heavily rely on the properties of the family $\mathcal{R}$.

The proof of Theorem 1.2 is quite lengthy and involved. On the other hand, we can have a 6-page-long proof [24] for showing the same stability statement with a weaker bound $\frac{1}{2}q(q+1)^2 - cq$ on the number of edges for some constant $c \geq 0.2$. It is worth pointing out that such a weaker stability can be applied to derive Corollary 1.6. However, we choose to present the current proof for $c = \frac{1}{2} - o(1)$ as the limit of our endeavours (there are significant difficulties in all three steps when considering $c \geq \frac{1}{2}$). For doing so, not only we think the value of c in the stability on its own is an intersecting and important problem, but also it has impacts on several conjectures and problems in extremal graph theory as we will discuss in Section 10. We now precess to the proof of Theorem 1.2 in Sections 4, 5 and 6.

4   Reducing to $\Delta = q + 1$

In this section, as outlined earlier, we present a proof which reduces Theorem 3.1 to Theorem 3.2.

Proof of Theorem 3.1 (Assuming Theorem 3.2). For any $c \in (0,1)$, we define ϵ to be any real in $(c, 1)$ and choose q_c such that $c + \frac{50}{\sqrt{q_c}} \leq \epsilon < 1$ and $q_c \geq \max\left\{q_\epsilon, \frac{2500}{(\epsilon-c)^2}\right\}$, where q_ϵ is from Theorem 3.2. Let $q \geq q_c$ be an even integer and let G be a C_4-free graph on $n = q^2 + q + 1$ vertices such that

$$e(G) \geq \frac{1}{2}q(q+1)^2 - \frac{c}{2}q.$$

8

We will show that there exists a unique polarity graph of order q containing G as a subgraph.

Let Δ denote the maximum degree of G. By the lower bound on $e(G)$, it is easy to see that $\Delta \geq q+1$. If $\Delta = q+1$, since $e(G) \geq \frac{1}{2}q(q+1)^2 - \frac{c}{2}q \geq \frac{1}{2}q(q+1)^2 - \frac{\epsilon}{2}q$ and $q \geq q_c \geq q_\epsilon$, our goal is accomplished by Theorem 3.2. So we may assume that $q+2 \leq \Delta \leq q^2 + q$.

Let $V(G) = \{v_1, ..., v_n\}$. We now process by showing a sequence of claims.

Claim 4.1. $\Delta = q + 2$.

Proof. Suppose that $d(v_1) = \Delta \geq q+3$. We now estimate the number T of 2-paths in G with none of its endpoints in $N(v_1)$. Since any two vertices have at most one common neighbor and any two vertices in $N(v_i)$ are contained in a 2-path, we have

$$\binom{q^2+q+1-\Delta}{2} = \binom{n-\Delta}{2} \geq T \geq \sum_{i=2}^{n} \binom{|N(v_i) \setminus N(v_1)|}{2}.$$

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 $q \geq q_c$ being sufficiently large, we have that

$$\sum_{i=2}^{n} |N(v_i) \setminus N(v_1)| \geq 2e(G) - \Delta - (n-1) \geq (q^2+q)(q-1) + (q^2+2-\Delta).$$

As $q^2 + 2 - \Delta \geq -(q^2+q)$, using Lemma 2.14 (with $m = n - 1 = q^2 + q$), we have

$$\binom{q^2+q+1-\Delta}{2} \geq \sum_{i=2}^{n} \binom{|N(v_i) \setminus N(v_1)|}{2} \geq (q^2+q)\binom{q-1}{2} + (q-1)(q^2+2-\Delta).$$

After simplification, this is equivalent to that $g(\Delta) := \Delta^2 - (2q^2+3)\Delta + (2q^3 + 5q^2 - 5q + 4) \geq 0$, where $q + 3 \leq \Delta \leq q^2 + q$. It can be verified that $g(q+3)$ and $g(q^2+q)$ both are negative. Since $g(\Delta)$ is quadratic, this shows that $g(\Delta) < 0$ for all choices of Δ, a contradiction. This completes the proof of this claim. $\square$

Claim 4.2. *Any two vertices of degree $q+2$ have one common neighbor.*

Proof. Suppose for a contradiction that there exist $v_1, v_2 \in S_{q+2}$ such that $N(v_1) \cap N(v_2) = \emptyset$. Then for $3 \leq i \leq n$ we have $|N(v_i) \setminus (N(v_1) \cup N(v_2))| = |N(v_i)| - |N(v_i) \cap N(v_1)| - |N(v_i) \cap N(v_2)| \geq d(v_i) - 2$. Similarly as above, we estimate the number of 2-paths with none of its endpoints in $N(v_1) \cup N(v_2)$. Using Jensen's inequality, we get that

$$\binom{n - 2(q+2)}{2} \geq \sum_{i=3}^{n} \binom{|N(v_i) \setminus (N(v_1) \cup N(v_2))|}{2} \geq (n-2)\binom{\frac{\sum_{i=3}^{n}(d(v_i)-2)}{n-2}}{2}$$

$$= (n-2)\binom{\frac{2e(G)-2(q+2)-2(n-2)}{n-2}}{2} \geq (q^2+q-1)\binom{\frac{q^3-3q-cq-2}{q^2+q-1}}{2}$$

which is equivalent to

$$(q^2+q-1)(q^2-q-3)(q^2-q-4) \geq (q^3-3q-cq-2)(q^3-q^2-4q-cq-1).$$

After further simplification, we can derive that

$$2(c-1)q^4 + (3-c)q^3 + (11-7c-c^2)q^2 - 3(2+c)q - 14 \geq 0.$$

Since $0 < c < 1$, this inequality can not hold for large q, a contradiction. $\square$

Claim 4.3. *Any common neighbor of three vertices of degree $q+2$ has degree at most $q/2$.*

Proof. Suppose on the contrary that $v_1, v_2, v_3 \in S_{q+2}$ have a common neighbor v_4 such that $d = d(v_4) > q/2$. Let $A = N(v_4)\backslash\{v_1, v_2, v_3\}$, $B = V\backslash N[v_4]$ and $C = N(v_1) \cup N(v_2) \cup N(v_3)$. So $|A| = d - 3$ and $|B| = q^2 + q - d$. By similar discussion as before, we have that $|N(v) \backslash C| \geq d(v) - 1$ for $v \in A$ and $|N(u) \backslash C| \geq d(u) - 3$ for $u \in B \cup \{v_4\}$. Therefore,

$$\sum_{v \in A}(d(v) - 1) + \sum_{v \in B}(d(v) - 3) = (2e(G) - 3(q+2) - d) - |A| - 3|B|$$

$$\geq (q(q+1)^2 - cq) - 3q^2 - 6q + d - 3 \geq (q^2 + q - 3)(q-2) - (q/2 + 9).$$

Note that $-(q/2 + 9) > -(q^2 + q - 3)$. Using Lemma 2.14, if we estimate the number of 2-paths T with none of its endpoints in C, then we can derive that

$$\binom{n - 3(q+2) + 2}{2} \geq T \geq \binom{d - 3}{2} + \sum_{v \in A}\binom{d(v) - 1}{2} + \sum_{v \in B}\binom{d(v) - 3}{2}$$

$$\geq \binom{q/2 - 3}{2} + (q^2 + q - 3)\binom{q - 2}{2} - (q - 2)(q/2 + 9).$$

This inequality is equivalent to $q^2 - 50q + 72 \leq 0$, which contradicts that q is large. □

Claim 4.4. *There are at most 7 vertices of degree at most $q/2$.*

Proof. Suppose on the contrary that there are at least 8 vertices of degree at most $q/2$. Recall the notations in Subsection 2.4. For any $v \in S_{q+1-k}$ for $k \in [q]$, we have

$$d_0(v) = (n - 1) - \sum_{u \in N(v)}(d(u) - 1) = (q^2 + 2q + 1 - k) - \sum_{u \in N(v)}d(u)$$

$$\geq (q^2 + 2q + 1 - k) - (q + 2)(q + 1 - k) = (k - 1)(q + 1),$$

where the inequality holds as $\Delta = q + 2$. By Proposition 2.10, we have

$$\binom{n}{2} - |P_2| = |UP| = \frac{1}{2}\sum_{v \in V}d_0(v) \geq \frac{q + 1}{2}\sum_{k=1}^{q}(k - 1)|S_{q+1-k}| \geq \frac{q + 1}{2} \cdot \frac{q}{2} \cdot 8 = 2q^2 + 2q.$$

We also have $|P_2| = \sum_{v \in V}\binom{d(v)}{2}$ and $\sum_{v \in V}d(v) \geq q(q + 1)^2 - cq = (q^2 + q + 1)(q + 1) - (q + cq + 1)$. Since $-(q + cq + 1) > -(q^2 + q + 1)$, by Lemma 2.14 we get

$$|P_2| \geq (q^2 + q + 1)\binom{q + 1}{2} - (q + 1)(q + cq + 1) = \binom{n}{2} - (q + 1)(q + cq + 1).$$

Combining with the above two inequalities, we derive that

$$2q^2 + 2q - (q + 1)(q + cq + 1) = (q + 1)((1 - c)q - 1) \leq 0.$$

Again, this contradicts the fact that q is large and thus completes the proof of Claim 4.4. □

Claim 4.5. $|S_{q+2}| \leq 22\sqrt{q}$.

Proof. Let $S' = \{v : d(v) \leq q/2\}$ and $S'_{q+2} = S_{q+2}\backslash N(S')$. By Claim 4.4 we have $|S'| \leq 7$, so $|S_{q+2} \cap N(S')| \leq 7q/2$ and $|S'_{q+2}| \geq |S_{q+2}| - 7q/2$. Recall $S = \{v : d(v) \leq q\}$.

We now define a weight function w on the edges xy with $x \in S'_{q+2}$ and $y \in S$ by assigning $w(xy)$ to be the deficiency $f(y)$. We consider the total weight W of these edges. On the one hand, by Proposition 2.11, every S'_{q+2} vertex contributes at least q and

$$f(S) = f(V) = (q + 1)n - 2e(G) + |S_{q+2}| \leq q + cq + 1 + |S_{q+2}|.$$

On the other hand, by Claim 4.3, every vertex in S is adjacent to at most two vertices in S'_{q+2} and thus contributes at most twice of its deficiency. Putting these together, we have

$$q(|S_{q+2}| - 7q/2) \le q|S'_{q+2}| \le W \le 2f(S) \le 2(q + cq + 1 + |S_{q+2}|), \tag{5}$$

from which we derive that $|S_{q+2}| \le \frac{3.5q^2 + 2(1+c)q + 2}{q-2} \le 4q$. For any $v \in S'$, let $N(v) \cap S_{q+2} = \{u_1, ..., u_t\}$. Since $N(u_i) \backslash \{v\}$ are disjoint for all $i \in [t]$, by Proposition 2.11 it follows that

$$f(V) \ge f(v) + \sum_{i=1}^{t} f(N(u_i) \backslash \{v\}) \ge f(v) + t(q - f(v)) = (d(v) - 1)(t - 1) + q.$$

Since $d(v) \ge t$ and $|S_{q+2}| \le 4q$, by (5) we have

$$(t-1)^2 + q \le f(V) \le q + cq + 1 + |S_{q+2}| \le 6q + 1,$$

implying that $|N(v) \cap S_{q+2}| = t \le \sqrt{5q+1} + 1 \le 3\sqrt{q}$ for any $v \in S'$.

We improve the estimation of $|S_{q+2} \cap N(S')| \le 21\sqrt{q}$ and can run the above procedure again. Since $|S'_{q+2}| \ge |S_{q+2}| - 21\sqrt{q}$, we have

$$q(|S_{q+2}| - 21\sqrt{q}) \le q|S'_{q+2}| \le W \le 2f(S) \le 2(q + cq + 1 + |S_{q+2}|).$$

This shows that $|S_{q+2}| \le \frac{21q\sqrt{q} + 2(1+c)q + 2}{q-2} \le 22\sqrt{q}$. $\qquad \square$

We are ready to complete the proof. Since $|S_{q+2}| \le 22\sqrt{q}$, we can delete at most $22\sqrt{q}$ edges from G to get a subgraph G' with maximum degree $q + 1$. By the choice of ϵ, we have

$$e(G') \ge e(G) - 22\sqrt{q} \ge \frac{1}{2}q(q+1)^2 - \frac{c}{2}q - 22\sqrt{q} \ge \frac{1}{2}q(q+1)^2 - \frac{\epsilon}{2}q. \tag{6}$$

By Theorem 3.2, there exists a unique polarity graph H containing G' as a subgraph. Let $e_1, ..., e_t$ be the edges deleted from G, where $t \le 22\sqrt{q}$. We may assume that $t \ge 1$, as otherwise $G = G'$ is a subgraph of H. So we have $e_1 = xy \notin E(H)$. By Lemma 2.8, $H \cup \{e_1\}$ contains at least $q - 1$ copies of C_4, all of which contain e_1 and are edge-disjoint otherwise.

Consider $G' \cup \{e_1\}$, which is a subgraph of G and thus is C_4-free. So any of these C_4's found in $H \cup \{e_1\}$ must have an edge not in $G' \cup \{e_1\}$, which are pairwise-distinct. This shows that $e(G') \le e(H) - (q-1) \le \frac{1}{2}q(q+1)^2 - (q-1)$, which contradicts (6). This finishes the proof of Theorem 3.1. $\blacksquare$

We point out that the reduction in this section works for *all* integers q, not necessarily for even q.

5 Finding a large 1-intersecting hypergraph

We prove Theorem 3.2 in the following two sections. The goal of this section is to construct a 1-intersecting $(q+1)$-hypergraph, which represents the C_4-free graph considered in Theorem 3.2 (see Lemma 5.1 below).

Let ϵ and G be from Theorem 3.2 throughout this section. Namely, $\epsilon \in (0, 1)$ is a fixed constant and $G = (V, \mathcal{E})$ is a C_4-free graph on $n = q^2 + q + 1$ vertices with maximum degree $q + 1$ and

$$e(G) \ge \frac{1}{2}q(q+1)^2 - \frac{\epsilon}{2}q, \tag{7}$$

where q is an even integer. Moreover, we choose q and $\delta = \delta(\epsilon)$ such that

$$\frac{1}{q} \ll \delta \ll 1 - \epsilon. \tag{8}$$

The following notations will play essential roles in the proofs of the coming two sections.

Definition 5.1. *Let $B = \{x \in V : |N(x) \cap S| \geq \delta q\}$ and $A = S_{q+1} \backslash B$. Let $\mathcal{R} = \{N(x) : x \in A\}$. We call any subset in V of size $q + 1$ a line.*

Now we are able to state the main result of this section.

Lemma 5.1. *There is a 1-intersecting $(q+1)$-hypergraph on the vertex set V, which contains $\mathcal{R}$ and at least q^2 lines.*

The full proof of this is involved, which we break into two subsections. However, Subsection 5.1 is concise and about 3-page long, which indicates that Lemma 5.1 holds for any real $\epsilon \in (0, 1/2)$.[2] Subsection 5.2 requires rather technical efforts to show Lemma 5.1 for all $\epsilon \in (0, 1)$. We like to mention that Subsection 5.2 is not necessary for the stability with a weaker bound such as in [24].

5.1 An initiatory bound

The proof of Lemma 5.1 will process by showing a sequence of claims. Before that, we first collect some basic properties on G. By Proposition 2.11 and Lemma 2.12, we have

$$q + 1 \leq |S| \leq \sum_{i=0}^{q} (i+1)|S_{q-i}| = f(V) = (q+1)n - 2e(G) \leq q + \epsilon q + 1 \tag{9}$$

and thus

$$q^2 - \epsilon q \leq |S_{q+1}| \leq q^2. \tag{10}$$

For any $T \subseteq S$, it holds that $q + \epsilon q + 1 \geq f(V) \geq f(T) + (|S| - |T|) \geq f(T) + (q + 1 - |T|)$. This implies that

$$f(T) \leq |T| + \epsilon q \quad \text{for any } T \subseteq S$$

and in particular, one can derive that

$$d(x) \geq (1 - \epsilon)q \quad \text{and} \quad d(x) + d(y) \geq (2 - \epsilon)q \quad \text{for any } x, y \in V. \tag{11}$$

We now prove the first claim by deriving some bounds on the sizes of A and B.

Claim 5.1. $|B| \leq \frac{2}{\delta}$ and $|A| \geq |S_{q+1}| - |B| \geq q^2 - \epsilon q - \frac{2}{\delta}$.

Proof. Let t be the number of adjacent ordered pairs (b, v) with $b \in B$ and $v \in S$. We have $|B| \cdot \delta q \leq t \leq |S| \cdot q \leq 2q^2$, implying that $|B| \leq 2q/\delta$. Now we consider the subgraph G_0 of G induced by the set $B \cup S$, where $|B \cup S| \leq 2q/\delta + 2q$. Since G_0 is C_4-free, we derive that $\frac{1}{2}|B|\delta q \leq e(G_0) \leq |B \cup S|^{3/2} = O(q^{3/2})$ and thus $|B| = O(\sqrt{q})$. For any $b, b' \in B$, we have $|N_S(b) \cap N_S(b')| \leq 1$. By (9) and the inclusion-exclusion principle, we can obtain

$$q + \epsilon q + 1 \geq |S| \geq |\cup_{b \in B} N_S(b)| \geq \sum_{b \in B} |N_S(b)| - \sum_{b, b' \in B} |N_S(b) \cap N_S(b')| \geq |B|\delta q - \binom{|B|}{2}.$$

Since $|B| = O(\sqrt{q})$, we can further derive that $|B| \leq \frac{2}{\delta}$. This finishes the proof. $\qquad \square$

Next we investigate properties on some special vertices of degree $q + 1$, defined as following. We remark that by Lemma 2.12, any vertex in S_{q+1} is adjacent to at least one vertex in S.

Definition 5.2. *A vertex $v \in V$ has **property 1**, if $v \in S_{q+1}$ satisfies that $|N(v) \cap S_{q+1}| = q$ and $|N(v) \cap S_q| = 1$. Let V_1 denote the set of all vertices of property 1 in G.*

Claim 5.2. $|V_1| \geq (1 - \epsilon)q^2 - (1 + 2\epsilon)q$.

---

[2]Note that Claim 5.8 yields a 1-intersecting $(q+1)$-hypergraph containing $\mathcal{R}$ and at least $q^2 + (1 - 2\epsilon)q + O(1)$ lines.

Proof. For $uv \in E(G)$ with $u \in S$ and $v \in S_{q+1}$, we assign a weight $w(uv)$ to be the deficiency $f(u)$. Let W denote the sum of the weights of these edges. We note that any vertex in V_1 contributes one to the sum W, while any vertex in $S_{q+1} \setminus V_1$ contributes at least two. Hence, by (9) we can derive that

$$|V_1| + 2(|S_{q+1}| - |V_1|) \leq W \leq \sum_{i=0}^{q} (q-i)(i+1)|S_{q-i}| \leq q \cdot f(V) \leq q(q + \epsilon q + 1).$$

Since $|S_{q+1}| \geq q^2 - \epsilon q$, we have $|V_1| \geq 2|S_{q+1}| - q(q + \epsilon q + 1) \geq (1 - \epsilon)q^2 - (1 + 2\epsilon)q$. □

The next claim describes the structure of the neighborhood of a vertex in V_1. Suppose $v \in V_1$ has $N(v) = \{v_1, ..., v_{q+1}\}$. Let

$$N_i = N(v_i) \setminus N[v] \quad \text{for} \quad i \in [q+1].$$

Claim 5.3. *For $v \in V_1$, the sets $N_1, ..., N_{q+1}$ form a partition of $V \setminus N[v]$, and $G[N(v)]$ consists of a matching of size $\frac{q}{2}$ plus an isolated vertex of degree q.*

Proof. Assume that the induced graph $G[N(v)]$ contains m edges, which clearly form a matching. Since G is C_4-free, any $x \in V \setminus N[v]$ has at most one neighbor in $N(v)$. By double-counting the number of edges between $N(v)$ and $V \setminus N(v)$, we have

$$(q^2 + q + 1) - (q + 2) + (q + 1) \geq \sum_{x \in N(v)} d(x) - 2m = q(q+1) + q - 2m, \tag{12}$$

implying that $m \geq \frac{q}{2}$. Since q is even, we derive that $m = \frac{q}{2}$ and moreover, (12) must be an equality. This further shows that $N_1, ..., N_{q+1}$ form a partition of $V \setminus N[v]$.

Suppose $d(v_{q+1}) = q$. It remains to show that v_{q+1} is an isolated vertex in $G[N(v)]$. Suppose for a contradiction that the edge set of $G[N(v)]$ is $\{v_2 v_3, ..., v_q v_{q+1}\}$. Then $|N_1| = q$, $|N_{q+1}| = q - 2$ and $|N_i| = q - 1$ for $2 \leq i \leq q$. Since G is C_4-free, every $G[N_i]$ contains at most $\lfloor |N_i|/2 \rfloor$ edges and there is no edge between N_{2i} and N_{2i+1} for $1 \leq i \leq \frac{q}{2}$. Also, there are at most $\min\{|N_i|, |N_j|\}$ edges between N_i and N_j for $i, j \in [q+1]$. Thus we have

$$\sum_{x \in N_2} d(x) \leq |N_2| + 2\lfloor |N_2|/2 \rfloor + \min\{|N_1|, |N_2|\} + \sum_{4 \leq i \leq q+1} \min\{|N_2|, |N_i|\} = (q+1)|N_2| - 2.$$

So $f(N_2) = (q+1)|N_2| - \sum_{x \in N_2} d(x) \geq 2$. Similarly, we have $f(N_i) \geq 2$ for all $2 \leq i \leq q-1$. Together with (9), we can obtained a contradiction as following

$$2q - 4 > q + \epsilon q + 1 \geq f(V) \geq \sum_{2 \leq i \leq q-1} f(N_i) \geq 2q - 4.$$

This completes the proof of Claim 5.3. □

The following is key for constructing a large $(q+1)$-uniform 1-interesting hypergraph.

Claim 5.4. *Suppose $v \in V_1$ has $N(v) = \{v_1, ..., v_{q+1}\}$. If $u \in S_{q+1} \setminus N[v]$ is adjacent to $S_{q+1} \cap N(v)$, then $|N(u) \cap N(v_i)| = 1$ for all $i \in [q+1]$.*

Proof. By Claim 5.3, we assume that $uv_1, v_1 v_2 \in E(G)$ for some $v_1, v_2 \in S_{q+1}$. Then u has exactly one neighbor in $N[v]$ and no neighbors in N_2. By Claim 5.3, $N_1, ..., N_{q+1}$ form a partition of $V \setminus N[v]$. Since $u \in S_{q+1}$ has at most one neighbor in each N_i for $i \neq 2$, it follows that u must have exactly one neighbor in each N_i for $i \neq 2$. Since $N(u) \cap N(v_2) = \{v_1\}$, we see that indeed $|N(u) \cap N(v_i)| = 1$ holds for all $i \in [q+1]$. □

We then show that the neighborhood of any vertex in A contains many vertices of property 1. To do so, for any $x \in A$ we define

$$S_x = N(x) \cap S \quad \text{and} \quad S_x^* = S_x \cup (N(S_x) \cap N(x)). \tag{13}$$

Since $x \in A$, we have $|S_x| \leq \delta q$. Every vertex in S_x has at most one neighbor in $N(x)$, so $|S_x^* \setminus S_x| \leq |S_x|$ and thus $|S_x^*| \leq 2|S_x| \leq 2\delta q$.

Claim 5.5. *For $x \in A$, there are at least $(1 - \epsilon - 3\delta)q + 1$ vertices of property 1 in $N(x) \backslash S_x^*$.*

Proof. Let $N(x) = \{x_1, ..., x_{q+1}\}$ and $N_i = N(x_i) \setminus N[x]$ for $i \in [q+1]$. We assert that $f(N_i) \geq 1$ for any $x_i \in N(x) \backslash S_x^*$. Indeed by definition, such $x_i \in S_{q+1}$ and every neighbor of x_i in S must lie outside of $N[x]$ (that is in N_i). Also by Lemma 2.12, x_i has at least one neighbor in S which belongs to N_i. So we have $f(N_i) \geq 1$. 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 $x_i \in V_1$. If we let m be the number of vertices of property 1 in $N(x) \backslash S_x^*$, then we have

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

Using $d(x) = q + 1$ and $2|S_x^*| - |S_x| \leq 3\delta q$, we can derive that $m \geq (1 - \epsilon - 3\delta)q + 1$. $\square$

Claim 5.6. *$\mathcal{R}$ is a 1-intersecting $(q + 1)$-hypergraph with $|\mathcal{R}| \geq q^2 - \epsilon q - 2/\delta$.*

Proof. It is clear that $\mathcal{R}$ is $(q+1)$-uniform and by Claim 5.1, $|\mathcal{R}| = |A| \geq q^2 - \epsilon q - 2/\delta$. So it is enough to show that $\mathcal{R}$ is 1-interesting. Suppose that there exist some $x, y \in A$ with no common neighbor. First consider the case $xy \in E(G)$. By Claim 5.5, there exists some $z \in N(x) \cap V_1 - \{y\}$. Clearly we have $yz \notin E(G)$. Applying Claim 5.4 (by viewing z as the vertex v therein), since $y \in S_{q+1} \backslash N[z]$ is adjacent to $x \in S_{q+1} \cap N(z)$, we can conclude that $|N(y) \cap N(x)| = 1$, a contradiction.

Now assume that $xy \notin E(G)$. 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 we have $y \in Y$. Since each x_i has at most one neighbor in $N(x)$, we get that

$$|Y| \leq n - (q + 2) - \sum_{i=1}^{q+1} (d(x_i) - 2) = \sum_{i=1}^{q+1} f(x_i).$$

Let $N_1(x)$ be the set of vertices in $N(x) \backslash S_x^*$ of property 1. By Claim 5.5, $|N_1(x)| \geq (1 - \epsilon - 3\delta)q + 1$. Further let $N_2(x) = N(x) \backslash (N_1(x) \cup S_x^*)$. Then, we have $f(N_i) = 1$ for each $x_i \in N_1(x)$ and $f(N_j) \geq 2$ for each $x_j \in N_2(x)$. Thus, we can derive that

$$|Y| \leq \sum_{i=1}^{q+1} f(x_i) = \sum_{x_i \in S_x} f(x_i) \leq q + \epsilon q + 1 - |N_1(x)| - 2|N_2(x)|.$$

Since $N(x) = N_1(x) \cup N_2(x) \cup S_x^*$, we see that the number of neighbors of y in those N_i's with $x_i \in N_1(x)$ is at least $d(y) - (|Y| - 1) - |S_x^*| - |N_2(x)|$, which is at least

$$(q + 2) - (q + \epsilon q + 1) + |N_1(x)| + |N_2(x)| - |S_x^*| \geq (1 - \epsilon - 4\delta)q + 2 \geq \delta q,$$

where we used the above estimation on $|Y|$ and the facts that $q + 1 = |N_1(x)| + |N_2(x)| + |S_x^*|$, $|S_x^*| \leq 2\delta q$ and $\delta \ll 1 - \epsilon$. Since $|N(y) \cap S| < \delta q$, among those neighbors of y, there is a vertex $z \in N(y) \cap S_{q+1}$. Suppose that $z \in N_j$ for some $x_j \in N_1(x) \subseteq V_1$. Applying Claim 5.4 (by viewing 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. Since G is C_4-free, x and y have exactly one common neighbor, finishing the proof of Claim 5.6. $\square$

Definition 5.3. *Let $V_1^* = V_1 \backslash N(B)$. For $v \in V_1^*$, let v' be the unique vertex in $N(v)$ of degree q. Let $\mathcal{L} = \{N[v'] : v \in V_1^*\}$.*

It is clear that $\mathcal{R} \cap \mathcal{L} = \emptyset$.

Claim 5.7. *For any $L \in \mathcal{L}$, there exist $R_1, ..., R_q \in \mathcal{R}$ such that $L \cup R_1 \cup ... \cup R_q = V$ and $|L \cap R_1 \cap ... \cap R_q| = 1$. Moreover, $|\mathcal{L}| \geq (1 - \epsilon)q - (2 + 2\epsilon + 2/\delta)$.*

Proof. Given $L \in \mathcal{L}$, there exists a vertex $v \in V_1^*$ with $N(v) = \{v_1, ..., v_{q+1}\}$ and $L = N[v_{q+1}]$. Since $N(v) \cap B = \emptyset$, we see that $v_1, ..., v_q \in A$ and thus $N(v_1), ..., N(v_q) \in \mathcal{R}$. By Claim 5.3, it implies that $L \cup N(v_1) \cup ... \cup N(v_q) = V$ and $L \cap N(v_1) \cap ... \cap N(v_q) = \{v\}$, as desired. It remains to show the lower bound of $|\mathcal{L}|$. By Claims 5.1 and 5.2, we have $|V_1^*| \geq |V_1| - |N(B)| \geq (1-\epsilon)q^2 - (1+2\epsilon)q - 2(q+1)/\delta$. It is obvious that each $L \in \mathcal{L}$ corresponds to a unique vertex $u \in S_q \backslash B$ with $L = N[u]$ (as otherwise it will force a C_4), however such a vertex u may be adjacent to at most q vertices in $V_1 \backslash N(B)$. Thus we have that $|\mathcal{L}| \geq \frac{|V_1^*|}{q} \geq \frac{1}{q} \left((1-\epsilon)q^2 - (1+2\epsilon)q - 2(q+1)/\delta\right) \geq (1-\epsilon)q - (2+2\epsilon+2/\delta)$. $\square$

Using the above claims, we now can construct a 1-intersecting $(q+1)$-hypergraph with considerably many edges.

Claim 5.8. $\mathcal{R} \cup \mathcal{L}$ *is a 1-intersecting $(q+1)$-hypergraph based on G with $|\mathcal{R} \cup \mathcal{L}| \geq q^2 + (1-2\epsilon)q - (2+2\epsilon+4/\delta)$.*

Proof. In view of Lemma 2.5, we see from Claims 5.6 and 5.7 that $\mathcal{R} \cup \mathcal{L}$ is a 1-intersecting $(q+1)$-hypergraph. Since $\mathcal{R} \cap \mathcal{L} = \emptyset$, we have

$$|\mathcal{R} \cup \mathcal{L}| \geq (q^2 - \epsilon q - 2/\delta) + (1-\epsilon)q - (2+2\epsilon+2/\delta) = q^2 + (1-2\epsilon)q - (2+2\epsilon+4/\delta),$$

completing the proof. $\square$

5.2 The completion of the proof of Lemma 5.1

Following the sequence of previous claims, we now continue and complete the proof of Lemma 5.1. By (10), we write

$$|S_{q+1}| = q^2 - \alpha q \text{ for some constant } 0 \leq \alpha \leq \epsilon. \tag{14}$$

Using $|V| = \sum_{i=0}^{q+1} |S_i|$ and $2e(G) = \sum_{i=0}^{q+1} i|S_i|$, by (7) and (14) we can conclude that

$$|S_q| \geq 2e(G) - (q-1)|V| - 2|S_{q+1}| \geq (2\alpha + 1 - \epsilon)q + 1. \tag{15}$$

Definition 5.4. *We say a vertex v has **property 2**, if $v \in S_{q+1} \backslash N(B)$ satisfies $|N(v) \cap S_{q+1}| = q-1$ and $|N(v) \cap S_q| = 2$. Let V_2 denote the set of all vertices of property 2 in G.*

Claim 5.9. *Any $u \in S_q \backslash (B \cup N(V_1^*))$ has at least $(1-\epsilon-3\delta)q - (1+6/\delta)$ neighbors in V_2.*

Proof. Let $S_u = N(u) \cap S$ and $S_u^* = S_u \cup (N(S_u) \cap N(u))$. Since $u \notin B$ and every vertex in S_u has at most one neighbor in $N(u)$, we have $|S_u^* \backslash S_u| \leq |S_u| \leq \delta q$ and $|S_u^*| \leq 2\delta q$.

Write $N(u) = \{u_1, ..., u_q\}$ and $N_i = N(u_i) \backslash N[u]$ for each $i \in [q]$. Since G is C_4-free, all the sets N_i are disjoint. Note that $N(u) \cap V_1^* = \emptyset$. So $N(u) \cap V_1 \subseteq N(u) \cap N(B)$. Since each vertex in B has at most one neighbor in $N(u)$, we have $|N(u) \cap V_1| \leq |N(u) \cap N(B)| \leq |B| \leq 2/\delta$ and thus $|N(u) \backslash V_1| \geq q - 2/\delta$. Each $u_i \in N(u) \backslash (V_1 \cup S_u^*)$ has degree $q+1$ and at least two neighbors in S, all of which (except u) are in $V \backslash N[u]$. Hence for such $u_i \in N(u) \backslash (V_1 \cup S_u^*)$, the deficiency $f(N_i)$ is at least 1 with equality if and only if $|N(u_i) \cap S_{q+1}| = q-1$ and $|N(u_i) \cap S_q| = 2$. Let M be the set of vertices $u_i \in N(u) \backslash (V_1 \cup S_u^*)$ with $f(N_i) = 1$. By counting the deficiency, we have

$$|M| + 2(|N(u) \backslash V_1| - |S_u^*| - |M|) + |S_u| \leq \sum_{i=1}^{q} f(N_i \cup \{u_i\}) \leq f(V) \leq q + \epsilon q + 1.$$

Using $|N(u) \backslash V_1| \geq q - 2/\delta$, we can derive that $|M| \geq (1-\epsilon-3\delta)q - (1+4/\delta)$. Notice that $M \backslash N(B) \subseteq V_2$. Since $M \subseteq N(u)$, each vertex in B has at most one neighbor in M and thus we have at least $|M \backslash N(B)| \geq |M| - |B| \geq (1-\epsilon-3\delta)q - (1+6/\delta)$ vertices in $N(u) \cap V_2$. $\square$

Claim 5.10. *For $v \in V_2$, $G[N(v)]$ consists of a matching of size $\frac{q}{2}$ plus an isolated vertex of degree q.*

Proof. We observe that every $v \in V_2$ is in A and every $x \in N(v) \cap S_{q+1}$ is also in A. By Claim 5.6, $|N(v) \cap N(x)| = 1$. Since v has $q - 1$ neighbors of degree $q + 1$ and q is even, it is easy to see that $G[N(v)]$ must contain exactly $\frac{q}{2}$ edges which form a matching and moreover, the only isolated vertex in $G[N(v)]$ has degree q. $\qquad\square$

In view of Claim 5.10, we now give some notations in relation to V_2 for later use.

Definition 5.5. *For $v \in V_2$, we write $N(v) = \{v_1, ..., v_{q+1}\}$ such that $d(v_1) = d(v_2) = q$ and the edge set of $G[N(v)]$ is $\{v_{2i}v_{2i+1} : 1 \le i \le \frac{q}{2}\}$. Let $N_i = N(v_i) \backslash N[v]$ for each i, and let the unique vertex not contained in $\cup_{1 \le i \le q+1} N[v_i]$ be $v^\star$.*

Note that the sets $N[v], N_1, ..., N_{q+1}$ and $\{v^\star\}$ form a partition of V. We recall that $\mathcal{R} = \{N(x) : x \in A\}$ is 1-intersecting and subsets in V of size $q + 1$ are called *lines*.

Definition 5.6. *For $v \in V_2$, we say v_1 and v_2 are the type-I vertex and type-II vertex for v, respectively. We also say that $L_I^v = N[v_1]$ is the type-I line for v and $L_{II}^v = N(v_2) \cup \{v^\star\}$ is the type-II line for v.*[3] *Furthermore, we say $v \in V_2$ is **extendable** if $\mathcal{R} \cup \{L_I^v, L_{II}^v\}$ is 1-intersecting; otherwise, we call it **non-extendable**.*

Claim 5.11. *For $v \in V_2$, we have $v_3, ..., v_{q+1} \in A$, $L_I^v \cup L_{II}^v \cup N(v_3) \cup ... \cup N(v_{q+1}) = V$ and $L_I^v \cap L_{II}^v \cap N(v_3) \cap ... \cap N(v_{q+1}) = \{v\}$. Moreover for each $N \in \mathcal{R}$, $|N \cap L_I^v| + |N \cap L_{II}^v| = 2$.*

Proof. It is clear that the first conclusion follows by definition. Consider any $v \in V_2$ and $N \in \mathcal{R}$. Note that $v_i \in A$ for $3 \le i \le q + 1$. If $N = N(v_i)$ for some $3 \le i \le q + 1$, then the second conclusion is clear. Otherwise, as $\mathcal{R}$ is 1-intersecting, we see that $|N \cap N(v_i)| = 1$ for $3 \le i \le q + 1$. This also infers that $|N \cap (L_I^v \cup L_{II}^v)| = 2$, completing the proof. $\qquad\square$

Therefore, if $v \in V_2$ is non-extendable, then there must be a line $N \in \mathcal{R}$ with $(|N \cap L_I^v|, |N \cap L_{II}^v|) \in \{(0, 2), (2, 0)\}$. This motivates the following definitions.

Definition 5.7. *If $v \in V_2$ is extendable, we say both vertices v_1, v_2 and both lines L_I^v, L_{II}^v are **good** (of type-I, type-II respectively) for v. Suppose $v \in V_2$ is non-extendable. If there exists $N \in \mathcal{R}$ with $|N \cap L_I^v| = 2$, then the type-I vertex v_1 is called **bad** for v. If there exists $N' \in \mathcal{R}$ with $|N' \cap L_{II}^v| = 2$, then the type-II vertex v_2 is called **bad** for v.*

We point out that any vertex of type-I or -II for some vertex in V_2 belongs to the set $S_q \backslash B$. Recall that $V_1^* = V_1 \backslash N(B)$ and $\mathcal{L} = \{N[u] : u \in S_q \cap N(V_1^*)\}$ from Definition 5.3.

Claim 5.12. *Let u be a vertex of type-I or -II for some $v \in V_2$. If $u \in N(V_1^*)$, then $N[u] \in \mathcal{L}$ and u is a good vertex of type-I for v.*

Proof. Suppose that u is adjacent to $w \in V_1 \backslash N(B)$. Then $N[u] \in \mathcal{L}$ and by Claim 5.8, $\mathcal{R} \cup \{N[u]\}$ is 1-intersecting. If u is a type-II vertex for v, then there exists some $u' \in N(v) \cap A$ adjacent to u. So $N(u') \in \mathcal{R}$, but $\{u, v\} \subseteq N(u') \cap N[u]$, a contradiction. Hence u is type-I for v. In particular, $L_I^v = N[u]$. By Claim 5.11, applying Lemma 2.5 with $\mathcal{H} = \mathcal{R} \cup \{N[u]\}$ and $\mathcal{F} = \{L_{II}^v\}$, we derive that $\mathcal{R} \cup \{L_I^v, L_{II}^v\}$ is 1-intersecting. This shows that v is extendable, finishing the proof. $\qquad\square$

We now investigate more properties for general $u \in S_q \backslash B$. It seems possible that u can be good or bad for different vertices in $V_2 \cap N(u)$. Nevertheless, we will show in the following claims that there are strong restrictions one can say for the goodness/badness.

Claim 5.13. *If $v \in V_2$ is non-extendable, then there always exists $N \in \mathcal{R}$ with $|N \cap L_I^v| = 2$.*

---

[3]We will also say that, for instance, v_1 is the vertex of type-I for v and L_{II}^v is the line of type-II for v.

Proof. Let $N_i = N(v_i) \backslash N[v]$ for $i \in [q + 1]$. Suppose not. Then by Claim 5.11, there exists some $N(c) \in \mathcal{R}$ with $|N(c) \cap L_{II}^v| = 2$. It is clear that $v^\star \in N(c)$. Since $c \in A$, by Claim 5.5 there are at least $(1 - \epsilon - 3\delta)q + 1$ many neighbors of c in $V_1 \backslash S_c^*$. As G is C_4-free, we can find at least $(1 - \epsilon - 3\delta)q - 2/\delta + 1$ vertices in $(V_1 \cap N(c)) \backslash N(B)$.

Now we take $b \in (V_1 \cap N(c)) \backslash (N(B) \cup N(v_1) \cup N(v_2) \cup N(v_3) \cup N(v) \cup N(v^\star))$. Suppose $N(b) \cap N(v) = \{v_i\}$ (by Claim 5.6 they intersect). By the choice of b, we may write $N(b) = \{b_1, ..., b_{q+1}\}$ with $b_1 = v_i$, where $d(b_1) = q + 1$ and $d(b_{q+1}) = q$. Since $e(N(b), N_2 \cup \{v_3\}) \le |N_2 \cup \{v_3\}| = q - 1$, there are at least two vertices in $N(b)$ with no neighbor in $N_2 \cup \{v_3\}$. So there exists $a \in N(b) \backslash \{v_i\} = N(b) \backslash N(v)$ with $N(a) \cap (N_2 \cup \{v_3\}) = \emptyset$.

We assert that $a \in N_1 \cup ... \cup N_{q+1}$ and $N(a) \cap L_{II}^v = \emptyset$. First, we see that $av \notin E(G)$ (as otherwise, $a \in N(v) \cap N(b) = \{v_i\}$). It then suffices to show $v^\star \notin N(a)$. To see this, suppose $v^\star \in N(a)$ and then one can find a C_4, namely $abcv^\star a$ in G, a contradiction.

We also assert that it has to be $d(a) = q$. It is clear that $d(a) \in \{q, q + 1\}$. Suppose for a contradiction that $d(a) = q + 1$. Since $b \in V_1 \backslash N(B)$, we have $a \in A$ and thus $N(a) \in \mathcal{R}$. Since $v_3, ..., v_{q+1} \in A$, by Claim 5.6 we have $|N(a) \cap N(v_\ell)| = 1$ for $3 \le \ell \le q + 1$. Then by Claim 5.11, $|N(a) \cap L_I^v| = 2$, contradicting our assumption.

We now further show that $a \in N_1$. Assume for a contradiction that $av_1 \notin E(G)$. As $av_3 \notin E(G)$, we have $a \in N_2 \cup N_4 \cup ... \cup N_{q+1}$. Suppose $a \in N_2$. By Claim 5.3, we see that $\{b_1, ..., b_q\}$ induces a matching in $G[N(b)]$; then at least two of $\{b_1, ..., b_q\}$ have no neighbor in $N_2 \cup \{v_3\}$. Hence we can choose $a' \in \{b_1, ..., b_q\} \backslash \{v_i\}$ with $N(a') \cap (N_2 \cup \{v_3\}) = \emptyset$. But such a' has degree $q + 1$, contradicting the previous assertion. Hence we may assume $a \in N_j$ for some $j \in \{4, ..., q + 1\}$. Since $b \in V_1 \backslash N(B)$ and $a \in N(b) \cap S_q$, by Claim 5.8, $\mathcal{R} \cup \{N[a]\}$ is 1-intersecting. Because $v_\ell \in A$ for $\ell \in \{3, 4, ..., q + 1\}$, we have $|N[a] \cap N(v_\ell)| = 1$. This, together with Claim 5.11 and the fact $N[a] \cap L_{II}^v = \emptyset$, imply that $|N[a] \cap L_I^v| \ge 2$. As $av_1 \notin E(G)$, we have $|N(a) \cap N(v_1)| \ge 2$, a contradiction as it would force a C_4.

Lastly, we observe that for every choice of such b, the above vertex a, which lies in $N_1 \cap S_q \subseteq N(v_1) \cap S_q$, must be distinct. This is because if there exist two vertices say b_1, b_2 corresponding to the same vertex a, then it provides a C_4 such as $b_1 a b_2 c b_1$ in G. There are at least $(1 - \epsilon - 3\delta)q - 2/\delta - 4$ choices for b, implying that $|N(v_1) \cap S_q| \ge (1 - \epsilon - 3\delta)q - 2/\delta - 4 \ge \delta q$, where the last inequality holds as $1/q \ll \delta \ll 1 - \epsilon$. This shows $v_1 \in B$, a contradiction to that $v \in S_{q+1} \backslash N(B)$, completing the proof. $\square$

Claim 5.14. *For any $u \in S_q \backslash B$, the number of non-extendable vertices $v \in V_2 \cap N(u)$ with u as the type-II vertex is at most $4\delta q$.*

Proof. Suppose on the contrary that there exists $u \in S_q \backslash B$ such that the set $T = \{$non-extendable $v \in V_2 \cap N(u) : u$ is the type-II vertex of $v\}$ has size $t > 4\delta q$. We write $T = \{v_i : i \in [t]\}$ and for each $v_i \in T$, we denote the type-I vertex of v_i by u_i. It is clear that all u_i are distinct (as otherwise it would force a C_4).

By Claim 5.13, for v_1 there exists $L = N(a) \in \mathcal{R}$ such that $|L \cap L_I^{v_1}| = 2$. By Claim 5.11, $L \cap N(u) = \emptyset$. This shows that exactly one of the following facts holds for L and any other v_i's in T:

(1). If $|L \cap L_I^{v_i}| = |L \cap L_{II}^{v_i}| = 1$, then $v_i^\star \in L$ (because $L_{II}^{v_i} = N(u) \cup \{v_i^\star\}$ and $L \cap N(u) = \emptyset$);

(2). Otherwise $|L \cap L_I^{v_i}| = 2$, then $u_i \in L$ (because $L_I^{v_i} = N[u_i]$ and $|L \cap N(u_i)| \le 1$).

Let T_1 consist of all $v_i \in T$ with $|L \cap L_I^{v_i}| = 2$. By (2), we see $u_i \in L = N(a)$ for all $v_i \in T_1$. Since $a \in A$ and $u_i \in S$ are distinct, we have $|T_1| \le |N(a) \cap S| < \delta q$. Now let $L' = N(a') \in \mathcal{R}$ be another line, other than L, such that $|L' \cap L_I^{v_j}| = 2$ for some $v_j \in T \backslash T_1$. Let T_2 consist of all $v_j \in T \backslash T_1$ with $|L' \cap L_I^{v_j}| = 2$. Similarly, we also have $|T_2| < \delta q$.

We also assert that there are at most δq vertices $v_i \in T$ sharing a common $v_i^\star$ (denoted by $v^\star$). Suppose $v_{j_1}, ..., v_{j_s} \in T$ share a common $v^\star$ and subject to this, s is maximum. We first find a line L_0 with $|L_0 \cap L_I^{v_{j_1}}| = 2$, which also satisfies $L_0 \cap (N(u) \cup \{v^\star\}) = \emptyset$. If L_0 satisfies that $|L_0 \cap L_I^{v_{j_i}}| = 2$ for all $i \in [s]$, then we can get that $s < \delta q$ as above. So we may assume that there is some v_{j_i} such

that (1) holds. By (1), we then derive that $v^\star \in L_0$, a contradiction to that $L_0 \cap (N(u) \cup \{v^\star\}) = \emptyset$. This proves the assertion.

By our choice, (1) holds for each of $L = N(a)$ and $L' = N(a')$ and for any $v_i \in T \backslash (T_1 \cup T_2)$. Since $|T \backslash (T_1 \cup T_2)| > 2\delta q$ and at most δq vertices $v_i \in T \backslash (T_1 \cup T_2)$ share a common $v_i^\star$, we can find two different $v_i^\star$, say x and y. By (1), we have $x, y \in N(a) \cap N(a')$. This forces a C_4 in G and finishes the proof. $\qquad\square$

The next claim shows that the type of a good vertex $u \in S_q \backslash B$ in fact is an invariance (that is, the type remains the same for all extendable vertices in $N(u) \cap V_2$).

Claim 5.15. *Let $u \in S_q \backslash B$. If u is a good vertex of type-I for some vertex in V_2, then u is the good vertex of type-I for all vertices in $N(u) \cap V_2$.*

If u is a good vertex of type-II for some vertex in V_2, then u is a vertex of type-I (which must be bad) for at most one vertex in $N(u) \cap V_2$; moreover, there are at most two good lines of type-II for all the vertices in $N(u) \cap V_2$.

Proof. First let us assume that u is a good vertex of type-I for some $v \in V_2$. Since v is extendable, $\mathcal{R} \cup \{N[u]\}$ is 1-intersecting. Consider any $v' \in N(u) \cap V_2 \backslash \{v\}$. Suppose that u is a type-II vertex for v'. If we let x be the unique vertex in $N(v')$ adjacent to u, then we see $N(x) \in \mathcal{R}$ and $N(x) \cap N[u] = \{u, v'\}$, a contradiction to that $\mathcal{R} \cup \{N[u]\}$ is 1-intersecting. Thus u is always type-I for all vertices in $N(u) \cap V_2$. Let $N(v') = \{u, a_2, ..., a_{q+1}\}$, where a_2 is type-II for v' and $a_3, ..., a_{q+1} \in A$. Then Claim 5.11 holds for v' analogously, where $N(a_3), ..., N(a_{q+1})$ and $L_I^{v'} = N[u]$ are q lines in the 1-intersecting hypergraph $\mathcal{R} \cup \{N[u]\}$. By Lemma 2.5, we infer that $\mathcal{R} \cup \{L_I^{v'}, L_{II}^{v'}\}$ is 1-intersecting and thus v' is also extendable. This shows that u is a good vertex of type-I for all vertices in $N(u) \cap V_2$.

Now we assume that u is a good vertex of type-II vertex for $v_1 \in V_2$. Let $F = N(u) \cup \{v_1^\star\}$. Since v_1 is extendable, $\mathcal{R} \cup \{F\}$ is 1-intersecting. Suppose that $v_2 \in N(u) \cap V_2$ is distinct from v_1 and u is a type-I vertex for v_2. By the first assertion, we see v_2 is non-extendable and u is bad for v_2. So $N(v_2) \in \mathcal{R}$ and $N(v_2) \cap N(u) = \emptyset$. Since $\mathcal{R} \cup \{F\}$ is 1-intersecting, the only possibility is that $v_1^\star \in N(v_2)$. Thus we have $u, v_1^\star \in N(v_2)$. Now suppose there exists another $v_3 \in N(u) \cap V_2 \backslash \{v_1, v_2\}$ which has u as its type-I vertex. Similarly we have $u, v_1^\star \in N(v_3)$, giving $uv_2 v_1^\star v_3 u$ as a C_4 in G, a contradiction. Hence by our discussion, u is a vertex of type-I for at most one vertex of $N(u) \cap V_2$.

By Claim 5.12, $u \in S_q \backslash (B \cup N(V_1^*))$. Then the previous paragraph, together with Claims 5.9 and 5.14, show that u appears as a good type-II vertex for at least $(1 - \epsilon - 3\delta)q - (1 + 6/\delta) - 1 - 4\delta q \geq 20$ vertices $w_i \in N(u) \cap V_2$. Suppose that there exist at least three good lines of type-II for these w_i's, say $F_1 = N(u) \cup \{w_1^\star\}$, $F_2 = N(u) \cup \{w_2^\star\}$ and $F_3 = N(u) \cup \{w_3^\star\}$. By renaming notations if necessary, we may assume that there are at least four $w_i \in N(u) \cap V_2$ whose $w_i^\star \notin \{w_1^\star, w_2^\star\}$. We observe that the type-I vertices u_i of these w_i are all distinct; indeed, otherwise say w_i and w_j have the same type-I vertex u', then it yields a 4-cycle $w_i u' w_j u w_i$, a contradiction. Hence, we can further find two of these w_i, say $w_4, w_5 \in N(u) \cap V_2$, such that their type-I vertices u_4, u_5 are distinct and not in $\{w_1^\star, w_2^\star\}$. Now we show $w_1^\star \in N(u_4)$ by considering the location of $w_1^\star$ in the local structure based on w_4. As $w_4 \in V_2$, any $v \in N(w_4) \cap S_{q+1}$ is in A, so w_4 is the unique vertex in $N(v) \cap F_1$ and thus $w_1^\star \notin N(v)$. Since $V = \left(\cup_{v \in N(w_4) \cap S_{q+1}} N(v) \right) \cup L_I^{w_4} \cup L_{II}^{w_4}$ and $w_1^\star \notin L_{II}^{w_4} = N(u) \cup \{w_4^\star\}$, we conclude that $w_1^\star \in L_I^{w_4} = N[u_4]$, which further shows $w_1^\star \in N(u_4)$, as wanted. Analogously, we can derive that $w_1^\star, w_2^\star \in N(u_4) \cap N(u_5)$. This provides a C_4 in G, a contradiction. $\qquad\square$

Claim 5.15 also indicates that the type of a good vertex $u \in S_q \backslash B$ is consistent with the type of all good lines containing $N(u)$. Consequently, just the same as a good vertex, a good line can only be of one particular type. To further study properties of good lines, we define an auxiliary graph as following.

Definition 5.8. *We denote $\mathcal{F}$ by the set of all good lines (of type-I or type-II for any vertex in V_2). Let $\mathcal{G}$ be the graph with vertex set $\mathcal{F}$, where $F, F' \in \mathcal{F}$ are adjacent if and only if they are the type-I and type-II lines for some extendable vertex in V_2.*

By the above discussion, we see that $\mathcal{G}$ is a bipartite graph on two parts $(\mathcal{F}_I, \mathcal{F}_{II})$, where $\mathcal{F}_I$ consists of all good lines of type-I and $\mathcal{F}_{II}$ consists of all good lines of type-II.

Claim 5.16. *For two independent edges* (L_I^v, L_{II}^v) *and* (L_I^w, L_{II}^w) *in* $\mathcal{G}$*, we have* $\overline{vw} = (|L_I^v \cap L_I^w|, |L_I^v \cap L_{II}^w|, |L_{II}^v \cap L_I^w|, |L_{II}^v \cap L_{II}^w|) \in \{(1,1,1,1), (2,0,0,2), (0,2,2,0), (0,2,2,q)\}$.

Proof. Clearly $v, w \in V_2$ are extendable. We define $v_1, ..., v_{q+1}, v^\star$ and $w_1, ..., w_{q+1}, w^\star$ according to Definition 5.5 for v and w, respectively.

First we observe that if any entry in the vector $\overline{vw}$ is one, then by applications of Lemma 2.5, we can infer that $\mathcal{R} \cup \{L_I^v, L_{II}^v, L_I^w, L_{II}^w\}$ is 1-intersecting and thus $\overline{vw} = (1,1,1,1)$. Hence, we may assume that none of the entries in $\overline{vw}$ is one.

Suppose $v \in L_I^w$. Then we have $v \in N(w_1)$ and thus $w_1 \in \{v_1, v_2\}$. If $w_1 = v_1$, then $L_I^w = L_I^v$, a contradiction as the two edges in $\mathcal{G}$ are independent. So $w_1 = v_2$. Then $\{v_2, v_3\} \subseteq N(v) \cap L_I^w$, however $\{N(v), L_I^w\} \subseteq \mathcal{R} \cup \{L_I^w\}$ is 1-intersecting, a contradiction. Hence we must have $v \notin L_I^w$. Also we have $|L_I^w \cap N(v_i)| = 1$ for $3 \le i \le q+1$. So $|L_I^w \cap (L_I^v \cup L_{II}^v)| = 2$. By symmetry between v and w, we also can get $|L_I^v \cap (L_I^w \cup L_{II}^w)| = 2$.

If $|L_I^v \cap L_I^w| = 2$, then $|L_I^v \cap L_{II}^w| = |L_{II}^v \cap L_I^w| = 0$. Note that $L_I^v, L_{II}^v, N(v_3), ..., N(v_{q+1})$ form a sun-flower with center v and vertex set V. Since $L_{II}^w \cap L_I^v = \emptyset$ and $|L_{II}^w \cap N(v_i)| = 1$ for $3 \le i \le q+1$, we derive that $|L_{II}^w \cap L_{II}^v| = 2$, i.e., $\overline{vw} = (2,0,0,2)$.

It remains to consider $|L_I^v \cap L_I^w| = 0$. Then $|L_I^v \cap L_{II}^w| = |L_{II}^v \cap L_I^w| = 2$. If $v \notin L_{II}^w$, using the same argument as above, we can easily get $\overline{vw} = (0,2,2,0)$. Hence we assume $v \in L_{II}^w$. If $v = w^\star$, then clearly $N(v) \cap N(w) = \emptyset$, contradicting that $\mathcal{R}$ is 1-intersecting. So $v \in N(w_2)$, implying that $w_2 \in \{v_1, v_2\}$. If $w_2 = v_1$, as $N(w_2) \subseteq L_{II}^w$ and $N(v_1) \subseteq L_I^v$, we have a contradiction that $2 = |L_{II}^w \cap L_I^v| \ge q$. So $w_2 = v_2$, which shows either $\overline{vw} = (0,2,2,q)$ or $L_{II}^w = L_{II}^v$ (a contradiction). $\square$

We say two good lines $F, F' \in \mathcal{F}$ are *friendly* if $|F \cap F'| = 1$ and two components $\mathcal{D}$ and $\mathcal{D}'$ of $\mathcal{G}$ are *friendly* if there is a friendly pair $\{F, F'\}$ with $F \in V(\mathcal{D})$ and $F' \in V(\mathcal{D}')$.

Claim 5.17. *If two components* $\mathcal{D}$ *and* $\mathcal{D}'$ *of* $\mathcal{G}$ *are friendly (possibly* $\mathcal{D} = \mathcal{D}'$*), then all pairs in* $\mathcal{D} \cup \mathcal{D}'$ *are friendly. In particular, any two vertices in the same component of* $\mathcal{G}$ *are friendly.*

Proof. First we point out that if two edges say (L_1, L_2) and (L_1, L_3) in $\mathcal{G}$ share a common vertex, then using Lemma 2.5, one can derive that $\mathcal{R} \cup \{L_1, L_2, L_3\}$ is 1-intersecting.

Assume that there exist $F \in V(\mathcal{D})$ and $F' \in V(\mathcal{D}')$ with $|F \cap F'| = 1$. To complete the proof, it suffices to show that for any vertex $L \in V(\mathcal{D})$, we have $|L \cap F'| = 1$ (unless $\mathcal{D} = \mathcal{D}'$ and $L = F'$). We will prove this by induction on the length d_L of the shortest path between L and F in $\mathcal{D}$. If $d_L = 0$, then $L = F$ and clearly it is true. Now suppose that for any $L^* \in V(\mathcal{D})$ with $d_{L^*} \le k$, the above statement holds. Consider any $L \in V(\mathcal{D})$ with $d_L = k+1$. Then there exists an edge (L, L^*) in $\mathcal{D}$ with $d_{L^*} = k$. By induction, we have $|L^* \cap F'| = 1$, unless $\mathcal{D} = \mathcal{D}'$ and $L^* = F'$. In the latter case, obviously we have $|L \cap F'| = 1$. Therefore $|L^* \cap F'| = 1$. Fix an edge incident to F', say (L', F') in $\mathcal{D}'$. If (L, L^*) and (L', F') share a common vertex, then $\mathcal{D} = \mathcal{D}'$ and by the first paragraph, it is easy to see that either $L = F'$ or $|L \cap F'| = 1$. So we may assume (L, L^*) and (L', F') are two independent edges. By Claim 5.16, as $|L^* \cap F'| = 1$, it infers that $|L \cap F'| = 1$. This finishes the proof. $\square$

Definition 5.9. *For each* $F \in \mathcal{F}$*, let* u_F *be the unique vertex in* $S_q \setminus B$ *satisfying* $N(u_F) \subseteq F$*. We say a component of* $\mathcal{G}$ *is* **rich** *if it contains some vertex* F *with* $u_F \in N(V_1^*)$*.*

We remark that for each $F \in \mathcal{F}$, there exists some extendable $v_F \in V_2$ such that u_F is a good vertex and F is a good line for v_F of the same type.

Claim 5.18. *A rich component of* $\mathcal{G}$ *is friendly with any component of* $\mathcal{G}$*.*

Proof. Let $\mathcal{D}$ be a rich component of $\mathcal{G}$ and let $F \in \mathcal{F}$ be a vertex in $\mathcal{D}$ with $u_F \in N(V_1^*)$. Then u_F is a good vertex and F is a good line for some $v \in V_2$ of the same type. By Claim 5.12, u_F is type-I for v and $F = N[u_F] \in \mathcal{L}$. Now consider any component $\mathcal{D}'$ of $\mathcal{G}$ and take any vertex $L \in V(\mathcal{D}') \cap \mathcal{F}$. Then $\mathcal{R} \cup \{L\}$ is 1-intersecting. Recall that Claim 5.7 holds for $F \in \mathcal{L}$. By Lemma 2.5, we can derive that $\mathcal{R} \cup \{L, F\}$ is 1-intersecting. This shows that $\mathcal{D}$ and $\mathcal{D}'$ are friendly. $\square$

Definition 5.10. *If a component of $\mathcal{G}$, which is not rich, contains at least 40 vertices in $\mathcal{F}_I$ and at least δq vertices in $\mathcal{F}_{II}$, then we say it is* **heavy**.

Claim 5.19. *Let $\mathcal{F}_r$ and $\mathcal{F}_h$ denote two unions of vertices in rich components and heavy components of $\mathcal{G}$, respectively. Then $|\mathcal{F}_r \cup \mathcal{F}_h \cup \mathcal{L}| \geq 2\alpha q + \frac{1}{2}(1-\epsilon)q$.*

Proof. We note that $\mathcal{L} = \{N[u] : u \in S_q \cap N(V_1^*)\}$ and $|\mathcal{L}| = |S_q \cap N(V_1^*)|$.

For $u \in S_q \backslash (B \cup N(V_1^*))$, if all vertices in $N(u) \cap V_2$ are non-extendable, then we say u is *poor*. Let P be the set of all poor vertices in $S_q \backslash (B \cup N(V_1^*))$. Consider the number M of pairs (v, u), where $v \in V_2$ is non-extendable and $u \in S_q \backslash (B \cup N(V_1^*))$ is the type-I vertex of v. Clearly M is at most the number of pairs (v, u'), where $v \in V_2$ is non-extendable and $u' \in S_q \backslash B$ is the type-II vertex of v. By Claims 5.9 and 5.14, $|P| \cdot ((1 - \epsilon - 3\delta)q - (1 + 6/\delta) - 4\delta q) \leq |M| \leq |S_q \backslash B| \cdot 4\delta q$. This implies that

$$|P| \leq \frac{4\delta q \cdot |S_q \backslash B|}{(1 - \epsilon - 3\delta)q - (1 + 6/\delta) - 4\delta q} = \frac{4\delta q \cdot |S_q \backslash B|}{(1 - \epsilon - 7\delta)q - (1 + 6/\delta)}. \tag{16}$$

We point out that any $u \in S_q \backslash (B \cup N(V_1^*) \cup P)$ has an extendable neighbor in V_2, which implies at least one good line (not in $\mathcal{L}$) containing $N(u)$. By Claim 5.15, if u is type-I, then the unique good line containing $N(u)$ is $N[u]$. If u is type-II, then there are at most two good lines containing $N(u)$, say F_1^u and F_2^u. We call these good lines as the *associated* lines of $u \in S_q \backslash (B \cup N(V_1^*) \cup P)$.

Let F be any line in $\mathcal{F}_I$ with $u_F \notin N(V_1^*)$. Clearly $u_F \in S_q \backslash (B \cup N(V_1^*))$ is type-I and $F = N[u_F]$. By Claims 5.9 and 5.15, u_F is the good vertex of type-I for at least $(1 - \epsilon - 3\delta)q - (1 + 6/\delta) \geq \delta q$ vertices in $N(u_F) \cap V_2$. This shows that there are at least δq good lines of type-II adjacent to F in $\mathcal{G}$, i.e., F has degree at least δq in $\mathcal{G}$.

Consider a type-II vertex $u \in S_q \backslash (B \cup N(V_1^*) \cup P)$. As mentioned above, there are at least one and at most two associated lines say F_1^u and F_2^u of u. By Claims 5.9, 5.14 and 5.15, we see that u is the good vertex of type-II for at least $(1 - \epsilon - 3\delta)q - (1 + 6/\delta) - 4\delta q - 1 \geq 80$ vertices in $N(u) \cap V_2$. Thus at least one of F_1^u and F_2^u has at least 40 neighbors in $\mathcal{F}_I$. By the previous paragraph, we see that at least one associated line of u is contained in a rich or heavy component of $\mathcal{G}$.

Now we show that the number of type-I vertices $u \in S_q \backslash (B \cup N(V_1^*) \cup P)$, which has no associated lines in any rich or heavy component of $\mathcal{G}$, is at most $\frac{40|S_q \backslash B|}{(1-\epsilon-3\delta)q-(1+6/\delta)}$. If a component $\mathcal{D}$ of $\mathcal{G}$ is neither rich nor heavy, then each $F \in V(\mathcal{D})$ has $u_F \notin N(V_1^*)$ and it has no more than 40 vertices in $\mathcal{F}_I$, implying that each $F \in V(\mathcal{D}) \cap \mathcal{F}_{II}$ has degree at most 40. Let e^*, ℓ_1 and ℓ_2 denote the numbers of edges, vertices of $\mathcal{F}_I$ and vertices of $\mathcal{F}_{II}$ contained in all components of $\mathcal{G}$ which are neither rich nor heavy, respectively. Then we have $\ell_1 \cdot ((1 - \epsilon - 3\delta)q - (1 + 6/\delta)) \leq e^* \leq 40 \cdot \ell_2 \leq 40|S_q \backslash B|$. This implies what we want.

Therefore we have $|(\mathcal{F}_r \cup \mathcal{F}_h) \backslash \mathcal{L}| \geq |S_q \backslash (B \cup N(V_1^*) \cup P)| - \frac{40|S_q \backslash B|}{(1-\epsilon-3\delta)q-(1+6/\delta)}$. This, together with (16) and $|\mathcal{L}| = |S_q \cap N(V_1^*)|$, imply that

$$|\mathcal{F}_r \cup \mathcal{F}_h \cup \mathcal{L}| \geq |S_q \backslash (B \cup N(V_1^*) \cup P)| - \frac{40|S_q \backslash B|}{(1 - \epsilon - 3\delta)q - (1 + 6/\delta)} + |S_q \cap N(V_1^*)|$$

$$\geq |S_q \backslash B| - \frac{4\delta q \cdot |S_q \backslash B|}{(1 - \epsilon - 7\delta)q - (1 + 6/\delta)} - \frac{40|S_q \backslash B|}{(1 - \epsilon - 3\delta)q - (1 + 6/\delta)}.$$

By (8) and (15), it shows that $|\mathcal{F}_r \cup \mathcal{F}_h \cup \mathcal{L}| \geq 2\alpha q + \frac{1}{2}(1 - \epsilon)q$, completing the proof. $\square$

Claim 5.20. *At least half of the lines in $\mathcal{F}_h$ are pair-wise friendly.*

Proof. We define H to be a graph whose vertices are heavy components of $\mathcal{G}$, where components $\mathcal{C}, \mathcal{D} \in V(H)$ are adjacent if and only if they are not friendly. To prove this, we may assume that any component has a non-friendly component in H (as otherwise, we can delete it and process). By Claim 5.17, it suffices to show that H is a bipartite graph.

Let $\mathcal{C}$ and $\mathcal{D}$ be two heavy components of $\mathcal{G}$ which are not friendly. Take a type-I line $F \in V(\mathcal{D})$. By Claim 5.16, for any edge $e = (L_1, L_2)$ in $\mathcal{C}$, $c(e) := (|F \cap L_1|, |F \cap L_2|)$ is either $(0, 2)$ or $(2, 0)$.

Repeatedly applying this, we see that all edges e in $\mathcal{C}$ have the same $c(e)$. So for any line $F \in V(\mathcal{D}) \cap \mathcal{F}_I$, exactly one of the following holds:

(A). All type-I lines L in $\mathcal{C}$ satisfy $|F \cap L| = 2$. Since $|N(u_F) \cap N(u_L)| \leq 1$, we have $u_L u_F \in E(G)$.

(B). All type-II lines L in $\mathcal{C}$ satisfy $|F \cap L| = 2$. Suppose L is a type-II line for $v_L \in V_2$. Then $L = N(u_L) \cup \{v_L^\star\}$ and we have either $v_L^\star = u_F$, $v_L^\star u_F \in E(G)$, or $u_L u_F \in E(G)$.

Suppose (A) holds for some $F \in V(\mathcal{D}) \cap \mathcal{F}_I$. Repeatedly applying the above conclusion, one would derive that in fact any $L \in V(\mathcal{C}) \cap \mathcal{F}_I$ and any $R \in V(\mathcal{D}) \cap \mathcal{F}_I$ satisfy $|L \cap R| = 2$ and $u_L u_R \in E(G)$. Since all u_L, u_R are distinct, it is easy to force a C_4 in G, a contradiction.

Hence (B) holds for all type-I lines $F \in V(\mathcal{D})$ and any heavy component $\mathcal{C}$ which is not friendly with $\mathcal{D}$. For every such F, we can partition $V(\mathcal{C}) \cap \mathcal{F}_{II}$ into three sets $X(F), Y(F)$ and $Z(F)$, where $X(F) = \{L \in V(\mathcal{C}) \cap \mathcal{F}_{II} : v_L^\star = u_F\}$, $Y(F) = \{L \in V(\mathcal{C}) \cap \mathcal{F}_{II} : v_L^\star u_F \in E(G)\}$ and $Z(F) = \{L \in V(\mathcal{C}) \cap \mathcal{F}_{II} : u_L u_F \in E(G)\}$. Let $F, F' \in V(\mathcal{D}) \cap \mathcal{F}_I$ be distinct. Then it is easy to see that $|X(F) \cap X(F')| = 0$ and $|Z(F) \cap Z(F')| \leq 1$. If there are $L_1, L_2 \in Y(F) \cap Y(F')$, then both $v_{L_1}^\star$ and $v_{L_2}^\star$ are adjacent to $\{u_F, u_{F'}\}$. This shows that $v_{L_1}^\star = v_{L_2}^\star$, as otherwise there is a C_4 in G. That is, all lines $L \in Y(F) \cap Y(F')$ have the same $v_L^\star$. By $f_{\mathcal{C}}(w)$, we denote the number of all lines $L \in V(\mathcal{C}) \cap \mathcal{F}_{II}$ with $v_L^\star = w$.

We first assert that there exists some vertex w with $f_{\mathcal{C}}(w) \geq (|V(\mathcal{C}) \cap \mathcal{F}_{II}| - 6)/6$. To see this, let us take four lines $F_1, F_2, F_3, F_4 \in V(\mathcal{D}) \cap \mathcal{F}_I$. We have $|V(\mathcal{C}) \cap \mathcal{F}_{II}| \geq |\cup_{1 \leq i \leq 4} X(F_i)| = \sum_{1 \leq i \leq 4} |X(F_i)|$, and by inclusion-exclusion,

$$|V(\mathcal{C}) \cap \mathcal{F}_{II}| \geq |\cup_{1 \leq i \leq 4} Y(F_i)| \geq \sum_{1 \leq i \leq 4} |Y(F_i)| - \sum_{1 \leq i < j \leq 4} |Y(F_i) \cap Y(F_j)|$$

and similarly, $|V(\mathcal{C}) \cap \mathcal{F}_{II}| \geq \sum_{1 \leq i \leq 4} |Z(F_i)| - 6$. Summing up the above three inequalities, using $|X(F_i)| + |Y(F_i)| + |Z(F_i)| = |V(\mathcal{C}) \cap \mathcal{F}_{II}|$ for each $i \in [4]$, we obtain that

$$3|V(\mathcal{C}) \cap \mathcal{F}_{II}| \geq 4|V(\mathcal{C}) \cap \mathcal{F}_{II}| - \sum_{1 \leq i < j \leq 4} |Y(F_i) \cap Y(F_j)| - 6.$$

Therefore one of $Y(F_i) \cap Y(F_j)$ contains at least $(|V(\mathcal{C}) \cap \mathcal{F}_{II}| - 6)/6$ lines L, all of which have the same $v_L^\star$. This proves our assertion.

We further assert that in fact for each heavy component $\mathcal{C}$, there exists a unique vertex $w_{\mathcal{C}}$ with $f_{\mathcal{C}}(w_{\mathcal{C}}) \geq 0.6 \cdot |V(\mathcal{C}) \cap \mathcal{F}_{II}|$. Let $c = |V(\mathcal{C}) \cap \mathcal{F}_{II}|$, which is at least δq. We choose w such that $f_{\mathcal{C}}(w)$ is maximum. So we have $f_{\mathcal{C}}(w) \geq (c - 6)/6$. Take any 20 lines say $F_1, ..., F_{20}$ in $V(\mathcal{D}) \cap \mathcal{F}_I$. Then we have $20c = \sum_{1 \leq i \leq 20} (|X(F_i)| + |Y(F_i)| + |Z(F_i)|) \leq \sum_{1 \leq i \leq 20} |Y(F_i)| + 2c + \binom{20}{2}$, implying that $\sum_{1 \leq i \leq 20} |Y(F_i)| \geq 18c - \binom{20}{2}$. Without loss of generality we may assume $w u_{F_i} \in E(G)$ if and only if $i \in [s]$ for some $s \in [20]$. Note that for $s < i \leq 20$, we have $|Y(F_i)| \leq c - f_c(w)$. If $s \leq 10$, then $\sum_{1 \leq i \leq 20} |Y(F_i)| \leq 20c - (20 - s) \cdot f_c(w) \leq 20c - 10(c - 6)/6 < 18c - \binom{20}{2}$, a contradiction. Hence $s \geq 11$. For $i \in [s]$, let $A_i = N_G(u_{F_i}) \setminus \{w\}$ and let $f_i = \sum_{x \in A_i} f_{\mathcal{C}}(x)$. We claim that A_i's are disjoint over $i \in [s]$; as otherwise say $x \in A_1 \cap A_2$, then both u_{F_1}, u_{F_2} are adjacent to w, x in G, a contradiction. So $\sum_{i \in [s]} f_i \leq c - f_{\mathcal{C}}(w)$. Then we can derive that $18c - \binom{20}{2} \leq \sum_{1 \leq i \leq 20} |Y(F_i)| \leq s \cdot f_{\mathcal{C}}(w) + \sum_{i \in [s]} f_i + (20 - s)c \leq (s - 1)f_{\mathcal{C}}(w) + (21 - s)c \leq 10 f_{\mathcal{C}}(w) + 10c$, implying that $f_{\mathcal{C}}(w) \geq (8c - \binom{20}{2})/10 \geq 0.6c$, as desired. Clearly such w is unique, denoted by $w_{\mathcal{C}}$.

We say $w_{\mathcal{C}}$ is the *associated* vertex of $\mathcal{C}$. We also point out that from the above proof, among any 20 lines $F_1, ..., F_{20}$ in $V(\mathcal{D}) \cap \mathcal{F}_I$, the vertex $w_{\mathcal{C}}$ is adjacent to at least 11 of u_{F_i}'s in G. Consider any two incident edges in H, say $\mathcal{C}\mathcal{C}_1, \mathcal{C}\mathcal{C}_2 \in E(H)$. Let w_1, w_2 be the associated vertices of $\mathcal{C}_1, \mathcal{C}_2$, respectively. Take any 20 lines $F_1, ..., F_{20}$ in $V(\mathcal{C}) \cap \mathcal{F}_I$. Then each of w_1 and w_2 is adjacent to at least 11 of u_{F_i}'s in G. So there are two vertices u_{F_i}, u_{F_j} adjacent to both w_1 and w_2 in G. If $w_1 \neq w_2$, then this forces a C_4, a contradiction. So we conclude that $w_1 = w_2$.

Now suppose on the contrary that H contains an odd cycle, say $\mathcal{C}_1 \mathcal{C}_2 ... \mathcal{C}_t \mathcal{C}_1$, where t is odd. Let w_i be the associated vertex of $\mathcal{C}_i$ for each $i \in [t]$. Applying the above conclusion, as t is odd, we can

21

derive that these w_i's are the same vertex, say w. Take any $L_1 \in V(\mathcal{C}_1) \cap \mathcal{F}_{II}$ and $L_2 \in V(\mathcal{C}_2) \cap \mathcal{F}_{II}$ with $v_{L_1}^\star = v_{L_2}^\star = w$. Then $L_1 = N(u_{L_1}) \cup \{w\}$ and $L_2 = N(u_{L_2}) \cup \{w\}$, implying $1 \le |L_1 \cap L_2| \le 2$. If $|L_1 \cap L_2| = 1$, then $\mathcal{C}_1$ and $\mathcal{C}_2$ are friendly, a contradiction. Hence $|L_1 \cap L_2| = 2$. Let $L_1 R_1$ be an edge in $\mathcal{C}_1$ and $L_2 R_2$ be an edge in $\mathcal{C}_2$, where R_1, R_2 are type-I. By Claim 5.16, as $|L_1 \cap L_2| = 2$, we have $|R_1 \cap R_2| = 2$. So (A) holds for the line $R_1 \in V(\mathcal{C}_1) \cap \mathcal{F}_I$ and the heavy component $\mathcal{C}_2$, a contradiction. This proves that H is bipartite, completing the proof of Claim 5.20. $\qquad\square$

Finally, we show how to add certain good lines into $\mathcal{R}$ to make a larger 1-intersecting hypergraph on the vertex set V. By Claim 5.20, there exists a 1-intersecting $\mathcal{F}_h' \subseteq \mathcal{F}_h$ with $|\mathcal{F}_h'| \ge |\mathcal{F}_h|/2$. By Claim 5.18, we see that $\mathcal{R} \cup \mathcal{F}_r \cup \mathcal{F}_h'$ is also 1-intersecting. Since $L \in \mathcal{L}$ satisfies Claim 5.7, by Lemma 2.5, we see that $\mathcal{R} \cup \mathcal{F}_r \cup \mathcal{F}_h' \cup \mathcal{L}$ remains 1-intersecting. Using Claim 5.19, we have

$$|\mathcal{R} \cup \mathcal{F}_r \cup \mathcal{F}_h' \cup \mathcal{L}| = |\mathcal{R}| + |\mathcal{F}_r \cup \mathcal{F}_h' \cup \mathcal{L}| \ge |\mathcal{R}| + \frac{1}{2}|\mathcal{F}_r \cup \mathcal{F}_h \cup \mathcal{L}|$$

$$\ge (q^2 - \alpha q + 1 - 2/\delta) + \frac{1}{2}(2\alpha q + (1-\epsilon)q/2) \ge q^2.$$

This finishes the proof of Lemma 5.1. $\qquad\blacksquare$

6 Finding a polarity

In this section, we complete the proof of Theorem 3.2. Fix $0 < \epsilon < 1$ and let q be an even integer with $q \gg \epsilon$. Let $G = (V, \mathcal{E})$ be a C_4-free graph on $q^2 + q + 1$ vertices with maximum degree $q + 1$ and at least $\frac{1}{2}q(q+1)^2 - \frac{\epsilon}{2}q$ edges. We aim to show that there exists a unique polarity graph of order q containing G as a subgraph.

By Lemma 5.1, there exists a 1-intersecting $(q+1)$-hypergraph $\mathcal{R}^*$ on the vertex set V with $\mathcal{R} \subseteq \mathcal{R}^*$ and $|\mathcal{R}^*| \ge q^2$. By Theorem 2.1, $\mathcal{R}^*$ (and thus $\mathcal{R}$) can be embedded into a projective plane $\mathcal{P}$ of order q. Since $|\mathcal{R}| = |A| \ge q^2 - \epsilon q - 2/\delta > q^2 - q + 1$ (by Claim 5.1), applying Theorem 2.2, we see that such $\mathcal{P}$ and the embedding of $\mathcal{R}$ into $\mathcal{P}$ both are unique. Let $\mathcal{R}^c = \mathcal{P} \backslash \mathcal{R}$. Now let us recall some basic facts about $\mathcal{P}$: every two lines intersect with exactly one vertex, every two vertices are contained in exactly one line, and every column or row of any incidence matrix of $\mathcal{P}$ has $q + 1$ 1's as entries.

We say $v \in V$ is *feasible*, if there exists a line $L \in \mathcal{P}$ with $N(v) \subseteq L$; otherwise, we say v is *non-feasible*. For non-feasible v, we say it is *near-feasible*, if there exist a line $L \in \mathcal{R}^c$ and a subset $K_v \subseteq N(v)$ such that $N(v) \backslash K_v \subseteq L$ and $|K_v| \le 5\sqrt{q}/\delta$. In both definitions, we say v and L are *associated* with each other. For feasible v, we let $K_v = \emptyset$. By (11) and since G is C_4-free, for any two feasible or near-feasible vertices u and v, we have

$$|(N(u) \backslash K_u) \cup (N(v) \backslash K_v)| \ge \big(d(u) - 5\sqrt{q}/\delta\big) + \big(d(v) - 5\sqrt{q}/\delta\big) - 1 \ge (2 - \epsilon)q - 10\sqrt{q}/\delta - 1 > q + 1.$$

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 $d(v) \ge (1 - \epsilon)q$ by (11), then 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 $\mathcal{P}$.

Next we study some properties on non-feasible vertices $v \in V$. Let $N(v) = \{v_1, ..., v_d\}$. Since v is non-feasible, we see $N(v) \nsubseteq L$ for any $L \in \mathcal{P}$ and thus $v \notin A$. Then any pair $\{v_i, v_j\}$ for $i, j \in [d]$ is not contained in any line $N(u) \in \mathcal{R}$. This is because that otherwise, we see that $v_i u v_j v v_i$ forms a C_4 in G, a contradiction. So every such pair $\{v_i, v_j\}$ is contained in a unique line $L \in \mathcal{R}^c$. Let $\mathcal{L}_v$ be the family of lines $L \in \mathcal{P}$ which contains at least two vertices of $N(v)$. Then we have $\mathcal{L}_v \subseteq \mathcal{R}^c$ and thus

$$|\mathcal{L}_v| \le |\mathcal{R}^c| = |\mathcal{P}| - |\mathcal{R}| \le (1 + \epsilon)q + 2/\delta + 1. \tag{17}$$

We also point out that any vertex in $N(v)$ appears in at least two lines of $\mathcal{L}_v$.

We process to show that all non-feasible vertices are near-feasible in the following claims. First we show any vertex has a neighbor which belongs to many lines in $\mathcal{R}$.

Claim 6.1. *Any vertex $v \in V$ has a neighbor v_j with $d_{\mathcal{R}}(v_j) = |N(v_j) \cap A| \geq q - \frac{\epsilon}{1-\epsilon} - 2/\delta - 2$. In addition, if $v \notin B$ has degree at least $\frac{1}{2}(1 + \epsilon + 4\delta)q + 6/\delta + 1$, then v has a neighbor v_j with $d_{\mathcal{R}}(v_j) \geq q - 1$.*

Proof. Let $N(v) = \{v_1, ..., v_d\}$. By (11), we have $d = d(v) \geq (1-\epsilon)q$. 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

$$q + \epsilon q + 1 \geq f(V) \geq \sum_{i \in [d]} f(N_i \cup \{v_i\}) + f(v) = \sum_{i \in [d]} f(N_i \cup \{v_i\}) + (q + 1 - d).$$

By averaging, there is some $j \in [d]$ with $f(N_j \cup \{v_j\}) \leq \frac{\epsilon q}{d} + 1 \leq \frac{\epsilon}{1-\epsilon} + 1$. Therefore,

$$d_{\mathcal{R}}(v_j) \geq |N_j \cap A| \geq |N_j| - |N_j \cap S| - |B| \geq \big(d(v_j) - 2\big) - f(N_j) - 2/\delta$$
$$= \big(q - 1 - f(v_j)\big) - f(N_j) - 2/\delta \geq q - \epsilon/(1 - \epsilon) - 2/\delta - 2,$$

as desired. Next we consider vertices $v \notin B$ with $d = d(v) \geq \frac{1}{2}(1 + \epsilon + 4\delta)q + 6/\delta + 1$. Let $B_v = N(v) \cap (S \cup B)$ and $B_v^* = B_v \cup (N(B_v) \cap N(v))$. Then we have $|B_v| \leq |N(v) \cap S| + |B| \leq \delta q + 2/\delta$. Since G is C_4-free, every vertex in B_v has at most one neighbor in $N(v)$, implying that $|B_v^*| \leq 2|B_v|$. Let $T = \{v_i \in N(v) \backslash B_v^* : N_i \cap B = \emptyset\}$. Since N_i's are disjoint and there are at most $|B|$ many N_i's containing some vertex in B, we get $|T| \geq |N(v) \backslash B_v^*| - |B| \geq d - 2\delta q - 6/\delta$. If $f(N_j) \geq 2$ for all $v_j \in T$, then $q + \epsilon q + 1 \geq f(V) \geq 2|T| \geq 2(d - 2\delta q - 6/\delta) \geq q + \epsilon q + 2$, a contradiction. Therefore, there exists a vertex $v_j \in T$ such that $f(N_j) \leq 1$. By the definition of T, we can see that $d_{\mathcal{R}}(v_j) = |N(v_j) \cap A| \geq d(v_j) - 1 - f(N_j) \geq q - 1$. This completes the proof. $\square$

We partition V into three disjoint sets $U_1 \cup U_2 \cup U_3$, where U_1 consists of all feasible vertices and U_2 consists of non-feasible vertices $v \notin B$ with $d(v) \geq \frac{1}{2}(1 + \epsilon + 4\delta)q + 6/\delta + 1$.

Claim 6.2. *There exists some $w \in V$ such that all $v \in U_2$ are near-feasible with $K_v = \{w\}$.*

Proof. For any $v \in U_2$, by Claim 6.1, there is a neighbor v_j of v with $d_{\mathcal{R}}(v_j) \geq q - 1$. By the above discussion, v_j appears in at least two lines in $\mathcal{L}_v \subseteq \mathcal{R}^c$. If $d_{\mathcal{R}}(v_j) \geq q$, then $d_{\mathcal{P}}(v_j) \geq q + 2$, a contradiction. So $d_{\mathcal{R}}(v_j) = q - 1$ and there are exactly two lines, say L_1 and L_2, in $\mathcal{L}_v \subseteq \mathcal{R}^c$ containing v_j. Let $N_1 = L_1 \cap N(v)$ and $N_2 = L_2 \cap N(v)$. Then we have $N_1 \cap N_2 = \{v_j\}$ and $N_1 \cup N_2 = N(v)$. Consider any other line $L_i \in \mathcal{L}_v \backslash \{L_1, L_2\}$ for $i \geq 3$. Set $N_i = L_i \cap N(v)$. We see that for any $i \geq 3$ and $j \in \{1, 2\}$, $|N_i \cap N_j| \leq 1$ and $|N_i \cap N_1| + |N_i \cap N_2| \geq |N_i \cap (N_1 \cup N_2)| = |N_i| \geq 2$. This shows that for any $i \geq 3$, N_i consists of two vertices, one from $N_1 \backslash \{v_j\}$ and the other from $N_2 \backslash \{v_j\}$. Hence, $|\mathcal{L}_v| = (|N_1| - 1)(|N_2| - 1) + 2$.

Let $d = d(v)$. We may assume that $d - 1 \geq |N_1| \geq |N_2| \geq 2$. If $|N_2| \geq 3$, then we have $|\mathcal{L}_v| = (|N_1| - 1)(|N_2| - 1) + 2 \geq 2(d - 3) + 2 = 2d - 4 \geq (1 + \epsilon + 4\delta)q + 12/\delta - 2 > (1 + \epsilon)q + 2/\delta + 1 \geq |\mathcal{L}_v|$, where the last inequality holds by (17), a contradiction. Thus, $|N_1| = d - 1$ and $|N_2| = 2$, implying $|\mathcal{L}_v| = d$. Suppose that $N_2 = \{v_j, w\}$. Then every N_i for $2 \leq i \leq d$ contains the vertex w. Also $N(v) \backslash \{w\} \subseteq L_1 \in \mathcal{R}^c$, implying that $v \in U_2$ is near-feasible with $K_v = \{w\}$.

Assume there is another non-feasible vertex $v' \in U_2$ with $K_{v'} = \{w'\}$, where $w' \neq w$. Let $d = d(v)$ and $d' = d(v')$. By the above arguments, we see w and w' appear in $d - 1$ and $d' - 1$ lines in $\mathcal{R}^c$, respectively. By (17), we have $|\mathcal{R}^c| + 2 \leq (1 + \epsilon)q + 2/\delta + 3 \leq (1 + \epsilon + 4\delta)q + 12/\delta \leq (d - 1) + (d' - 1)$, which shows that w and w' appear in at least two lines of $\mathcal{R}^c$ in common. This contradicts that $\mathcal{P}$ is a projective plane. $\square$

Claim 6.3. *All non-feasible vertices are near-feasible.*

Proof. Let $v \in V$ be any non-feasible vertex. We have $d(v) \geq (1-\epsilon)q$. By Claim 6.1, v has a neighbor u with $d_{\mathcal{R}}(u) = q + 1 - m$, where $m \leq \frac{\epsilon}{1-\epsilon} + 2/\delta + 3$. Let $\mathcal{U} = \{L \in \mathcal{L}_v : u \in L \cap N(v)\}$. We have $|\mathcal{U}| \leq m$ and $\cup_{L \in \mathcal{U}} N_L = N(v)$, where $N_L := L \cap N(v)$. We assert that for all but at most one $L \in \mathcal{U}$, the size of N_L is at most $2\sqrt{q}$. Suppose on the contrary that there are $L_1, L_2 \in \mathcal{U}$ with $|N_{L_1}| \geq 2\sqrt{q} + 1$

23

and $|N_{L_2}| \geq 2\sqrt{q} + 1$. Then all pairs (x, y) with $x \in N_{L_1}\backslash\{u\}$ and $y \in N_{L_2}\backslash\{u\}$ should appear in distinct lines in $\mathcal{L}_v$. By (17), this shows that $(1 + \epsilon)q + 2/\delta + 1 \geq |\mathcal{L}_v| \geq (|N_{L_1}| - 1)(|N_{L_2}| - 1) \geq 4q$, a contradiction.

Let L_1 be the line in $\mathcal{U}$ with the maximum N_{L_1} and let $K_v = \bigcup_{L\in\mathcal{U}\backslash\{L_1\}}(N_L\backslash\{u\})$. Then $N(v)\backslash K_v \subseteq L_1 \in \mathcal{R}^c$ with $|K_v| \leq \sum_{L\in\mathcal{U}\backslash\{L_1\}}(|N_L| - 1) \leq (m-1)\cdot 2\sqrt{q} \leq 2\left(\frac{\epsilon}{1-\epsilon} + \frac{2}{\delta} + 2\right)\sqrt{q} \leq \frac{5\sqrt{q}}{\delta}$. Therefore, v is near-feasible. $\square$

We express $V = \{v_1, ..., v_n\}$ such that $U_1 = \{v_1, ..., v_a\}$, $U_2 = \{v_{a+1}, ..., v_b\}$ and $U_3 = \{v_{b+1}, ..., v_n\}$ for $1 \leq a < b \leq n$. Since all vertices in G are feasible or near-feasible, by the discussion before Claim 6.1, we can conclude that each $v_i \in V$ is associated with a unique line denoted by L_i in $\mathcal{P}$. 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.

Let $s := |U_3|$. We point out that any $v \in U_3$ either is in B or has $d(v) \leq (1 + \epsilon + 4\delta)q/2 + 6/\delta$. In the latter case, we have the deficiency $f(v) = q + 1 - d(v) \geq (1 - \epsilon - 4\delta)q/2 - 6/\delta$. Hence by (9), we have
$$s \leq |B| + \frac{f(V)}{(1-\epsilon-4\delta)q/2 - 6/\delta} \leq \frac{2}{\delta} + \frac{q + \epsilon q + 1}{(1-\epsilon-4\delta)q/2 - 6/\delta} \leq C^*,$$
where C^* is a constant depending on ϵ and δ only. Let K be the union of K_v's over all $v \in V$. By Claims 6.2 and 6.3, we know that $K_v = \emptyset$ for $v \in U_1$, $K_v = \{w\}$ for $v \in U_2$ and $|K_v| \leq 5\sqrt{q}/\delta$ for $v \in U_3$. Hence $|K| \leq 1 + s \cdot 5\sqrt{q}/\delta = O(\sqrt{q})$.

Claim 6.4. $\mathcal{M}$ is symmetric.

Proof. We assert that if $v_i \in A\backslash K$, then $m_{ij} = m_{ji}$ for all $j \in [n]$. If $m_{ij} = 1$, then 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. Since $|A\backslash K| \geq |A| - |K| \geq (q^2 - \epsilon q - 2/\delta) - O(\sqrt{q}) \geq q^2 - q + 3$, by Lemma 2.4, the whole matrix $\mathcal{M}$ is symmetric. $\square$

Hence we see from (2) that the above function $\pi : V \leftrightarrow \mathcal{P}$ is a polarity of the projective plane $\mathcal{P}$. Let H be the polarity graph of π. For any $k \times \ell$ matrices $\mathcal{X} = (x_{ij})$ and $\mathcal{Y} = (y_{ij})$, we say $\mathcal{X}$ is at most $\mathcal{Y}$ if $x_{ij} \leq y_{ij}$ for all i, j and we express this by $\mathcal{X} \leq \mathcal{Y}$.

Now we are finishing the proof of Theorem 3.2 by showing that G is a subgraph of H. Let $\mathcal{A} = (a_{ij})$ be the adjacent matrix of the graph G. It suffices to shows that $\mathcal{A} \leq \mathcal{M}$. We call these (i, j)-entries with $a_{ij} = 1$ and $m_{ij} = 0$ *problematic*. Since both $\mathcal{A}$ and $\mathcal{M}$ are 0/1 matrices, it is equivalent for us to show that there is no problematic entries.

For every $v_i \in U_1$, as it is feasible, we see that $N(v_i) \subseteq L_i$ and thus the i'th row of $\mathcal{A}$ is at most the i'th row of $\mathcal{M}$. Since both $\mathcal{A}$ and $\mathcal{M}$ are symmetric, the i'th column of $\mathcal{A}$ is also at most the i'th column of $\mathcal{M}$, whenever $v_i \in U_1$. Now consider vertices $v_i \in U_2$. By Claim 6.2, $N(v_i)\backslash\{w\} \subseteq L_i$, where $w = v_\ell$ is fixed. Consider $a_{ij} = 1$ for possible j which is not ℓ. Then we have $v_j \in N(v_i)\backslash\{w\} \subseteq L_i$. This shows that the i'th row of $\mathcal{A}$ is at most the i'th row of $\mathcal{M}$, except the (i, ℓ)-entry. By symmetry, we see that for all $v_i \in U_2$, the i'th column of $\mathcal{A}$ is at most the i'th column of $\mathcal{M}$, except the possible (ℓ, i)-entry. We also know w is feasible or near-feasible. So $|K_w| \leq 5\sqrt{q}/\delta$ and the number of problematic (ℓ, i)-entries is clearly at most $|K_w| \leq 5\sqrt{q}/\delta$. This further shows that the number of problematic (i, j)- or (j, i)-entries for all $v_i \in U_2$ is at most $10\sqrt{q}/\delta$. Note that $|U_3| = s$ is at most a constant C^* depending on only ϵ and δ. Putting all the above together, we see that the number of problematic (i, j)-entries for $i, j \in [n]$ is at most $10\sqrt{q}/\delta + s^2 = O(\sqrt{q})$.

Let E_0 be the set of $v_i v_j$ for all problematic (i, j)-entries. It is easy to see that $E_0 = E(G)\backslash E(H)$ and $|E_0| = O(\sqrt{q})$. Suppose that there is some edge say $e = v_i v_j \in E_0$. By Lemma 2.8, $H \cup \{e\}$ contains at least $q - 1$ copies of C_4, all of which contain the edge e and are edge-disjoint otherwise. Hence in order to turn $H \cup \{e\}$ into a subgraph of G containing e (which is C_4-free), one needs to delete at least $q - 1$ edges in $H \cup \{e\}$. On the other hand, since H is a polarity graph, we have

$e(H) \le \frac{1}{2}q(q+1)^2$ and $|E(H)\backslash E(G)| - |E_0| = e(H) - e(G) \le \frac{1}{2}\epsilon q$. So one can delete $|E(H)\backslash E(G)| \le \frac{1}{2}\epsilon q + |E_0| \le \frac{1}{2}\epsilon q + O(\sqrt{q}) < q - 1$ edges to turn $H \cup \{e\}$ into a subgraph of G while preserving the edge e. This is a contradiction. Therefore, $E_0 = \emptyset$ and G is a subgraph of H.

It only remains to show that the polarity graph H is unique. Recall that the projective plane $\mathcal{P}$ containing $\mathcal{R}$ has been shown to be unique. So it is equivalent to show that the polarity π is unique. Suppose for a contradiction that there exists another polarity $\pi' : V \leftrightarrow \mathcal{P}$, where $\pi' : v_i \leftrightarrow L_{\sigma(i)}$ for some permutation σ on $[n]$. Let $\mathcal{M}' = (m'_{ij})$ be the incidence matrix of $\mathcal{P}$ with respect to π'. By the same proof as above, we can deduce that $\mathcal{A} \le \mathcal{M}'$. By (11), we see that any vertex $v_i \in V$ has degree at least $(1 - \epsilon)q \ge 2$. Choose any pair $\{x_i, y_i\} \subseteq N(v_i)$. Since the i'th row of $\mathcal{A}$ is at most the i'th row of $\mathcal{M}'$, we see $\{x_i, y_i\} \subseteq N(v_i) \subseteq L_{\sigma(i)} \in \mathcal{P}$. Also we have $\{x_i, y_i\} \subseteq N(v_i) \subseteq L_i \in \mathcal{P}$. Since $\mathcal{P}$ is a projective plane, it is clear that $L_{\sigma(i)} = L_i$ for all $i \in [n]$. This shows that $\pi = \pi'$ and indeed the polarity graph H is unique. The proof of Theorem 3.2 (and thus Theorem 1.2) is completed. ∎

We remark that it would suffice to choose $q_\epsilon = \frac{10^{10}}{(1-\epsilon)^6}$ in the statement of Theorem 3.2.

7 Turán numbers

In this section, we discuss the consequences of Theorem 1.2 on Turán numbers. First, let us restate and prove Corollary 1.3. Recall the definition of $\lambda(q)$.

Corollary 7.1. *Let q be even. If $\lambda(q) \ge \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)$, then $\mathrm{ex}(q^2 + q + 1, C_4) = \lambda(q)$, where the equality holds only for polarity graphs of order q with $\lambda(q)$ edges; otherwise, $\mathrm{ex}(q^2 + q + 1, C_4) < \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)$. In particular, $\mathrm{ex}(q^2 + q + 1, C_4) \le \max\left\{\lambda(q), \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)\right\}$.*

Proof. Let q be even and G be an extremal graph for $\mathrm{ex}(q^2 + q + 1, C_4)$. First suppose that $\lambda(q) \ge \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)$. As $e(G) \ge \lambda(q) \ge \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)$, by Theorem 1.2, there exists a polarity graph H of order q containing G as a subgraph. Then we have $\lambda(q) \le e(G) \le e(H) \le \lambda(q)$, which implies that $G = H$ must be a polarity graph of order q with $\lambda(q)$ edges. Now assume $\lambda(q) < \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)$. By Theorem 1.2, it is easy to conclude that $e(G) < \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)$. ∎

A quick inference of this corollary is that: For all even integers q such that there is no projective planes of order q, it holds that

$$\mathrm{ex}(q^2 + q + 1, C_4) \le \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q).$$

We point out that by Theorem 2.3, there are infinitely many such integers q, including all integers $q \equiv 2 \mod 4$ which cannot be expressed as a sum of two square numbers.

Another inference can be stated related to the existence of orthogonal polarity graphs of order q.

Corollary 7.2. *Let q be a large even integer. If there exists an orthogonal polarity graph of order q, then $\mathrm{ex}(q^2 + q + 1, C_4) = \frac{1}{2}q(q+1)^2$; otherwise, we have $\mathrm{ex}(q^2 + q + 1, C_4) \le \frac{1}{2}q(q+1)^2 - \frac{1}{2}\sqrt{q}$ and in addition if q is not a square number, then $\mathrm{ex}(q^2 + q + 1, C_4) \le \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)$.*

Proof. By Proposition 2.6, any polarity graph of order q has $\frac{1}{2}q(q+1)^2 - \frac{1}{2}m\sqrt{q}$ edges for some integer $m \ge 0$. The first assertion follows by Theorem 1.1. Now we may assume that $m \ge 1$ for any polarity graph of order q and thus $\lambda(q) \le \frac{1}{2}q(q+1)^2 - \frac{1}{2}\sqrt{q}$. By Corollary 1.3, $\mathrm{ex}(q^2 + q + 1, C_4) \le \max\left\{\lambda(q), \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)\right\} \le \frac{1}{2}q(q+1)^2 - \frac{1}{2}\sqrt{q}$. In addition, if q is not a square number, then this implies that there is no polarity graphs of order q and the conclusion follows easily. ∎

We conclude this section with an explicit lower bound of $\mathrm{ex}(n, C_4)$ for later use.

Proposition 7.3. *For any sufficiently large n, there exists some prime number p with $\sqrt{n} - n^{0.2625} - 1 \le p \le \frac{1}{2}(-1 + \sqrt{4n - 3})$ such that $\mathrm{ex}(n, C_4) \ge \mathrm{ex}(p^2 + p + 1, C_4) \ge \frac{1}{2}(n^{1.5} - 3n^{1.2625} + n)$.*

Proof. Let $x = \frac{1}{2}(-1 + \sqrt{4n - 3})$. As n is sufficiently large, by Theorem 2.13, there exists a prime number $p \in [x - x^{0.525}, x]$. So $n = x^2 + x + 1 \ge p^2 + p + 1$, where $p \ge x - x^{0.525} \ge \sqrt{n} - n^{0.2625} - 1$. By Theorem 1.1, we derive $\mathrm{ex}(n, C_4) \ge \mathrm{ex}(p^2 + p + 1, C_4) = \frac{1}{2}p(p+1)^2 \ge \frac{1}{2}(n^{1.5} - 3n^{1.2625} + n)$. ∎

8 Supersaturation: one additional edge

8.1 Proof of Theorem 1.5

We devote this subsection to the proofs of Theorem 1.5 and Corollary 1.6.

Proof of Theorem 1.5. Let G be any graph on $q^2 + q + 1$ vertices with $e(G) = \frac{1}{2}q(q+1)^2 + 1$ edges where $q \geq 10^{12}$ is an even integer. Our goal is to show that either G has at least $2q - 3$ copies of C_4, or G is obtained from an orthogonal polarity graph of order q by adding an edge. In the latter case, by Lemma 2.8, we see that G has $q - 1, q$ or $q + 1$ copies of C_4.

By $\#C_4$, we denote the number of copies of C_4 in G. For $v \in V := V(G)$, let $c(v)$ be the number of copies of C_4 containing v. In the following, we assume that

$$\#C_4 \leq 2q - 4 \quad \text{and} \quad \sum_{v \in V} c(v) = 4 \cdot \#C_4 < 8q. \tag{18}$$

We say a pair $\{u, v\} \subseteq V$ is *opposite* if $d(u, v) \geq 2$. The deficiency $f(v)$ is defined by $f(v) = \max\{q + 1 - d(v), 0\}$. So $f(v) + d(v) \geq q + 1$ holds for every $v \in V$.

Recall the definitions of sets UP and P_2. For $A \subseteq V$, let $UP \cap A$ be the set of uncovered pairs $\{u, v\} \subseteq A$ of G and let $P_2 \cap A$ be the set of 2-paths of G with both endpoints in A.

Claim 8.1. *For any $A \subseteq V$, we have $2 \cdot \#C_4 \geq |P_2 \cap A| + |UP \cap A| - \binom{|A|}{2}$.*

Proof. For any subset $A \subseteq V$, it holds that

$$2 \cdot \#C_4 \geq \sum_{\text{covered pairs } \{u,v\} \subseteq A} \binom{d(u, v)}{2} \geq \sum_{\text{covered pairs } \{u,v\} \subseteq A} \left(d(u, v) - 1\right),$$

the righthand side of which equals $|P_2 \cap A| + |UP \cap A| - \binom{|A|}{2}$. $\qquad\square$

Similarly as in earlier sections, we let S_i be the set of all vertices of degree i in G and let $S = \cup_{i=0}^{q} S_i$. For $v \in V$, let $d_0(v)$ be the number of vertices $u \in V$ with $d(u, v) = 0$.

Claim 8.2. *Every vertex v has $c(v) \geq (d(v) - q - 1)q - f(N(v)) + d_0(v)$. If $d(v) \geq q + 1$ and $N(v) \cap S = \emptyset$, then $c(v) \geq 1$.*

Proof. By counting the 2-paths with the fixed endpoint v, we get

$$\sum_{u \in V \setminus \{v\}} d(v, u) = \sum_{w \in N(v)} (d(w) - 1) \geq \sum_{w \in N(v)} (q - f(w)) = d(v)q - f(N(v)). \tag{19}$$

Then we see that $c(v) = \sum_{u \in V \setminus \{v\}} \binom{d(v, u)}{2}$ is at least

$$\sum_{u \in V \setminus \{v\}} (d(v, u) - 1) + d_0(v) \geq (d(v) - q - 1)q - f(N(v)) + d_0(v).$$

Now consider $v \in V$ with $d(v) \geq q + 1$ and $N(v) \cap S = \emptyset$. Clearly $f(N(v)) = 0$. If $d(v) \geq q + 2$, then the above inequality immediately implies that $c(v) \geq q \geq 1$. So let $d(v) = q + 1$. Suppose that v is not contained in any C_4. If $d_0(v) \geq 1$, then the above inequality again implies that $c(v) \geq d_0(v) \geq 1$. So $d_0(v) = 0$, that is, every vertex in $V \setminus \{v\}$ has exactly one common neighbor with v. However this is impossible, as $d(v) = q + 1$ is odd and thus $G[N(v)]$ can not consist of a perfect matching. $\qquad\square$

We now show that the maximum degree of G is at most $q + 3$. Let $V = \{v_1, ..., v_n\}$.

Claim 8.3. $\Delta(G) \leq q + 3$.

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

$$2q - 4 \geq c(v_1) = \sum_{i=2}^{n} \binom{a_i}{2} \geq \sum_{i=2}^{n} (a_i - 1),$$

implying that $\sum_{i=2}^{n} a_i \leq q^2 + 3q - 4$. Also we have

$$\sum_{i=2}^{n} (d(v_i) - a_i) = 2e(G) - d(v_1) - \sum_{i=2}^{n} a_i = (q^2 + q)(q - 1) + (2q^2 + q + 2 - k - \sum_{i=2}^{n} a_i),$$

where clearly $2q^2 + q + 2 - k - \sum_{i=2}^{n} a_i \geq -(q^2 + q)$. By Lemma 2.14, we can deduce that

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

By Claim 8.1, we have $2 \cdot \#C_4 \geq |P_2 \cap (V \backslash N(v_1))| - \binom{n - d(v_1)}{2}$. This further implies that

$$2 \cdot \#C_4 \geq q^3 - 2q^2 + 2q - (q - 1)\sum_{i=2}^{n} a_i - 0.5k^2 + k(q^2 - q + 1.5) - 2.$$

When $k = 4$ and $\sum_{i=2}^{n} a_i = q^2 + 3q - 4$, the righthand side of this inequality achieves its minimum value, which is $5q - 8 > 4q - 8$. This is a contradiction to (18). $\qquad \square$

Using Claim 8.3, we see that the deficiency of V can be expressed as

$$f(V) = (q + 1)n + |S_{q+2}| + 2|S_{q+3}| - 2e(G) = q - 1 + |S_{q+2}| + 2|S_{q+3}|. \tag{20}$$

Claim 8.4. $|S_{q+2} \cup S_{q+3}| \leq 3\sqrt{q}$.

Proof. We show that any two vertices say v_1, v_2 in $S_{q+2} \cup S_{q+3}$ with $c(v_1), c(v_2) < 0.2q$ form an opposite pair. Suppose for a contradiction that $d(v_1, v_2) \leq 1$. Let $d(v_1) = q + 2 + \delta_1$ and $d(v_2) = q + 2 + \delta_2$ where $\delta_1, \delta_2 \in \{0, 1\}$. Let $a_i = |N(v_i) \cap (N(v_1) \cup N(v_2))|$ for all $3 \leq i \leq n$. Then there are at least $a_i - 2$ copies of C_4 with the opposite pair $\{v_i, v_1\}$ or $\{v_i, v_2\}$ in G. First we assume that v_1, v_2 have exactly one common neighbor say v_3. In this case, if $v_i \in N(v_3) \backslash \{v_1, v_2\}$, then there are at least $a_i - 1$ copies of C_4 with the opposite pair $\{v_i, v_1\}$ or $\{v_i, v_2\}$. 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 Claim 8.1 and Lemma 2.14, if we write $V' = V \backslash (N(v_1) \cup N(v_2))$, then one can derive

$$2 \cdot \#C_4 \geq |P_2 \cap V'| - \binom{|V'|}{2} \geq \sum_{i=4}^{n} \binom{d(v_i) - a_i}{2} - \binom{q^2 - q - 2 - \delta_1 - \delta_2}{2}$$

$$\geq (q^2 + q - 2)\binom{q - 1}{2} - (0.4q + 6)(q - 1) - \binom{q^2 - q - 2}{2} = 0.1q^2 - 4.1q + 1 > 4q - 8,$$

a contradiction to (18). Now we can assume that v_1, v_2 have no common neighbor. By similar arguments, we have $\sum_{i=3}^{n}(a_i - 2) \leq 0.4q$ and thus $\sum_{i=3}^{n} a_i \leq 2q^2 + 2.4q - 2$. This shows that

$$\sum_{i=3}^{n} (d(v_i) - a_i) = 2e(G) - d(v_1) - d(v_2) - \sum_{i=3}^{n} a_i \geq (q^2 + q - 1)(q - 1) + (-1.4q - 3),$$

where $-1.4q - 3 \geq -(q^2 + q - 1)$. Using Claim 8.1 and Lemma 2.14 again, we have

$$2 \cdot \#C_4 \geq |P_2 \cap V'| - \binom{|V'|}{2} = \sum_{i=3}^{n} \binom{d(v_i) - a_i}{2} - \binom{q^2 - q - 3 - \delta_1 - \delta_2}{2}$$

$$\geq (q^2 + q - 1)\binom{q-1}{2} - (1.4q + 3)(q - 1) - \binom{q^2 - q - 3}{2} = 0.6q^2 - 2.6q - 4 > 4q - 8,$$

a contradiction. This shows that v_1, v_2 form an opposite pair.

If there are more than $\sqrt{8q}$ vertices $u \in S_{q+2} \cup S_{q+3}$ with $c(u) < 0.2q$, then we have at least $\binom{\sqrt{8q}+1}{2} \geq 4q$ opposite pairs and thus $2q$ copies of C_4, a contradiction. Hence there are at most $\sqrt{8q}$ vertices $u \in S_{q+2} \cup S_{q+3}$ with $c(u) < 0.2q$. Also by (18), there are at most $8q/0.2q = 40$ vertices $w \in V$ with $c(w) \geq 0.2q$. So in total we have $|S_{q+2} \cup S_{q+3}| \leq \sqrt{8q} + 40 \leq 3\sqrt{q}$, finishing the proof. □

Since $|S_{q+2} \cup S_{q+3}| \leq 3\sqrt{q}$, by (20) we have $|S| \leq f(V) = q - 1 + |S_{q+2}| + 2|S_{q+3}| \leq q + 6\sqrt{q} - 1$. Then $|S_{q+1}| = n - |S| - |S_{q+2} \cup S_{q+3}| > q^2 - 9\sqrt{q}$. If $|S| \leq q - 9$, there are at least $|S_{q+1}| - q|S| \geq 8q$ vertices in S_{q+1} with no neighbors in S, by Claim 8.2 each of which is contained in at least one C_4 and thus there are at least $2q$ copies of C_4 in G, a contradiction. Therefore we have $q - 8 \leq |S| \leq f(V) \leq q + 6\sqrt{q} - 1$. Furthermore, for any set $T \subseteq V$, it follows $q + 6\sqrt{q} - 1 \geq f(V) \geq f(T) + (|S| - |S \cap T|) \geq f(T) - |S \cap T| + (q - 8)$, implying that

$$f(T) \leq |S \cap T| + 6\sqrt{q} + 7. \tag{21}$$

We now improve Claim 8.4 to the following.

Claim 8.5. $|S_{q+2} \cup S_{q+3}| \leq 5$.

Proof. Suppose on the contrary that there are six vertices $v_1, v_2, v_3, v_4, v_5, v_6 \in S_{q+2} \cup S_{q+3}$. For any $1 \leq i < j \leq 6$, we know $d(v_i, v_j) \leq 2\sqrt{q}$ (as otherwise there are at least $2q$ copies of C_4 in G). For any $j \in [6]$, by (19) and (21) we see that $\sum_{i=7}^{n} \binom{d(v_j, v_i)}{2}$ is at least

$$\sum_{i=7}^{n} (d(v_j, v_i) - 1) \geq \sum_{i \in [n] \setminus \{j\}} (d(v_j, v_i) - 1) - 10\sqrt{q} \geq (q + 2)q - f(N(v_j)) - (n - 1) - 10\sqrt{q}$$

$$\geq q - (|N(v_j) \cap S| + 6\sqrt{q} + 7) - 10\sqrt{q} = q - 16\sqrt{q} - 7 - |N(v_j) \cap S|.$$

Since each C_4 contains two opposite pairs, there are at least $\frac{1}{2} \sum_{j=1}^{6} \sum_{i=7}^{n} (d(v_j, v_i) - 1)$ copies of C_4 containing opposite pairs $\{v_j, v_i\}$ for $1 \leq j \leq 6$ and $7 \leq i \leq n$. This gives

$$4q \geq \sum_{j=1}^{6} \sum_{i=7}^{n} (d(v_j, v_i) - 1) \geq 6q - 96\sqrt{q} - 42 - \sum_{j=1}^{6} |N(v_j) \cap S|.$$

Thus $\sum_{j=1}^{6} |N(v_j) \cap S| \geq 2q - 96\sqrt{q} - 42$. By the inclusion-exclusion principle,

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

where last inequality holds since $d(v_i, v_j) \leq 2\sqrt{q}$ for $1 \leq i < j \leq 6$. Since $|S| \leq q + 6\sqrt{q} - 1$ and q is large, this is a contradiction and finishes the proof of Claim 8.5. □

Equipped with Claim 8.5, the discussion after Claim 8.4 can easily deduce that $q - 8 \leq |S| \leq f(V) \leq q + 9$ and for any $T \subseteq V$, $f(T) \leq |S \cap T| + 17$. In particular, any $v \in V$ has $f(v) \leq 18$ and $d(v) \geq q + 1 - f(v) \geq q - 17$.

Claim 8.6. *For any $v \in V$, either $|N(v) \cap S| \leq 20$ or $|N(v) \cap S| \geq q - 28$.*

Proof. Let $N(v)\backslash S = \{v_1, ..., v_t\}$. So $t = d(v) - |N(v) \cap S|$. For $i \in [t]$, $v_i \in S_{q+1} \cup S_{q+2} \cup S_{q+3}$ and we let $N_i = N(v_i)\backslash\{v\}$. If N_i, N_j have a common vertex x, then $vv_i x v_j v$ forms a distinct C_4. Since G has less than $2q$ copies of C_4, we derive that

$$|\bigcup_{i=1}^{t} N_i| \geq \sum_{i=1}^{t} |N_i| - 2q \geq qt - 2q.$$

Let $B_i = N_i \cap S$ and C_i be the set of vertices $x \in N_i$ with $c(x) = 0$. Then C_i's are disjoint over $i \in [t]$; indeed, if $x \in C_i \cap C_j$, then $vv_i x v_j v$ forms a C_4, a contradiction. Thus we have

$$\sum_{1\leq i \leq t} |B_i \cap C_i| = |\bigcup_{1\leq i \leq t} (B_i \cap C_i)| \leq |\bigcup_{1\leq i \leq t} B_i| \leq |S| \leq q + 9.$$

We now claim that every $x \in C_i \backslash B_i$ has at least one neighbor in $S\backslash N(v)$. By Claim 8.2, since $d(x) \geq q + 1$ and $c(x) = 0$, we see that x has at least one neighbor say $y \in S$. If $y \in N(v) \cap S$, then $vv_i x y v$ is a C_4, a contradiction. This proves the claim. Also it is clear that every vertex in $S\backslash N(v)$ has at most one neighbor in $C_i \backslash B_i$. Hence $|C_i \backslash B_i| \leq |S\backslash N(v)| \leq q + 9 - |S \cap N(v)|$. Totally, we have

$$|\bigcup_{1\leq i \leq t} C_i| = |\bigcup_{1\leq i \leq t} (C_i \backslash B_i)| + |\bigcup_{1\leq i \leq t} (B_i \cap C_i)| \leq t(q + 9 - |S \cap N(v)|) + q + 9.$$

By definitions of C_i, all vertices in $\cup_{1\leq i \leq t}(N_i \backslash C_i)$ are contained in some C_4. Putting the above inequalities together with (18), we have

$$8q \geq |\bigcup_{1\leq i \leq t} (N_i \backslash C_i)| \geq |\bigcup_{1\leq i \leq t} N_i| - |\bigcup_{1\leq i \leq t} C_i| \geq t(|N(v) \cap S| - 9) - 3q - 9$$
$$= (d(v) - |N(v) \cap S|) \cdot (|N(v) \cap S| - 9) - 3q - 9.$$

Note that $d(v) \geq q - 17$ and q is large. Solving the above inequality, it is easy to infer that $|N(v) \cap S| \leq 20$ or $|N(v) \cap S| \geq q - 28$. $\square$

We point out that there is at most one vertex (say z if it exists) in V with $|N(z) \cap S| \geq q - 28$. Indeed, if there are z_1, z_2 with $|N(z_i) \cap S| \geq q - 28$ for $i \in [2]$, as $|S| \leq q + 9$, we see z_1, z_2 have at least $q - 65$ common neighbors in S, which would give at least $\binom{q-65}{2} \gg 2q$ copies of C_4. Hence, any vertex in $V\backslash\{z\}$ has at most 20 neighbors in S.

Let $c'(v)$ denote the number of vertices $x \in N(v)$ with $c(x) \geq 1$. We also let $W = S_{q+2} \cup S_{q+3} \cup \{z\}$. So $|W| \leq 6$.

Claim 8.7. *If $\{u, v\}$ is an opposite pair with $u \in V\backslash W$ and $v \in V\backslash\{z\}$, then $c(u) + 19 \cdot c'(v) \geq q - 740$.*

Proof. We have $|S \cap N(u)| \leq 20$ and $|S \cap N(v)| \leq 20$. Let $d(u) = q + 1 - a$. Then $0 \leq a \leq 18$, $f(u) = a$ and $f(N(u)) \leq |S \cap N(u)| + 17 \leq 37$. Let V_u be the set of vertices $x \in V$ with $d(x, u) = 0$. By Claim 8.2, we get $|V_u| = d_0(u) \leq c(u) + aq + f(N(u)) \leq c(u) + aq + 37$.

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 $N_3 \subseteq B$. Take 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 there are at least $|N(x)\backslash V_u| + 1 = d(x) - |N(x) \cap V_u| + 1 \geq (q + 1) - a + 1$ edges wy with $w \in N(x)$ and $y \in N(u)$. Since $|N(u)| = q + 1 - a$, there is a 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 find a C_4 containing x. This shows $N_3 \subseteq B$.

Note that $v \notin V_u$. Since 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 $x \in N_4 \cap A$. For any $x \in N_4 \cap A$, we also have $|N(x) \cap V_u| \geq a+1$. Since $a \geq 0$, $|N_4 \cap A| \leq \frac{|V_u|}{a+1} \leq \frac{c(u)+aq+37}{a+1}$. This shows

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

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. $\square$

Now we choose the integer $\ell = 900$ such that $0.01\ell q - \binom{\ell}{2} \cdot 2\sqrt{q} > 8q$.

Claim 8.8. *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. Suppose $c(u_i) \leq 0.8q$ for all $i \in [\ell]$. For each $i \in [\ell]$, as the pair $\{u_i, v_i\}$ satisfies the conditions of Claim 8.7, 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 $\cup_{i \in [\ell]} N(v_i)$ 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 (18). $\square$

The next claim shows that deleting just a constant number of edges will result in a subgraph which contains a bounded number of 4-cycles.

Claim 8.9. *There exists an edge set $E^* \subseteq E(G)$ with $|E^*| \leq 105$ such that $G' = G - E^*$ has at most $0.1q$ copies of C_4.*

Proof. Let $A = \{v \in V : c(v) > 0.8q\}$, and let $X = A \cup W$. Then we have that $|X| < \frac{8q}{0.8q} + 6 = 16$. Let $E^* = E(G[X])$. Then $|E^*| \leq \binom{15}{2} = 105$. We next show that E^* is the edge set we wanted.

Suppose that $G' = G - E^*$ has more than $0.1q$ copies of C_4. Let the set of 4-cycles in G' be $\mathcal{C}$. Suppose first that there exists a vertex $x \in X$ contained in more than $0.001q$ copies of C_4 in $\mathcal{C}$. Since each copy of them offers an opposite pair (u_i, v_i) with $u_i, v_i \notin X$ and these opposite pairs span at least $\sqrt{0.001q} > \ell$ vertices in $V \backslash X$, we may choose ℓ opposite pairs among them say $\{u_i, v_i\}$ for $i \in [\ell]$ such that all v_i are distinct. By Claim 8.9, there is a vertex $u \in V \backslash X$ with $c(u) > 0.8q$ which contradicts the definition of X. Hence we may assume that every $x \in X$ is contained in at most $0.001q$ copies of C_4 in $\mathcal{C}$. Since $|X| \leq 15$, there are at least $0.085q$ copies of C_4 in $\mathcal{C}$ disjoint with X. These C_4's span at least $\sqrt[4]{8 \times 0.085q} > \ell$ vertices in $V \backslash X$. Using Claim 8.9 again, we can easily find a vertex $u \in V \backslash X$ with $c(u) > 0.8q$, which contradicts the definition of X. $\square$

By Claim 8.9, we know that $e(G') \geq e(G) - 105$ and G' has at most $0.1q$ copies of C_4. We further define a graph G'' to be obtained from G' by deleting one edge from each 4-cycle of G'. Thus we have

$$e(G'') \geq e(G') - 0.1q \geq q(q+1)^2/2 - 0.1q - 104.$$

It is clear that G'' is C_4-free. By our stability theorem (Theorem 3.1), as $c = 0.21q$ holds when q is even with $q \geq 10^{12}$, there exists a unique polarity graph H of order q such that $G'' \subseteq H$.

We claim that $G' \subseteq H$. Suppose for a contradiction that $G' \nsubseteq H$. Then there exists an edge $e \in E(G') \backslash E(G'')$ such that $e \notin E(H)$. By Lemma 2.8, there are at least $q - 1$ copies of C_4 in $H + e$, any two of which share e as their unique common edge. We note that $e(H) \leq q(q+1)^2/2$ and $G'' + e \subseteq H + e$. Since $G'' + e \subseteq G'$ 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 $G'' + e$. Hence

$$e(G'') = e(G'' + e) - 1 \leq e(H + e) - (0.9q - 1) - 1 \leq q(q+1)^2/2 - 0.9q + 1,$$

which contradicts the above lower bound on $e(G'')$. This proves that $G' \subseteq H$.

Suppose there are at least three edges from E^* which are not in H, say $e, e', e'' \in E^* \backslash E(H)$. By Lemma 2.8, there are $3(q-1)$ distinct copies of C_4 in $H + \{e, e', e''\}$, $q - 1$ 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 see each of edges in H can appear in at most three of these $3(q-1)$ cycles. We have $G' + \{e, e', e''\} \subseteq H + \{e, e', e''\}$ and $e(H + \{e, e', e''\}) - e(G' + \{e, e', e''\}) \le (e(G)+2) - (e(G')+3) \le |E^*| - 1 \le 104$. So $G' + \{e, e', e''\}$ can be obtained from $H + \{e, e', e''\}$ by deleting at most 104 edges. This shows that $G \supseteq G' + \{e, e', e''\}$ has at least $3(q-1) - 312 > 2q$ copies of C_4, a contradiction to (18).

Therefore we have $|E^* \backslash E(H)| \le 2$. Also we have $|E(G \cap H)| + |E^* \backslash E(H)| = |E(G)| \ge |E(H)| + 1$. This shows that $1 \le |E^* \backslash E(H)| \le 2$ and by Proposition 2.6, $e(H) = \frac{1}{2}q(q+1)^2$ and thus H is orthogonal. If $|E^* \backslash E(H)| = 1$, then we can derive from the above that G is a graph obtained from H by adding one new edge, as desired. Hence we have $|E^* \backslash E(H)| = 2$ 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'. First suppose that both e, e' are new edges between vertices of degree q. If e, e' are independent, then by Lemma 2.8, $H + \{e, e'\}$ has exactly $2(q-1)$ copies of C_4 and by Proposition 2.9, any edge in H is contained in at most one copy of these 4-cycles. This shows that G contains $2q - 2$ or $2q - 3$ copies of C_4. Now let e, e' share an endpoint. In this case, we can derive that $H + \{e, e'\}$ has exactly $2q - 1$ copies of C_4 and then G contains at least $2q - 3$ copies of C_4. It remains to consider the case that at least one of the endpoints of e and e' has degree $q + 1$. By Lemma 2.8, $H + \{e, e'\}$ has at least $2q - 1$ copies of C_4, at most two of which contain the edge e''. This proves that G contains at least $2q - 3$ copies of C_4, finishing the proof of Theorem 1.5. ∎

We are ready to prove Corollary 1.6.

Proof of Corollary 1.6. Let $q = 2^k$ where $k \ge 40$. Since $q \ge 2^{40} \ge 10^{12}$, by Theorem 1.1 and Theorem 1.5, we can immediately get the result. ∎

8.2 Strengthening Theorem 1.5

Now we show that using roughly the same proof, Theorem 1.5 can be strengthened as the following. Clearly this implies Theorem 1.7.

Theorem 8.1. *Let q be a large even integer and G be a graph on $q^2 + q + 1$ vertices with $\frac{1}{2}q(q+1)^2 + 1$ edges. Then either G has at least $q^{9/8}/30$ copies of C_4, or there exists an orthogonal polarity graph H of order q such that $|E(G) \backslash E(H)| = s$ and $|E(H) \backslash E(G)| = s - 1$ for some $1 \le s \le q^{1/8}/30$. In the latter case, the number of copies of C_4 in G is between $sq - s^2$ and $sq + s^2$.*

Proof. We will prove this by following the proof of Theorem 1.5 closely. It should be mentioned that most claims there can be generalized in this setting easily and thus, in these cases, we often only mention the modified statements without providing many details. For other cases where extra arguments are needed, we give self-contained proofs.

Throughout this proof, let $t = q^{1/8}/30$ and similarly, we may assume that

$$\#C_4 < tq \text{ and } \sum_{v \in V} c(v) = 4 \cdot \#C_4 < 4tq. \tag{22}$$

We retain Claim 8.1 and Claim 8.2 as they are. By adjusting the proof of Claim 8.3, one can derive $\Delta(G) \le q + 2 + t$. Indeed, suppose on the contrary that there is a vertex $v_1 \in V$ with $d(v_1) = q + k$ for some $t + 3 \le k \le q^2$. Following the proof therein, we can get $\sum_{i=2}^{n} a_i \le q^2 + (t+1)q$ and

$$2 \cdot \#C_4 \ge q^3 - 2q^2 + 2q - (q-1)\sum_{i=2}^{n} a_i - 0.5k^2 + k(q^2 - q + 1.5) - 2.$$

So we see that when $k = t + 3$ and $\sum_{i=2}^{n} a_i = q^2 + (t+1)q$, the above inequality achieves its minimum $q^2 - 0.5t^2 - 1.5t - 2$, which is a contradiction to (22) for large q. This proves $\Delta(G) \le q + 2 + t$. We then see that the deficiency of V is

$$f(V) = (q+1)n + \sum_{k=0}^{t}(k+1)|S_{q+2+k}| - 2e(G) = q - 1 + \sum_{k=0}^{t}(k+1)|S_{q+2+k}|. \tag{23}$$

31

Let $S' = S_{q+2} \cup \cdots \cup S_{q+2+t}$. By a similar proof as Claim 8.4, one can show $|S'| \le 3\sqrt{tq}$. Following the deductions after Claim 8.4, one can conclude that $q - 4t \le |S| \le q - 1 + 3(t+1)\sqrt{tq}$ and

$$\text{for any } T \subseteq V, \ f(T) \le |S \cap T| + 3(t+1)\sqrt{tq} + 4t - 1. \tag{24}$$

We also can derive the following analogue of Claim 8.5.

Claim A. $\sum_{k=0}^{t}(k+1)|S_{q+2+k}| \le 2t + 2$.

Proof. For any $u, v \in V$, we have $d(u,v) \le 2\sqrt{tq}$; as otherwise there are at least $2tq$ copies of C_4, a contradiction to (22). Choose $v_1, ..., v_s \in S'$ for $s = \min\{4t, |S'|\}$. By the inclusion-exclusion principle,

$$|S| \ge \Big| \bigcup_{j=1}^{s}(N(v_j) \cap S)\Big| \ge \sum_{1 \le j \le s}|N(v_j) \cap S| - \sum_{1 \le i < j \le s}|N(v_i) \cap N(v_j)| \ge \sum_{1 \le j \le s}|N(v_j) \cap S| - 2\binom{s}{2}\sqrt{tq}$$

which implies that $\sum_{1 \le j \le s}|N(v_j) \cap S| \le q + (3t + s^2 + 3)\sqrt{tq} - 1$. Since each C_4 has two opposite pairs, we see there are at least $\frac{1}{2}\sum_{j=1}^{s}\sum_{i=s+1}^{n}\binom{d(v_j,v_i)}{2}$ copies of C_4 containing opposite pairs $\{v_j, v_i\}$ for $1 \le j \le s$ and $s + 1 \le i \le n$. For any $j \in [s]$, by (19) and (24) we derive

$$\sum_{i=s+1}^{n}\binom{d(v_j,v_i)}{2} \ge \sum_{i=s+1}^{n}(d(v_j,v_i) - 1) \ge \sum_{i \in [n]\setminus\{j\}}(d(v_j,v_i) - 1) - 2(s-1)\sqrt{tq}$$

$$\ge (d(v_j) - q - 1)q - f(N(v_j)) - 2(s-1)\sqrt{tq} \ge (d(v_j) - q - 1)q - |N(v_j) \cap S| - X',$$

where $X' = (3t + 2s + 1)\sqrt{tq} + 4t - 1$. Putting the above together, we have that

$$2tq > 2\#C_4 \ge \sum_{j=1}^{s}\sum_{i=s+1}^{n}\binom{d(v_j,v_i)}{2} \ge \sum_{j=1}^{s}(d(v_j) - q - 1)q - \sum_{j=1}^{s}|N(v_j) \cap S| - sX'$$

$$\ge q\sum_{k=0}^{t}(k+1)|S_{q+2+k} \cap \{v_1, ..., v_s\}| - \big(q + (3t + s^2 + 3)\sqrt{tq} - 1\big) - sX'.$$

Since $s/4 \le t \le q^{1/8}/30$ and q is large, we can derive $\sum_{k=0}^{t}(k+1)|S_{q+2+k} \cap \{v_1, ..., v_s\}| \le 2t + 2$. This shows that $s \le 2t + 2$ and by the choice of s, we have $S' = \{v_1, ..., v_s\}$. Therefore $\sum_{k=0}^{t}(k+1)|S_{q+2+k}| \le 2t + 2$, completing the proof of Claim A. $\qquad \square$

Using Claim A and (23), we can easily deduce that $q - 4t \le |S| \le f(V) \le q + 2t + 1$ and for any $T \subseteq V$, $f(T) \le |S \cap T| + 6t + 1$. In particular, for any $v \in V$, we have $f(v) \le 6t + 2$ and $d(v) \ge q + 1 - f(v) \ge q - 6t - 1$.

Following the arguments in the proof of Claim 8.6, we can show that for any $v \in V$, either $|N(v) \cap S| \le 7t + 2$ or $|N(v) \cap S| \ge q - 11t - 2$. Again we note that there is at most one vertex (denoted by z if it exists) in V with $|N(z) \cap S| \ge q - 11t - 2$. Let

$$W = S' \cup \{z\}.$$

Claim A shows $|W| \le 2t + 3$. Then one can generalize Claim 8.7 into the following: if $\{u, v\}$ is an opposite pair with $u \in V\setminus W$ and $v \in V\setminus\{z\}$, then $c(u) + (6t+3)\cdot c'(v) \ge q - (13t+3)(6t+4) > q - 200t^2$.

Choose $\ell = q^{1/4}/2$ such that $\ell q/(100t) - \binom{\ell}{2}\cdot 2\sqrt{tq} > 4tq$ holds for large q. Then Claim 8.8 can be extended as following.

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

Let $A = \{v \in V : c(v) > 0.8q\}$ and $X = A \cup W$. So $|X| < 4tq/(0.8q) + 2t + 3 = 7t + 3$. One can easily obtain the following analogue of Claim 8.9.

Claim C. There exists an edge set $E^* \subseteq E(G)$ with $|E^*| \leq 45t^2$ such that $G' = G - E^*$ has at most $0.1q$ copies of C_4.

Having the above, the arguments after Claim 8.9 can be directly converted to show that there exists an orthogonal polarity graph H of order q such that $G' \subseteq H$. Let $|E(G)\backslash E(H)| = s$. Then $|E(H)\backslash E(G)| = s - 1$ and $1 \leq s \leq |E(G)\backslash E(G')| = |E^*| \leq 45t^2$. By Lemma 2.8, there are at least $s(q - 1)$ copies of C_4 in $H \cup (E(G)\backslash E(H))$ such that each edge in $E(H)\backslash E(G)$ can appear in at most s of them. This shows that G has at least $s(q - 1) - (s - 1)s = sq - s^2$ copies of C_4. If $s \geq t + 1$, then since $s \leq 45t^2$, G has at least tq copies of C_4, contradicting our assumption. Hence we have $1 \leq s \leq t$.

It remains to show that the number L of copies of C_4 in G is at most $sq + s^2$. Let $\mathcal{C}_0$ be the collection of all C_4's in G using exactly one edge in $E(G)\backslash E(H)$ and $\mathcal{C}_1$ be the collection of the remaining C_4's in G. By Lemma 2.8, we have $|\mathcal{C}_0| \leq s(q + 1)$. There are three types of C_4's in $\mathcal{C}_1$, namely using two edges, three edges or four edges in $E(G)\backslash E(H)$. For each 4-cycle C in $\mathcal{C}_1$, we define one or two pairs of edges in $(E(G) \cap E(C))\backslash E(H)$ as following. If C has exactly two edges in $E(G)\backslash E(H)$, then we take these two edges to form a pair. If C has three edges in $E(G)\backslash E(H)$ which form a path of length three, then we take the two non-incident edges to form a pair. Otherwise all four edges in C are from $E(G)\backslash E(H)$, then we take two pairs, each of which is formed by two non-incident edges of C. Let us call all such pairs *feasible*. It is easy to verify that each feasible pair can be contained in at most two 4-cycles in $\mathcal{C}_1$. Thus we have $|\mathcal{C}_1| \leq 2\binom{s}{2}$, implying that $L = |\mathcal{C}_0| + |\mathcal{C}_1| \leq s(q + 1) + s(s - 1) = sq + s^2$. This completes the proof of Theorem 8.1. ∎

9 Supersaturation: the general case

In this section, we establish two supersaturation results - Theorems 9.1 and 9.2, which together imply Theorem 1.8. We also give a proof of Proposition 1.9.

9.1 A generalization of Theorem 1.5

We now deduce a supersaturation result for a wider range on the number of edges from Theorem 8.1. This is also optimal for infinitely many values q (as powers of two).

Theorem 9.1. *Let q be a large even integer and t be any integer such that $1 \leq t < q^{1/8}/30$. Let G be a graph on $q^2 + q + 1$ vertices with $\frac{1}{2}q(q + 1)^2 + t$ edges. Then either G has at least $(t + 1)q - (t + 1)^2$ copies of C_4, or G is obtained from an orthogonal polarity graph of order q by adding t new edges. In particular, G has at least $t(q - 1)$ copies of C_4.*

Proof. Suppose on the contrary that G has less than $(t+1)q - (t+1)^2$ copies of C_4. We denote G' to be any spanning subgraph of G with $\frac{1}{2}q(q+1)^2 + 1$ edges. Thus G' has less than $(t+1)q - (t+1)^2 < q^{9/8}/30$ copies of C_4. By Theorem 8.1, there exists an orthogonal polarity graph H of order q such that $|E(G')\backslash E(H)| = s$ and $|E(H)\backslash E(G')| = s - 1$ for some $1 \leq s \leq t$.

Let $|E(G)\backslash E(H)| = j$. Then $j - t = |E(H)\backslash E(G)| \leq |E(H)\backslash E(G')| = s - 1 \leq t$, implying that $t \leq j \leq 2t$. So G has at least $j(q - 1) - (j - t)j$ copies of C_4. If $j \geq t + 1$, then G has at least $(t + 1)(q - 1) - (t + 1)$ copies of C_4, a contradiction. So $j = t$ and G is obtained from H by adding t new edges. By Lemma 2.8, G has at least $t(q - 1)$ edges. This finishes the proof of Theorem 9.1. ∎

9.2 A half-way bound

Theorem 9.2. *Let q be a positive even integer. If G is a graph on $q^2 + q + 1$ vertices with $\frac{1}{2}q(q+1)^2 + t$ edges for $t \geq 1$, then G contains at least $\frac{1}{2}(tq - 2.5q - t)$ copies of C_4.*

Proof. As earlier, we let $\#C_4$ be the number of copies of C_4 in G, $c(v)$ be the number of C_4 containing the vertex v, and S_i be the set of vertices of degree i in G. We may assume that $\#C_4 \leq \frac{1}{2}tq - \frac{1}{2}q$ (as otherwise we have the desired number of C_4's).

Let $V = V(G) = \{v_1, ..., v_n\}$, where $n = q^2 + q + 1$.

Claim 9.1. *Let $v_i \in V$ be any vertex with $d(v_i) = q + 2 + k$ for some $k \geq 0$. Then $k \leq \frac{q}{2}$. Moreover, if $k = 0$ then $c(v_i) > t$, and otherwise $c(v_i) > kq$.*

Proof. Without loss of generality, we consider v_1 with $d(v_1) = q + 2 + k$ for $k \geq 0$. First suppose that $t \geq q + 1$. We have $\sum_{v \in V} d(v) = 2e(G) = (q^2 + q + 1)q + (q^2 + 2t)$, where $q^2 + 2t \geq -(q^2 + q + 1)$. By Claim 8.1 and Lemma 2.14, we see $2 \cdot \#C_4$ is at least

$$\sum_{v \in V} \binom{d(v)}{2} - \binom{n}{2} \geq (q^2 + q + 1)\binom{q}{2} + q(q^2 + 2t) - \binom{q^2 + q + 1}{2} = 2tq - q^2 - q \geq tq.$$

This implies $\#C_4 \geq \frac{1}{2}tq$, contradicting our assumption. Hence we may assume that $t \leq q$.

Let P_2' be the set of 2-paths in G with none of its endpoints in $N(v_1)$. Since each C_4 contains two covered pairs, we see that $2 \cdot \#C_4 \geq \sum_{\{u,v\} \subseteq V \setminus N(v_1)} \binom{d(u,v)}{2} + c(v_1)$ is at least

$$\sum_{\{u,v\} \subseteq V \setminus N(v_1)} \big(d(u,v) - 1\big) + c(v_1) \geq |P_2'| - \binom{n - d(v_1)}{2} + c(v_1).$$

Let $a_i = |N(v_i) \cap N(v_1)| - 1$ for $2 \leq i \leq n$. Then we have

$$|P_2'| = \sum_{i=2}^{n} \binom{d(v_i) - a_i - 1}{2} \quad \text{and} \quad \sum_{i=2}^{n} a_i \leq \sum_{i=2}^{n} \binom{a_i + 1}{2} \leq c(v_1) \leq tq/2 \leq q^2/2.$$

On the other hand, we can derive that

$$\sum_{i=2}^{n} (d(v_i) - a_i - 1) = \big(2e(G) - d(v_1)\big) - \sum_{i=2}^{n} (a_i + 1) = (q^2 + q)(q - 1) + X.$$

Here $X = q^2 + 2t - 2 - k - \sum_{i=2}^{n} a_i \geq q^2 + 2t - 2 - k - c(v_1) \geq -(q^2 + q)$, where the inequalities hold because $1 \leq t \leq q$, $k \leq q^2 - 2$ and $\sum_{i=2}^{n} a_i \leq c(v_1) \leq q^2/2$. Putting the above all together, by Lemma 2.14, we infer that $tq - q \geq 2 \cdot \#C_4$ is at least

$$|P_2'| - \binom{n - d(v_1)}{2} + c(v_1) \geq (q^2 + q)\binom{q - 1}{2} + (q - 1)X - \binom{q^2 - k - 1}{2} + c(v_1)$$

$$\geq k(q^2 - q) - \frac{1}{2}(k^2 + k) - c(v_1)(q - 2) + 2tq - q - 2t + 1.$$

Simplifying the above one can derive $c(v_1) \geq \frac{1}{q-2}\big(k(q^2 - q) - \frac{1}{2}(k^2 + k) + t(q - 2) + 1\big)$ and

$$c(v_1) - \frac{1}{2}tq \geq \frac{1}{q - 2}\Big(k(q^2 - q) - \frac{1}{2}(k^2 + k) - \frac{1}{2}tq^2 + 2t(q - 1) + 1\Big) := \frac{g(k, t)}{q - 2}.$$

Suppose $\frac{q}{2} < k \leq q^2 - 2$. Then $\frac{\partial g}{\partial t}(k, t) = -\frac{1}{2}(q - 2)^2 < 0$ and $\frac{\partial^2 g}{\partial k^2}(k, t) = -1 < 0$. This implies $\#C_4 - \frac{1}{2}tq \geq c(v_1) - \frac{1}{2}tq \geq \min\{g(\frac{q}{2}, q)/(q - 2), g(q^2 - 2, q)/(q - 2)\} > 0$, a contradiction. Therefore we obtain that $k \leq \frac{q}{2}$.

If $k = 0$, then $c(v_1) \geq (t(q - 2) + 1)/(q - 2) > t$, as desired. We also have

$$c(v_1) - kq \geq \frac{1}{q - 2}\Big(kq - \frac{1}{2}(k^2 + k) + t(q - 2) + 1\Big) := \frac{h(k, t)}{q - 2}.$$

Consider $1 \leq k \leq \frac{q}{2}$. Since $\frac{\partial h}{\partial t}(k, t) = q - 2 > 0$ and $\frac{\partial h}{\partial k}(k, t) = q - k - \frac{1}{2} > 0$, we have $c(v_1) - kq \geq h(1, 1)/(q - 2) > 0$. This proves the claim. $\square$

Let $\Delta(G) = q + 2 + m$. By Claim 9.1, we see that $m \leq \frac{q}{2}$.

Claim 9.2. $|S_{q+2}| < 2q$ *and* $\sum_{k=1}^{m} k|S_{q+2+k}| < 2t$.

Proof. By Claim 9.1, we see that if $v \in S_{q+2}$, then $c(v) > t$. So $2tq > 4\#C_4 > t|S_{q+2}|$, implying that $|S_{q+2}| < 2q$. For $v \in S_{q+2+k}$ with $1 \le k \le m$, by Claim 9.1 we have $c(v) > kq$. Then $2tq > 4\#C_4 > \sum_{k=1}^m kq \cdot |S_{q+2+k}|$, which implies that $\sum_{k=1}^m k|S_{q+2+k}| < 2t$. $\qquad \square$

Now we are ready to complete the proof of Theorem 9.2. First let us estimate $|UP|$, i.e., the number of uncovered pairs in G. Consider a vertex $v \in S_{q+1}$, where the maximum degree of the vertices in $N(v)$ is at most $q+1$. We assert that v is contained in at least one uncovered pair. Otherwise, as q is even, the number e_v of edges in $G[N(v)]$ is at least $q/2 + 1$ and thus there are at most $(q+1)q - 2e_v \le q^2 - 2$ vertices adjacent to $N[v]$, which again forces an uncovered pair containing v. For $1 \le k \le q+1$, if $v \in S_{q+1-k}$ and the maximum degree of the vertices in $N(v)$ is at most $q+1$, we can get at least

$$(n-1) - \sum_{u \in N(v)} (d(u) - 1) = (q^2 + 2q + 1 - k) - \sum_{u \in N(v)} d(u) \ge kq$$

uncovered pairs containing v. If some vertex in $N(v)$ has degree $q + 1 + \ell$ for $\ell \ge 1$, then the above number of uncovered pairs containing v will decrease by ℓ. Thus, by double-counting, we can get

$$\begin{aligned}
|UP| &\ge \frac{1}{2}\Big(|S_{q+1}| + \sum_{k=1}^{q+1} kq|S_{q+1-k}| - \sum_{u \in V}\sum_{k=1}^{m+1} k|N(u) \cap S_{q+1+k}|\Big) \\
&= \frac{1}{2}\Big(|S_{q+1}| + \sum_{k=1}^{q+1} kq|S_{q+1-k}| - \sum_{k=1}^{m+1} k(q+1+k)|S_{q+1+k}|\Big).
\end{aligned} \tag{25}$$

Next, we give a lower bound on $|P_2| - \binom{n}{2}$. Let $S' = S_{q+2} \cup \cdots \cup S_{q+2+m}$. We have

$$\sum_{v \in V \setminus S'} d(v) = 2e(G) - \sum_{k=1}^{m+1} (q+1+k)|S_{q+1+k}| = \Big(n - \sum_{k=1}^{m+1} |S_{q+1+k}|\Big)q + Y.$$

Here $Y = 2t + q^2 - \sum_{k=1}^{m+1} (1+k)|S_{q+1+k}| \ge 2t + q^2 - 2|S_{q+2}| - 3\sum_{k=1}^m k|S_{q+2+k}| \ge q^2 - 4q - 6t > -(n-|S'|)$, where the second inequality holds because of Claim 9.2. By Lemma 2.14,

$$\begin{aligned}
\sum_{v \in V \setminus S'} \binom{d(v)}{2} - \binom{n}{2} &\ge \Big(q^2 + q + 1 - \sum_{k=1}^{m+1} |S_{q+1+k}|\Big)\binom{q}{2} + qY - \binom{q^2 + q + 1}{2} \\
&= 2tq - q^2 - q - \sum_{k=1}^{m+1}\Big(\binom{q}{2} + qk + q\Big)|S_{q+1+k}|.
\end{aligned}$$

Recall that P_2 is the set of all 2-paths in G. We have

$$\begin{aligned}
|P_2| - \binom{n}{2} &= \sum_{k=1}^{m+1} |S_{q+1+k}|\binom{q+1+k}{2} + \sum_{v \in V \setminus S'} \binom{d(v)}{2} - \binom{n}{2} \\
&\ge 2tq - q^2 - q + \frac{1}{2}\sum_{k=1}^{m+1} k(k+1)|S_{q+1+k}|.
\end{aligned} \tag{26}$$

Let $M = 2tq - q^2 - q + \frac{1}{2}|S_{q+1}|$. Combining (25) and (26), by Claim 8.1 we derive that

$$\begin{aligned}
2\#C_4 &\ge |P_2| + |UP| - \binom{n}{2} \ge M - \frac{q}{2}\Big(\sum_{k=1}^{m+1} k|S_{q+1+k}| - \sum_{k=1}^{q+1} k|S_{q+1-k}|\Big) \\
&= M - \frac{q}{2}\big(2e(G) - (q+1)n\big) = tq - \frac{1}{2}(q^2 + q - |S_{q+1}|).
\end{aligned} \tag{27}$$

Finally, by Claim 9.2 we have

$$q^2 + q - |S_{q+1}| = \sum_{k=1}^{q+1} |S_{q+1-k}| + |S_{q+2}| + \sum_{k=1}^{m} |S_{q+2+k}| - 1$$

$$\leq \sum_{k=-(q+1)}^{m+1} \Big((q+1) - (q+1+k)\Big) \cdot |S_{q+1+k}| + 2|S_{q+2}| + 2\sum_{k=1}^{m} k|S_{q+2+k}| - 1$$

$$\leq (q+1)n - 2e(G) + 4q + 4t - 1 = 5q + 2t.$$

This together with (27) show that $\#C_4 \geq \frac{1}{2}(tq - 2.5q - t)$. The proof of Theorem 9.2 is finished. ∎

We point out that Theorem 9.2 only works for $t \geq 3$ and becomes invalid when $t \in \{1, 2\}$.

9.3 Proofs of Theorem 1.8 and Proposition 1.9

Before presenting the proofs, we show a upper bound on $h(q^2 + q + 1, t)$ for any prime power q and $t \geq 1$, using a random construction based on polarity graphs.

Lemma 9.3. *Let q be a prime power and t be an integer such that $4t \leq q^3(q+1)$. Then there exists a graph on $q^2 + q + 1$ vertices, which contains at least $\frac{1}{2}q(q+1)^2 + t$ edges and at most $500(tq + t^4/q^8)$ copies of C_4.*

Proof. We may assume $q \geq 3$ and $t \geq 1$. Let H be an orthogonal polarity graph on $n = q^2 + q + 1$ vertices. Let $\alpha = \frac{4t}{q^3(q+1)} \in [0, 1]$ and let G be obtained from H by adding an edge for each non-adjacent pair of vertices independently and randomly with probability α. Denote by X the number of new edges added to H. Since the number of non-adjacent pairs in H is $N = \binom{n}{2} - e(H) = \frac{q^3(q+1)}{2}$, we have $E[X] = N\alpha = 2t$. Here, X is a binomial random variable $X \sim Bin(N, \alpha)$. Then the Chernoff bound states that $P(X < (1-\epsilon)N\alpha) \leq e^{-\epsilon^2 N\alpha/2}$. Choosing $\epsilon = 1/2$, we can get that $P(X < t) \leq e^{-\frac{t}{4}} \leq 0.78$.

Let Y be the number of copies of C_4's in G. For $1 \leq i \leq 4$, let Y_i be the number of copies of C_4's in G consisting of exactly i new edges. We estimate $E[Y] = \sum_{i=1}^{4} E[Y_i]$ as follows. Note that every vertex in H has degree q or $q+1$. For $E[Y_1]$, each of these C_4's corresponds to a unique path of length three in H. Thus we have $E[Y_1] \leq \frac{1}{2}n(q+1)q^2 \cdot \alpha \leq 3tq$. For $E[Y_2]$, each of these C_4's contains two edges in H which are incident or not. Thus we have $E[Y_2] \leq \frac{1}{2}n^2(q+1)q \cdot \alpha^2 + \binom{e(H)}{2} \cdot 2\alpha^2 \leq 20t^2/q^2$. Similarly, we can get that $E[Y_3] \leq e(H)\binom{n}{2} \cdot 2\alpha^3 \leq 60t^3/q^5$ and $E[Y_4] \leq n^4\alpha^4/8 \leq 50t^4/q^8$. Since $t^2/q^2 + t^3/q^5 \leq tq + t^4/q^8$, we can get that $E[Y] = \sum_{i=1}^{4} E[Y_i] \leq 110(tq + t^4/q^8)$.

Since $P(X \geq t) \geq 0.22$, we have $E[Y] = P(X \geq t) \cdot E[Y|X \geq t] + P(X < t) \cdot E[Y|X < t] \geq 0.22 \cdot E[Y|X \geq t]$. So $E[Y|X \geq t] \leq E[Y]/0.22 \leq 500(tq + t^4/q^8)$. This shows that there exists an n-vertex graph with at least $\frac{1}{2}q(q+1)^2 + t$ edges and at most $500(t\sqrt{n} + t^4/n^4)$ copies of C_4's. ∎

In aid of Lemma 9.3, we are ready to derive Theorem 1.8 from Theorems 9.1 and 9.2.

Proof of Theorem 1.8. Let $q = 2^k$ be sufficiently large. So $ex(q^2 + q + 1, C_4) = \frac{1}{2}q(q+1)^2$.

First we consider (A). Suppose that $1 \leq t < q^{1/8}/30$. Let G be a $(q^2 + q + 1)$-vertex graph with $ex(q^2 + q + 1, C_4) + t$ edges. By Theorem 9.1 and Lemma 2.8, G has at least $t(q-1)$ copies of C_4, with equality only if G is obtained from an orthogonal polarity graph of order q by adding t edges between vertices of degree q. As any two vertices of degree q in a polarity graph has a common neighbor, it is straightforward to see that when G has exactly $t(q-1)$ copies of C_4, these aforementioned t new edges must form a matching (in fact it is also an induced matching by Lemma 2.7). This proves (A).

For the first assertion of (B), it suffices to consider when $t \geq q^{1/8}/30$ and this follows from Theorem 9.2 that $h(q^2 + q + 1, t) \geq \frac{1}{2}(tq - 2.5q - t) = \left(\frac{1}{2} + o(1)\right)tq$, where $o(1) \to 0$ as $q \to \infty$.

Finally we prove the second assertion of (B) that $h(q^2 + q + 1, t) = \Theta(tq + t^4/q^8)$. It is well known (see [16]) that for any $c > 0$ there exists some $c' > 0$ such that $h(q^2 + q + 1, t) \geq c' \cdot t^4/q^8$ for any $t \geq cq^3$. Also as q is large, from the above proof, we have $h(q^2 + q + 1, t) \geq \left(\frac{1}{2} + o(1)\right)tq \geq tq/3$ for

any $t \geq 1$. Note that $tq \geq t^4/q^8$ if and only if $t \leq q^3$. Combining the above all together, we see that there exists some absolute constant $d > 0$ such that for any $t \geq 1$, $h(q^2 + q + 1, t) \geq d \cdot (tq + t^4/q^8)$. The upper bound easily follows from Lemma 9.3. The proof of Theorem 1.8 is completed. $\blacksquare$

We now complete the proof of Proposition 1.9.

Proof of Proposition 1.9. Let n be sufficiently large and $t \geq 3n^{1.2625}$. By Proposition 7.3, there exists some prime p with $\sqrt{n} - n^{0.2625} - 1 \leq p \leq \frac{1}{2}(-1 + \sqrt{4n - 3})$ such that $\mathrm{ex}(n, C_4) \geq \mathrm{ex}(p^2 + p + 1, C_4) \geq (n^{1.5} - 3n^{1.2625} + n)/2$. We first consider the lower bound of $h(n, t)$. Consider any n-vertex graph G with $\mathrm{ex}(n, C_4) + t$ edges. Let s be such that $\mathrm{ex}(n, C_4) + t = \frac{1}{2}(n^{1.5} + n) + s$. So $s \geq t - \frac{3}{2}n^{1.2625} \geq 0.5t$. By Jensen's inequality, we see that the number M of 2-paths in G is at least

$$\sum_{v \in V(G)} \binom{d(v)}{2} \geq n\binom{\frac{\sum d(v)}{n}}{2} = n\binom{\frac{n^{1.5}+n+2s}{n}}{2} \geq \binom{n}{2} + 2s\sqrt{n} + \frac{2s^2}{n}.$$

As $s \geq 0.5t$, this implies that $h(n, t) \geq \#C_4(G) \geq \frac{1}{2}\left(M - \binom{n}{2}\right) \geq \frac{1}{2}\left(\sum_{v \in V(G)} \binom{d(v)}{2} - \binom{n}{2}\right) \geq \frac{t\sqrt{n}}{2}$. It is known that if $t \geq n^{3/2}$, then $h(n, t) \geq c \cdot t^4/n^4$ holds for some $c > 0$. Combining these facts, we infer $h(n, t) = \Omega(t\sqrt{n} + t^4/n^4)$ for $t \geq 3n^{1.2625}$. For the upper bound of $h(n, t)$, let r be the integer such that $\mathrm{ex}(n, C_4) + t = \frac{1}{2}p(p+1)^2 + r$. By our choice of p, we can derive that $r \leq t + \frac{3}{2}n^{1.2625} \leq 2t$. Using Lemma 9.3, there exists a graph on $p^2 + p + 1$ vertices with $p(p+1)^2/2 + r = \mathrm{ex}(n, C_4) + t$ edges and at most $O(rp + r^4/p^8) = O(t\sqrt{n} + t^4/n^4)$ copies of C_4's. Since $p^2 + p + 1 \leq n$, this also shows that $h(n, t) = O(t\sqrt{n} + t^4/n^4)$, finishing the proof of Proposition 1.9. $\blacksquare$

10 Concluding remarks

In this paper, we focus on extremal problems of 4-cycles and prove several stability and supersaturation theorems. These imply some exact or near-optimal extremal results on C_4 for infinite instances. In what follows we discuss related problems, some of which in fact partially motivate the results here.

Theorem 1.2 provides a stability type result for dense C_4-free graphs G on $q^2 + q + 1$ vertices where q is even. It states that if $e(G) \geq \frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)$, then G is contained in some (unique) polarity graph of order q. We wonder if some other form of stability (e.g., in the sense of "edit distance", which counts edges adding and deleting between G and the extremal configuration) can hold for a much weaker condition on the number of edges. This stability also indicates that there exists some hierarchy on the number of edges for *maximal C_4-free graphs* G, in the interval starting from $\frac{1}{2}q(q+1)^2 - \frac{1}{2}q + o(q)$. This is because any of these graphs G must be some polarity graph and according to Proposition 2.6, $e(G) = \frac{1}{2}q(q+1)^2 - \frac{1}{2}m\sqrt{q}$ holds for some integer $m \geq 0$. We remark that using the result of Metsch [29] and similar arguments here, one also can establish an analogous stability result for C_4-free balanced bipartite graphs where the size of two parts equals $q^2 + q + 1$ for any (even or odd) large integer q. (As a side note here, we would like to mention that recently Nagy [31] proved some supersaturation results on 4-cycles in the bipartite setting.)

Arguably, Theorem 1.2 provides some (very weak) evidence to the following conjecture of McCuaig.

Conjecture 10.1 (McCuaig, 1985; see [19, 22])**.** *Each extremal graph which achieves the maximum number $\mathrm{ex}(n, C_4)$ is a subgraph of some polarity graph.*

On the other hand, one of the authors in [22] "strongly disagrees and he believes just the opposite that for e.g. $n = q^2 + q + 2$ maybe the extremal graphs are obtained by adding an extra vertex and some edges to a polarity graph".[4] Yet, if what is described as above is true, then there are at most two new edges which can be added to remain C_4-free (this is because any two non-adjacent vertices in a polarity graph must share a common neighbor). Following this suggestion from [22], we tend to believe that this indeed is the case for $n = q^2 + q + 2$ (at least for large $q = 2^k$).

---

[4]We quote this sentence from [22].

Conjecture 10.2. *For large $q = 2^k$, $\mathrm{ex}(q^2 + q + 2, C_4) = \frac{1}{2}q(q+1)^2 + 2$.*

We point out that there are many extremal graphs for this problem. Let H be an orthogonal polarity graph of order q and choose $v \in V(H)$ such that $N_H(v)$ is independent (for $q = 2^k$, any vertex of degree q or one special vertex of degree $q + 1$ satisfies this property; see [19]). Take any subsets A and B such that $A \cup B = N_H(v)$ and $A \cap B = \{z\}$ for some z. Let G be obtained from H by deleting v, adding two new vertices x, y and adding new edges in $\{xa : a \in A\} \cup \{yb : b \in B\} \cup \{xy\}$. It is easy to see that G is C_4-free and has $q^2 + q + 2$ vertices and $\frac{1}{2}q(q+1)^2 + 2$ edges. Let us note that this construction works as long as no edges are between A and B (so $N_H(v)$ need not be independent).

A possible approach to attack Conjecture 10.2 is to use the stability on $(q^2 + q + 1)$-vertex graphs. Observe that the minimum degree of all so-called extremal graphs G of $\mathrm{ex}(q^2 + q + 2, C_4)$ defined in the previous paragraph satisfies $2 \le \delta(G) \le \frac{q}{2} + 2$. If indeed in some extremal graph G, one can ensure a vertex u of degree at most $\frac{q}{2} + o(q)$, then $G - u$ becomes a $(q^2 + q + 1)$-vertex C_4-free graph with $e(G - u) \ge \frac{1}{2}q(q+1)^2 - \frac{q}{2} + o(q)$; so one can apply Theorem 1.2 on $G - u$ to obtain a host polarity graph, from which one may derive the structure of G and solve the conjecture. This is one of the reasons why we make serious efforts to prove the stability statement with the edge-bound $\frac{1}{2}q(q+1)^2 - \frac{q}{2} + o(q)$. However, having this in mind, we still think to find such a low-degree vertex requires new ideas. Another possible route for solving Conjecture 10.2 is to extend Theorem 1.2 so that the same conclusion holds for graphs G with $e(G) \ge \frac{1}{2}q(q+1)^2 - q + 3$ (this would be sharp for the stability statement, as we mentioned earlier that $e(G)$ cannot be lowered to $\frac{1}{2}q(q+1)^2 - q + 2$).

As indicated, Conjecture 10.2 may provide potential counterexamples to Conjecture 10.1. Another interesting consequence is that the resolution of Conjecture 10.2 would disprove the supersaturation conjecture of Erdős-Simonovits on C_4 (i.e., Conjecture 1.4), even in its weakest version.

Proposition 10.3. *If Conjecture 10.2 holds for $q = 2^k$, then $h(q^2 + q + 2) = 1$.*

Proof. Let G be the graph defined in the paragraph following Conjecture 10.2, for which $N_H(v)$ is independent. If Conjecture 10.2 holds, then G is an extremal graph. If we add an extra edge between y and any vertex $w \in A \setminus \{z\}$ (or symmetrically between x and any vertex in $B \setminus \{z\}$), this will only create one copy of C_4, namely, $xzywx$. This shows that $h(q^2 + q + 2) = 1$. □

We point out that a result of [38] implies that $h(q^2 + q + 2) = 1$ for $q = 4$. The above proposition has further implication to a question asked by Erdős [13] for general graphs.

Question 10.4 (Erdős [13]). *For which graphs G is it true that every graph on n vertices and $\mathrm{ex}(n, G) + 1$ edges contains at least two G's? Perhaps this is always true.*

We see that by definition itself, $h_G(n, 1) \ge 1$ holds for any G. It is natural to ask if $h_G(n, 1) \ge 2$ holds – this is the question of Erdős. By Proposition 10.3, we see that if Conjecture 10.2 holds, then C_4 would serve as a counterexample and refuse this.

Suggested by orthogonal polarity graphs, the following supersaturation problem seems plausible.

Question 10.5. *For large $q = 2^k$, is it true that $h(q^2 + q + 1, t) = t(q - 1)$ holds for every $1 \le t \le \frac{q}{2}$?*

Theorem 1.8 confirms this for $1 \le t \le O(q^{1/8})$, while its proof perhaps can be generalized further. Another problem is to determine all integers $t \ge 1$ such that every graph achieving the maximum $h(q^2 + q + 1, t)$ contains an orthogonal polarity graph of order q. Regarding to Proposition 1.9, a similar supersaturation result for C_4 perhaps can hold under a more general condition.

Question 10.6. *Is there a constant t_0 such that $h(n, t) = \Theta(t\sqrt{n} + t^4/n^4)$ holds whenever $t \ge t_0$?*

It is known [27] that every n-vertex graph with $\mathrm{ex}(n, K_3) + 1$ edges contains an edge that is in at least $n/6$ triangles, which is sharp. Inspired by this, one may ask the analogous question for C_4.

Question 10.7. *For large $q = 2^k$, what is the maximum number t such that every $(q^2 + q + 1)$-vertex graph with $\mathrm{ex}(q^2 + q + 1, C_4) + 1$ contains an edge that is in at least t copies of C_4?*

We do not even know for sure if $t \geq 2$, though we proved that there are many copies of C_4. Let B_s be the collection of s copies of C_4 that share a common edge. If $t \geq 2$ holds, then perhaps it also makes sense to ask if $\mathrm{ex}(n, C_4) = \mathrm{ex}(n, B_2)$ holds (for infinite many integers n).

Many problems here only ask for $n = q^2 + q + 1$ and $q = 2^k$. Certainly it will be interesting to ask for other values of n when it is applicable.

References

[1] R. Baer, Null systems in projective space, *Bull. Amer. Math. Soc.* **51** (1945), 907–910.

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

[3] R. C. Baker, G. Harman and J. Pintz, The difference between consecutive primes, II, *Proceedings of the London Mathematical Society* **83(3)** (2001), 532–562.

[4] W. G. Brown, On graphs that do not contain a Thomsen graph, *Canad. Math. Bull.* **9** (1966), 281–289.

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

[6] F. Chung, Open problems of Paul Erdős in graph theory, *J. Graph Theory* **25** (1997), 3–36.

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

[8] 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)

[9] P. Erdős, On a theorem of Rademacher-Turán, *Illinois J. Math.* **6** (1962), 122–127.

[10] P. Erdős, On the number of complete subgraphs contained in certain graphs, *Magy. Tud. Acad. Mat. Kut. Int. Közl.* **7** (1962), 459–474.

[11] P. Erdős, Extremal problems on graphs and hypergraphs, in Hypergraph Seminar (Proc. First Working Sem., Ohio State Univ., Columbus, Ohio, 1972) *Lecture Notes in Math.* **411**, Springer, Berlin, 1974. pp. 75–84.

[12] P. Erdős, Problems and results in combinatorial analysis and graph theory, *Discrete Math.* **72** (1988), 81–92.

[13] P. Erdős, Some of my favourite unsolved problems, *A Tribute to Paul Erdős.* Cambridge University Press, New York (1990), 467–478.

[14] P. Erdős, A. Rényi and V. T. Sós, On a problem of graph theory, *Studia Sci. Math. Hungar.* **1** (1966), 51–57.

[15] P. Erdős and M. Simonovits, Supersaturated graphs and hypergraphs, *Combinatorica* **3** (1983), 181–192.

[16] 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.

[17] A. Ferber, G. McKinley and W. Samotij, Supersaturated sparse graphs and hypergraphs, *International Mathematics Research Notices*, to appear.

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

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

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

[21] 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`

[22] 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.

[23] 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.

[24] J. He, J. Ma and T. Yang, A short proof on stability of 4-cycles, manuscript, 2019 November, `http://staff.ustc.edu.cn/~jiema/short-stability-C4.pdf`

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

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

[27] N. Khadžiivanov and V. Nikiforov, Solution of a problem of P. Erdős about the maximum number of triangles with a common edge in a graph, *C. R. Acad. Bulgare Sci.* **32** (1979), 1315–1318 (in Russian).

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

[29] K. Metsch, On the maximum size of a maximal partial plane, *Rend. Mat. Appl. (7)* **12** (1992), 345–355.

[30] D. Mubayi, Counting substructures I: color critical graphs, *Adv. Math.* **225(5)** (2010), 2731–2740.

[31] Z. Nagy, Supersaturation of C_4: From Zarankiewicz towards ErdősCSimonovitsCSidorenko, *European J. Combin.* **75** (2019), 19–31.

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

[33] O. Pikhurko and Z. Yilma, Supersaturation problem for color-critical graphs, *J. Combin. Theory Ser. B* **123** (2017), 148–185.

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

[35] C. Reiher, The clique density theorem, *Annals of Math.* **184** (2016), 683–707.

[36] I. Reiman, Über ein Problem von K. Zarankiewicz, *Acta Math. Acad. Sci. Hungar.* **9** (1958), 269–278.

[37] M. Simonovits, Extremal graph problems, degenerate extremal problems and supersaturated graphs, *in Prog. Graph Theory,* (Academic Press, New York, 1984), 419–437.

[38] Y. Yuansheng and P. Rowlinson, On extremal graphs without four-cycles, *Util. Math.* **41** (1992) 204–210.