WILEY

On the number of linear hypergraphs of large girth

József Balogh | Lina Li

Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, Illinois

Correspondence
Lina Li, Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801.
Email: linali2@illinois.edu

Abstract

An r-uniform *linear cycle* of length ℓ, denoted by C_ℓ^r, is an r-graph with edges $e_1, ..., e_\ell$ such that for every $i \in [\ell - 1]$, $|e_i \cap e_{i+1}| = 1$, $|e_\ell \cap e_1| = 1$, and $e_i \cap e_j = \varnothing$ for all other pairs $\{i, j\}$, $i \neq j$. For every $r \geq 3$ and $\ell \geq 4$, we show that there exists a constant C depending on r and ℓ such that the number of linear r-graphs of girth ℓ is at most $2^{Cn^{1+1/\lfloor \ell/2 \rfloor}}$. Furthermore, we extend the result for $\ell = 4$, proving that there exists a constant C depending on r such that the number of linear r-graphs without C_4^r is at most $2^{Cn^{3/2}}$. The idea of the proof is to reduce the hypergraph enumeration problems to some graph enumeration problems, and then apply a variant of the graph container method, which may be of independent interest. We extend a breakthrough result of Kleitman and Winston on the number of C_4-free graphs, proving that the number of graphs containing at most $n^2/32 \log^6 n$ C_4's is at most $2^{11n^{3/2}}$, for sufficiently large n. We further show that for every $r \geq 3$ and $\ell \geq 2$, the number of graphs such that each of its edges is contained in only $O(1)$ cycles of length at most 2ℓ, is bounded by $2^{3(\ell+1)n^{1+1/\ell}}$ asymptotically.

KEYWORDS
girth, linear cycles, linear hypergraphs, the graph container method

1 | INTRODUCTION

For an integer $r \geq 2$, an r-uniform hypergraph (or r-graph) $H = (V, E)$ consists of a set V of vertices and a set E of edges, where each edge is an r-element subset of V. For a family of r-graphs $\mathcal{H}$, the *Turán number (function)* of $\mathcal{H}$, denoted by $\text{ex}_r(n, \mathcal{H})$, is the maximum number of edges among r-graphs on n vertices which contain no r-graph from $\mathcal{H}$ as a subgraph. Write $\text{Forb}_r(n, \mathcal{H})$ for the set of r-graphs with vertex set $[n]$ which contain no r-graph from $\mathcal{H}$ as a subgraph. When $\mathcal{H}$ consists of a single graph H, we simply write $\text{ex}_r(n, H)$ and $\text{Forb}_r(n, H)$ instead. Since every subgraph of an H-free graph is also H-free, we have a trivial bound

$$2^{\text{ex}_r(n,H)} \leq |\text{Forb}_r(n, H)| \leq \sum_{i \leq \text{ex}_r(n,H)} \binom{\binom{n}{r}}{i} \leq 2^{n^r \cdot \text{ex}_r(n,H)}. \tag{1}$$

The study on determination of $|\text{Forb}_r(n, H)|$ has a very rich history. Recently, the case when H is a linear cycle received more attention. For integers $r \geq 2$ and $\ell \geq 3$, an r-uniform *linear cycle* of length ℓ, denoted by C_ℓ^r, is an r-graph with edges $e_1, \ldots, e_\ell$ such that for every $i \in [\ell - 1]$, $|e_i \cap e_{i+1}| = 1$, $|e_\ell \cap e_1| = 1$, and $e_i \cap e_j = \varnothing$ for all other pairs $\{i, j\}$, $i \neq j$. Kostochka, Mubayi, and Verstraëte [17], and independently, Füredi and Jiang [10] proved that for every $r, \ell \geq 3$, $\text{ex}_r(n, C_\ell^r) = \Theta(n^{r-1})$. Then by (1), we trivially have

$$|\text{Forb}_r(n, C_\ell^r)| = 2^{\Omega(n^{r-1})} \text{ and } |\text{Forb}_r(n, C_\ell^r)| = 2^{O(n^{r-1} \log n)} \tag{2}$$

for every $r, \ell \geq 3$. Guided and motivated by this development on the extremal numbers of linear cycles, recently, Mubayi and Wang [21] showed that $|\text{Forb}_3(n, C_\ell^3)| = 2^{O(n^2)}$ for all even ℓ and improved the trivial upper bound in (2) for $r > 3$. Inspired by Mubayi and Wang's [21] method, Han and Kohayakawa [11] subsequently improved the general upper bound to $2^{O(n^{r-1} \log \log n)}$. Very recently, Balogh, Narayanan, and Skokan [3] provided a balanced supersaturation theorem for linear cycles and finally proved $|\text{Forb}_r(n, C_\ell^r)| = 2^{O(n^{r-1})}$, for every $r, \ell \geq 3$, using the hypergraph container method [2,24].

In this paper, we study the enumeration problem of linear hypergraphs containing no linear cycle of fixed length. An r-graph H is said to be *linear* if for every $e, e' \in E(H)$, $|e \cap e'| \leq 1$. For a family of linear r-graphs $\mathcal{H}$, the *linear Turán number* of $\mathcal{H}$, denoted by $\text{ex}_L(n, \mathcal{H})$, is the maximum number of edges among linear r-graphs on n vertices which contain no r-graph from $\mathcal{H}$ as a subgraph. Write $\text{Forb}_L(n, \mathcal{H})$ for the set of linear r-graphs with vertex set $[n]$ which contain no r-graph from $\mathcal{H}$ as a subgraph. Again, when $\mathcal{H}$ consists of a single graph H, we simply write $\text{ex}_L(n, H)$ and $\text{Forb}_L(n, H)$ instead. Similarly to (1), a trivial bound on the size of $\text{Forb}_L(n, H)$ is given as follows:

$$2^{\text{ex}_L(n,\mathcal{H})} \leq |\text{Forb}_L(n, \mathcal{H})| \leq \sum_{i \leq \text{ex}_L(n,\mathcal{H})} \binom{\binom{n}{r}}{i} \leq 2^{n^r \cdot \text{ex}_L(n,\mathcal{H})}. \tag{3}$$

It is known from the famous (6, 3)-problem that $n^{2-c\sqrt{\log n}} < \text{ex}_L(n, C_3^3) = o(n^2)$, where the lower bound is given by Behrend [4] and the upper bound is given by Ruzsa and Szemerédi [23]. In 1968, Erdős, Frankl, and Rödl [7] showed that for every $r \geq 3$, $\text{ex}_L(n, C_3^r) = o(n^2)$, and $\text{ex}_L(n, C_3^r) = \Omega(n^c)$ for every $c < 2$. Using the so-called 2-fold Sidon sets, Lazebnik and Verstraëte [19] constructed linear 3-graphs with girth 5 and $\Omega(n^{3/2})$ edges. On the other hand, it

is not hard to show that $\mathrm{ex}_L(n, C_4^3) = O(n^{3/2})$. Hence, $\mathrm{ex}_L(n, C_4^3) = \Theta(n^{3/2})$. Kostochka, Mubayi, and Verstraëte [16] proved $\mathrm{ex}_L(n, C_5^3) = \Theta(n^{3/2})$ and conjectured that

$$\mathrm{ex}_L(n, C_\ell^r) = \Theta\left(n^{1 + \frac{1}{\lfloor \ell/2 \rfloor}}\right)$$

for every $r \geq 3$ and $\ell \geq 4$. Later, Collier-Cartaino, Graber, and Jiang [6] proved that $\mathrm{ex}_L(n, C_\ell^r) = O\left(n^{1 + \frac{1}{\lfloor \ell/2 \rfloor}}\right)$ for $r \geq 3$ and $\ell \geq 4$. Although the lower bound on the linear Turán number of linear cycles is still far from what is conjectured, following the same logic with the usual Turán problem of cycles, it is natural to guess that

$$|\mathrm{Forb}_L(n, C_\ell^r)| = 2^{\Theta\left(n^{1 + \frac{1}{\lfloor \ell/2 \rfloor}}\right)} \tag{4}$$

for every $r \geq 3$ and $\ell \geq 4$. In this paper, we confirm the above conjecture for $\ell = 4$.

Theorem 1.1. *For every $r \geq 3$, there exists $C = C(r) > 0$ such that*

$$|\mathrm{Forb}_L(n, C_4^r)| \leq 2^{Cn^{3/2}}.$$

The upper bound for C_4^3 is sharp in order of magnitude given by $\mathrm{ex}_L(n, C_4^3) = \Theta(n^{3/2})$ and (3). In general, since the sharp bound of related linear Turán number remains open, we are not able to confirm the sharpness now.

For $\ell = 3$, the work of Erdős, Frankl, and Rödl [7] could be extended to show that $\mathrm{Forb}_L(n, C_3^r) = 2^{o(n^2)}$ for every $r \geq 3$. For $\ell > 4$, although we are not ready to prove (4), we provide a result on the girth version. Recall that the girth of a graph is the length of a shortest cycle contained in the graph. Kleitman and Wilson [13], and independently Kreuter [18], and Kohayakawa, Kreuter, and Steger [15] proved that there are $2^{O(n^{1+1/\ell})}$ graphs with no even cycles of length 2ℓ, which made a step towards proving a longstanding conjecture of Erdős, who asked for determining the number of $C_{2\ell}$-free graphs. Motivated by the above work, we introduce an analogous girth problem on linear hypergraphs. For a linear r-graph H, the *girth* of H is the smallest integer k such that H contains a C_k^r. We remark that for linear r-graphs, our girth definition is equivalent to a more classical girth definition, *Berge girth*, that is, the smallest number k such that the r-graph contains a Berge-C_k^r, as a linear Berge-C_k^r must contain a linear cycle of length i for some $3 \leq i \leq k$. For every $r \geq 3$ and $\ell \geq 4$, let $\mathrm{Forb}_L(n, r, \ell)$ denote the set of all linear r-graphs on $[n]$ with girth larger than ℓ. Our second main result is as follows.

Theorem 1.2. *For every $r \geq 3$ and $\ell \geq 4$, there exists a constant $C = C(r, \ell) > 0$ such that*

$$|\mathrm{Forb}_L(n, r, \ell)| \leq 2^{Cn^{1 + 1/\lfloor \ell/2 \rfloor}}.$$

Recently, Palmer et al [22] considered extremal problems for Berge-hypergraphs and proved our theorem for the case $\ell = 4$. Note that for every $\ell \geq 4$, we have $\mathrm{Forb}_L(n, r, \ell + 1) \subseteq \mathrm{Forb}_L(n, r, \ell)$. Therefore, it is sufficient to prove Theorem 1.2 for all even ℓ and we provide the following equivalent theorem instead.

Theorem 1.3. *For every $r \geq 3$ and $\ell \geq 2$, there exists a constant $C = C(r, \ell) > 0$ such that*

$$|\mathrm{Forb}_L(n, r, 2\ell)| \leq 2^{Cn^{1 + 1/\ell}}.$$

Once again, the above upper bounds are possibly sharp, but we are not able to confirm it now.

The proofs of Theorems 1.1 and 1.3 are based on two graph enumeration results related to even cycles. A classical result of Bondy and Simonovits [5] yields $ex_2(n, C_{2\ell}) = O(n^{1+1/\ell})$ for all $\ell \geq 2$. By a series of papers of Kleitman and Winston [14], Kleitman and Wilson [13], Kreuter [18], Kohayakawa, Kreuter, and Steger [15], and Morris and Saxton [20], we now know that the number of $C_{2\ell}$-free graphs is at most $2^{O(n^{1+1/\ell})}$. Inspired by these works, we prove that the number of graphs containing some but not many short cycles is still at most $2^{O(n^{1+1/\ell})}$, which may be of independent interest. We state our results as follows.

Theorem 1.4. *Let n be a sufficiently large integer and $a = 32 \log^6 n$. The number of n-vertex graphs with at most n^2/a 4-cycles is at most $2^{11n^{3/2}}$.*

Given a graph G on $[n]$, for every integer $k \geq 3$ and every edge $uv \in E(G)$, denote by $c_k(u, v; G)$, the number of k-cycles in G containing edge uv. When the underlying graph is clear, we simply write $c_k(u, v)$. For an integer $\ell \geq 3$ and a constant $L > 0$, write $\mathcal{G}_n(\ell, L)$ for the family of graphs G on $[n]$ such that for every $3 \leq k \leq \ell$ and $uv \in E(G)$, $c_k(u, v; G) \leq L$.

Theorem 1.5. *For an integer $\ell \geq 3$ and a constant $L > 0$, let n be a sufficiently large integer and then we have*

$$|\mathcal{G}_n(2\ell, L)| \leq 2^{3(\ell+1)n^{1+1/\ell}}.$$

Like many of these advances, our approach to proving Theorems 1.4 and 1.5 relies on the graph container method developed in [14], in which one assigns a *certificate* for each target graph. The certificate should be able to uniquely determine the target graph, and then we can estimate the number of certificates instead of graphs. However, the previous applications of the graph container method address the problems for graphs forbidding short cycles, while we concern with the graphs with sparse short cycles. Therefore, the means by which we apply this technique is quite nonstandard, and requires some new ideas.

It is not hard to extend Theorem 1.4 to $a = \Theta(\log^5 n)$ by proving a similar statement for $\mathcal{G}_n(4, \sqrt{n}/\log^4 n)$ as in Theorem 1.5. We choose to display the proof of Theorem 1.4 since it contains some ideas which may bring more insights of this method to readers. Let $p = \omega/(\sqrt{n} \log n)$. Note that the number of graphs on $[n]$ with $p\binom{n}{2}$ edges is about $2^{\omega n^{3/2}}$ and they typically contain $\Theta(n^4 p^4) = \Theta(\omega^4 n^2/\log^4 n)$ 4-cycles. Therefore, $a = \Theta(\log^4 n)$ would be the best possible in Theorem 1.4 and we believe that it should be the truth. Given by the connection between Sidon sets and graphs without 4-cycles, this problem is closely related with an enumeration problem on generalized Sidon set which was recently studied in the authors' another paper [1].

Throughout this paper, we let $[n]$ denote the set $\{1, 2,...,n\}$. For a graph G and a set $S \subseteq V(G)$, the *induced subgraph $G[S]$* is the subgraph of G whose vertex set is S and whose edge set consists of all of the edges with both endpoints in S. Let $\delta(G)$ denote the minimum degree of graph G and $\Delta(G)$ denote the maximum degree of G. For a multigraph G and a vertex $v \in V(G)$, the *neighborhood $N_G(v)$* of v is the set of all vertices adjacent to v in G and the *degree $d_G(v)$* of v is the number of edges incident to v. For a set $S \subseteq V(G)$, the *neighborhood of v restricted to S* is $N_G(v, S) = N_G(v) \cap S$; the *degree of v restricted to S*, denoted by $d_G(v, S)$, is the number of edges incident to v with another endpoint in S. When the underlying graph is clear, we simply write $N(v, S)$ and $d(v, S)$ instead. All logarithms have base 2.

2 | GRAPHS CONTAINING A FEW 4-CYCLES

2.1 | Preliminary results

Definition 2.1 (Min-degree ordering, Min-degree sequence). For a graph G on $[n]$, a *min-degree ordering* is an ordering $v_n < v_{n-1} < \cdots < v_1$, such that v_i is a vertex of minimum degree in the graph $G_i = G[v_i,...,v_1]$, for every $i \in [n]$ (if there are more than one vertices of the minimum degree, choose the one with the largest label). Let $d_i = d_{G_i}(v_i)$, then $d_n, d_{n-1},...,d_1$ is called the *min-degree sequence*.

Lemma 2.2. *Let G be an n-vertex graph with average degree d. If $d \geq 2\sqrt{n}$, then G contains at least $d^4/36$ copies of 4-cycles.*

Proof. Let $v_1, v_2, ...,v_n$ be the vertices in G and $b_i = d_G(v_i)$ for every $i \in [n]$. Let S be the set of paths of length 2 (or 3-paths) in G. We will count 3-paths in two ways.

First, for a vertex v_i, the number of 3-paths containing v_i as the middle point is exactly $\binom{b_i}{2}$. Therefore, we have

$$|S| = \sum_{i=1}^{n} \binom{b_i}{2} \geq n\left(\frac{(\sum_{i=1}^{n} b_i)/n}{2} \right) = n\binom{d}{2} \geq \frac{1}{3}d^2 n.$$

On the other hand, for $1 \leq i < j \leq n$, let c_{ij} be the number of common neighbors of v_i and v_j. Then $|S| = \sum_{1 \leq i < j \leq n} c_{ij}$. Therefore, the number of 4-cycles in G is equal to

$$\frac{1}{2} \sum_{1 \leq i < j \leq n} \binom{c_{ij}}{2} \geq \frac{1}{2}\binom{n}{2}\left(\frac{(\sum_{i<j} c_{ij})/\binom{n}{2}}{2} \right) = \frac{1}{2}\binom{n}{2}\left(\frac{|S|/\binom{n}{2}}{2} \right) \geq \frac{|S|^2}{4n^2} \geq \frac{d^4}{36}. \qquad \square$$

From Lemma 2.2, we immediately obtain the following corollary.

Corollary 2.3. *Let G be a n-vertex graph which contains at most $4n^2/9$ 4-cycles, and $d_n, ...,d_1$ be the min-degree sequence of G. Then for every $i \in [n]$,*

$$d_i \leq 2\sqrt{n}.$$

Proof. Suppose that there exists $k \in [n]$, such that $d_k > 2\sqrt{n}$. Then by Lemma 2.2, the number of 4-cycles in G_k is at least $d_k^4/36 > \frac{4}{9}n^2$, which contradicts our assumption. $\square$

We also provide an estimation for the following binomial coefficients, which will be used repeatedly later.

Lemma 2.4. *For integers n, k, ℓ and a constant c satisfying $cn/k^\ell \geq k$,*

$$\binom{cn/k^\ell}{k} \leq 2^{\frac{\ell+1}{2^{1/\ln 2}\ln 2}(cen)^{\frac{1}{\ell+1}}},$$

where $2^{1/\ln 2}\ln 2 \approx 1.88$.

Proof. Let $f(x) = (\log cen - (\ell + 1)\log x)x$ on $(0, +\infty)$. Since $f(x)$ is a concave function, it is maximized at the point x^*, where

$$f'(x^*) = \log cen - \frac{\ell + 1}{\ln 2} - (\ell + 1)\log x^* = 0,$$

that is,

$$\log x^* = \frac{\log cen}{\ell + 1} - \frac{1}{\ln 2}.$$

Therefore, we have

$$f(k) \leq f(x^*) = \left(\log cen - (\ell + 1)\left(\frac{\log cen}{\ell + 1} - \frac{1}{\ln 2}\right)\right)2^{\frac{\log cen}{\ell+1} - \frac{1}{\ln 2}} = \frac{\ell + 1}{2^{1/\ln 2}\ln 2}(cen)^{\frac{1}{\ell+1}}.$$

Since $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$ for every $1 \leq k \leq n$, we obtain that

$$\binom{cn/k^\ell}{k} \leq \left(\frac{cen}{k^{\ell+1}}\right)^k = 2^{f(k)} \leq 2^{2^{\frac{\ell+1}{2^{1/\ln 2}\ln 2}(cen)^{\frac{1}{\ell+1}}}}. \qquad \square$$

2.2 | Certificate lemma

This section is devoted to prove our main lemma, which is a key step to build the certificates for graphs with sparse 4-cycles. This lemma can be viewed as a generalization of the Kleitman-Winston algorithm [14], which builds certificates for graphs without 4-cycles. Before we proceed, we first need a counting lemma, which will be used later in the proof.

For a graph F, denote by F^2 the multigraph defined on $V(F)$ such that for every distinct $u, v \in V(F^2)$, the multiplicity of uv in F^2 is the number of (u, v)-paths of length 2 in F.

Lemma 2.5. *For integers $n > m \geq d \geq 8$, let F be an m-vertex graph with $\delta(F) \geq d - 1$ and $H = F^2$. Then for every $J \subseteq V(H)$ of size at least $4n/d$, we have*

$$e(H[J]) \geq \frac{d^2|J|^2}{4n}.$$

Proof. Write $V(F) = \{v_1,...,v_m\}$. For every $j \in [m]$, let $b_j = d_F(v_j, J)$. Then we have $\sum_{j=1}^m b_j = \sum_{v \in J} d_F(v) \geq |J|(d - 1) \geq (4(d - 1)/d)n > 3n > 3m$. Therefore, we obtain that

$$e(H[J]) = \sum_{j=1}^m \binom{b_j}{2} \geq m\binom{\frac{\sum b_j}{m}}{2} \geq m\binom{\frac{|J|(d-1)}{m}}{2} \geq \frac{|J|^2(d-1)^2}{3m} \geq \frac{d^2|J|^2}{4n}. \qquad \square$$

Lemma 2.6 (Certificate lemma). *For a sufficiently large integer n, define $b = 16\log^4 n$ and $g = 32\log^5 n$. Let m and d be the integers satisfying $m \leq n - 1$ and $(\sqrt{n}/\log n) \leq d \leq 2\sqrt{n}$. Suppose that F is an m-vertex graph with $\delta(F) \geq d - 1$ and $H = F^2$. Additionally, assume that for every $u, v \in V(F)$, $|N_F(u) \cap N_F(v)| \leq \sqrt{n}/b$. Then for every set $I \subseteq V(F)$ of size d which satisfies $e(H[I]) \leq n/g$, there exist a set T and a set $C(T)$ depending only on T, not on I, such that*

(i) $T \subseteq I \subseteq C(T)$,
(ii) $|T| \leq 2\sqrt{n}/\log n$,
(iii) $|C(T)| \leq 5n/d$.

Proof. Let I be a subset of $V(F)$ of size d which satisfies $e(H[I]) \leq n/g$. Following the ideas of Kleitman and Winston [14], we describe a deterministic algorithm that associates to the set I a pair of sets T and $C(T)$, which shall be treated as the "fingerprint" and the "container," respectively.

Let $\quad I_h = \{v \in I: d_H(v, I) > \sqrt{n}/b\} \quad$ and $\quad I_l = \{v \in I: d_H(v, I) \leq \sqrt{n}/b\}$. Since $e(H[I]) \leq n/g$, the size of I_h is at most

$$\frac{2e(H[I])}{\sqrt{n}/b} \leq \frac{2\sqrt{n} \cdot b}{g} = \frac{\sqrt{n}}{\log n},$$

which is sufficiently small. Therefore, we only need to concern the vertices in I_l.

The core algorithm. We start the algorithm with sets $A_0 = V(H) - I_h$, $T_0 = \emptyset$ and the function $t_0(v) = 0$ for every $v \in V(H) - I_h$. As the algorithm proceeds, one should view A_i as the set of "candidate" vertices, T_i as the set of "representative" vertices, and $t_i(v)$ as a "state" function which is used to control the process. In the ith iteration step, we pick a vertex $u_i \in A_i$ of maximum degree in $H[A_i]$. In case there are multiple choices, we give preference to vertices that come earlier in some arbitrary predefined ordering of $V(H)$ as we always do, even if it is not pointed out at each time. If $u_i \in I_l$, we define

$$t_{i+1}(v) = \begin{cases} t_i(v) + d_H(v, u_i) & \text{if } v \in A_i, \\ t_i(v) & \text{if } v \notin A_i, \end{cases}$$

and $Q = \{v|t_{i+1}(v) > \sqrt{n}/b\}$, and let $T_{i+1} = T_i + u_i$, $A_{i+1} = A_i - u_i - Q$. Otherwise, let $T_{i+1} = T_i$, $A_{i+1} = A_i - u_i$, and $t_{i+1}(v) = t_i(v)$, for every $v \in V(H) - I_h$. The algorithm terminates at step K once we get a set A_K of size at most $4n/d$. We also assume that $u_{K-1} \in T_K$ as otherwise we can continue the algorithm until it is satisfied.

The algorithm outputs a vertex sequence $\{u_1, u_2,...,u_{K-1}\}$, a set of 'representative' vertices T_K and a strictly decreasing set sequence $\{A_0, A_1, A_2, A_3,...,A_K\}$. Let

$$T = T_K \cup I_h, \quad \text{and} \quad C(T) = A_K \cup T.$$

From the algorithm, we have $T_K \subseteq I_l$ and therefore $T \subseteq I$. Furthermore, if a vertex v satisfies $t_i(v) > \sqrt{n}/b$ for some i, then we have $d_H(v, I) \geq t_i(v) > \sqrt{n}/b$, which implies $v \notin I_l$. Therefore, we maintain $I_l \subseteq A_i \cup T_i$ for every $i \leq K$ and in particular we have $I \subseteq A_K \cup T_K \cup I_h = A_K \cup T = C(T)$. Hence, Condition (i) is satisfied. Another crucial fact is that $C(T)$ depends only on T, not on I. The reason is that for a given underlying graph, its max degree sequence is fixed once we break the tie by some predefined ordering on vertices. Therefore, for two sets I_1, I_2 with the same "fingerprint" T, the algorithm outputs the same vertex sequence $\{u_1, u_2,...,u_{K-1}\}$, which uniquely determines the set $C(T)$ by the mechanics of the algorithm.

To verify Conditions (ii) and (iii), it is sufficient to show that $|T_K| \leq \sqrt{n}/\log n$. Once we prove it, we immediately obtain

$$|T| = |T_K| + |I_h| \leq \frac{\sqrt{n}}{\log n} + \frac{\sqrt{n}}{\log n} = \frac{2\sqrt{n}}{\log n},$$

and

$$|C(T)| = |A_K| + |T| \leq \frac{4n}{d} + \frac{2\sqrt{n}}{\log n} \leq \frac{5n}{d},$$

completing the proof.

Denote q the integer such that $n/2^q \leq |A_K| < n/2^{q-1}$. By the choice of A_K, we have $q < \log n$. For every integer $1 \leq l \leq q$, define A^l to be the first A-set satisfying

$$\frac{n}{2^l} \leq |A^l| < \frac{n}{2^{l-1}}$$

if it exists, and let T^l be the corresponding T-set and $t^l(v)$ be the corresponding t-function of A^l. Note that A^l may not exist for every l, but A^q always exists and it could be that $A^q = A_K$. Suppose that

$$A^{l_1} \supset A^{l_2} \supset \cdots \supset A^{l_p}$$

are all the well-defined A^l, where $p \leq q$. By the above definition, we have $A^{l_1} = A_0$, $T^{l_1} = T_0$, and $l_p = q$. Define $A^{l_{p+1}} = A_K$, $T^{l_{p+1}} = T_K$. Now, we have

$$T_K = \bigcup_{j=2}^{p+1} (T^{l_j} - T^{l_{j-1}}). \tag{5}$$

To achieve our goal, we are going to estimate the size of $T^{l_j} - T^{l_{j-1}}$ for every $2 \leq j \leq p + 1$.

From the algorithm, we have $t^{l_j}(v) \leq \sqrt{n}/b$, for every $v \in A^{l_j} \cup T^{l_j}$. Moreover, for $v \in A^{l_{j-1}} - A^{l_j} - T^{l_j}$, suppose that v is removed in step i, then we have

$$t^{l_j}(v) \leq t_{i-1}(v) + d_H(v, u_i) \leq \frac{\sqrt{n}}{b} + |N_F(u_i) \cap N_F(v)| \leq \frac{2\sqrt{n}}{b},$$

where u_i is the selected vertex in step i. Therefore, we obtain

$$\sum_{v \in A^{l_{j-1}}} t^{l_j}(v) \leq \frac{2\sqrt{n}}{b} |A^{l_{j-1}}| \leq \frac{2n^{3/2}}{2^{l_{j-1}-1}b}. \tag{6}$$

Let $2 \leq j \leq p$. For every $u_i \in T^{l_j} - T^{l_{j-1}}$, u_i is chosen of maximum degree in $H[A_i]$, where A_i is a set between $A^{l_{j-1}}$ and A^{l_j}. By the choice of A^{l_j}, we have $|A_i| \geq n/2^{l_{j-1}}$. By Lemma 2.5, we have

$$d_H(u_i, A_i) \geq \frac{d^2|A_i|}{4n} \geq \frac{d^2}{2^{l_{j-1}+2}}.$$

Note that $d_H(u_i, A_i)$ only contributes to $t^{l_j}(v)$ for $v \in A_i \subseteq A^{l_{j-1}}$. Then we obtain

$$|T^{l_j} - T^{l_{j-1}}| \frac{d^2}{2^{l_{j-1}+2}} \leq \sum_{u_i \in T^{l_j} - T^{l_{j-1}}} d_H(u_i, A_i) \leq \sum_{v \in A^{l_{j-1}}} t^{l_j}(v). \tag{7}$$

Combining (6) and (7), we have

$$|T^{l_j} - T^{l_{j-1}}| \frac{d^2}{2^{l_{j-1}+2}} \le \frac{2n^{3/2}}{2^{l_{j-1}-1}b},$$

which implies

$$|T^{l_j} - T^{l_{j-1}}| \le \frac{16n^{3/2}}{bd^2} \le \frac{16\sqrt{n}\log^2 n}{b} = \frac{\sqrt{n}}{\log^2 n}$$

for $2 \le j \le p$. For $j = p + 1$, since we have

$$\frac{n}{2^q} \le |A^{l_{p+1}}| \le |A^{l_p}| \le \frac{n}{2^{q-1}}$$

by a similar argument, we obtain that

$$|T^{l_{p+1}} - T^{l_p}| \frac{d^2|A^{l_{p+1}}|}{4n} \le \sum_{u_i \in T^{l_{p+1}} - T^{l_p}} d(u_i, A_i) \le \sum_{v \in A^{l_p}} t^{l_{p+1}}(v) \le \frac{2\sqrt{n}}{b}|A^{l_p}|,$$

which gives

$$|T^{l_{p+1}} - T^{l_p}| \le \frac{16n^{3/2}}{bd^2} \le \frac{16\sqrt{n}\log^2 n}{b} = \frac{\sqrt{n}}{\log^2 n}.$$

Finally, by (5), we get

$$|T_K| = \bigcup_{j=2}^{p+1} |T^{l_j} - T^{l_{j-1}}| \le p \cdot \frac{\sqrt{n}}{\log^2 n} \le q \cdot \frac{\sqrt{n}}{\log^2 n} \le \frac{\sqrt{n}}{\log n}. \qquad \square$$

2.3 | Proof of Theorem 1.4

In this section, we give an upper bound on the number of graphs containing only "few" 4-cycles. Before we proceed to prove Theorem 1.4, we need to do a cleaning process for the target graphs in order to apply Lemma 2.6.

Let $a = 32\log^6 n$, $g = 32\log^5 n$, and $b = 16\log^4 n$. Given a graph G on $[n]$, for every $1 \le i < j \le n$, define $N_G(i, j)$ to be the set of common neighbors of i and j in G. Let

$$m_G(i, j) = \begin{cases} |N_G(i, j)| & \text{when } |N_G(i, j)| > \frac{\sqrt{n}}{b}, \\ 0 & \text{when } |N_G(i, j)| \le \frac{\sqrt{n}}{b}. \end{cases}$$

We delete all edges from i to $N_G(i, j)$, for all $1 \le i < j \le n$ with $m_G(i, j) \ne 0$. Then the resulting subgraph, denoted by $\widehat{G}$, satisfies $|N_{\widehat{G}}(i, j)| \le \sqrt{n}/b$ for every $1 \le i < j \le n$. Let $\mathcal{G}_n$ be the family of graphs on $[n]$ with at most n^2/a 4-cycles and $\widehat{\mathcal{G}}_n = \{\widehat{G} : G \in \mathcal{G}_n\}$.

Lemma 2.7. *Let n be a sufficiently large integer. Then for every $G \in \mathcal{G}_n$, we have*

$$|E(G) - E(\widehat{G})| \leq \frac{4n^{3/2}}{\log^2 n}.$$

Proof. By counting 4-cycles in G, we obtain that

$$\frac{1}{2} \sum_{i<j} \binom{m_G(i,j)}{2} \leq \frac{n^2}{a},$$

which gives

$$\sum_{i<j} m_G(i,j)^2 \leq 8\frac{n^2}{a}. \tag{8}$$

Let $B = \{(i,j): 1 \leq i < j \leq n \text{ and } m_G(i,j) \neq 0\}$. By the definition of $m_G(i,j)$ and (8), we have

$$|B| \leq 8\frac{n^2}{a} / \left(\frac{\sqrt{n}}{b}\right)^2 = 8b^2n/a.$$

Therefore, by the convexity, we get

$$\sum_{(i,j)\in B} m_G(i,j)^2 \geq \frac{(\sum_{(i,j)\in B} m_G(i,j))^2}{|B|} = \frac{(\sum_{i<j} m_G(i,j))^2}{|B|} \geq \frac{(\sum_{i<j} m_G(i,j))^2}{8b^2n/a}. \tag{9}$$

Combining (8) and (9), we obtain

$$\sum_{i<j} m_G(i,j) \leq \frac{8n^{3/2}b}{a} = \frac{4n^{3/2}}{\log^2 n}.$$

Finally, by the definition of $\widehat{G}$, we have $|E(G) - E(\widehat{G})| = \sum_{i<j} m_G(i,j) \leq (4n^{3/2}/\log^2 n)$. $\square$

Lemma 2.8. *Let n be a sufficiently large integer. Then*

$$|\mathcal{G}_n| \leq |\widehat{\mathcal{G}}_n| \cdot 2^{\frac{4n^{3/2}}{\log n}}.$$

Proof. For every $F \in \widehat{\mathcal{G}}_n$, let $\mathcal{S}_F = \{G \in \mathcal{G}_n | \widehat{G} = F\}$. By Lemma 2.7, for every $G \in \mathcal{S}_F$, we have $|E(G) - E(F)| \leq (4n^{3/2}/\log^2 n)$. Therefore, the size of $\mathcal{S}_F$ is bounded by

$$|\mathcal{S}_F| \leq \binom{\binom{n}{2}}{0} + \binom{\binom{n}{2}}{1} + \cdots + \binom{\binom{n}{2}}{\lfloor\frac{4n^{3/2}}{\log^2 n}\rfloor} \leq 2\binom{\binom{n}{2}}{\lfloor\frac{4n^{3/2}}{\log^2 n}\rfloor} \leq 2^{\frac{4n^{3/2}}{\log n}}.$$

Finally, we obtain that

$$|\mathcal{G}_n| \leq \sum_{F \in \widehat{\mathcal{G}}_n} |S_F| \leq |\widehat{\mathcal{G}}_n| \cdot 2^{\frac{4n^{3/2}}{\log n}}. \qquad \square$$

Theorem 2.9. *Let n be a sufficiently large integer. Then $|\widehat{\mathcal{G}}_n| \leq 2^{10n^{3/2}}$.*

Proof. We construct the certificate of a graph G in the following way. Let $Y_G := v_n < v_{n-1} < \cdots < v_1$ be the min-degree ordering of G and $D_G := \{d_n, d_{n-1},...,d_1\}$ be the min-degree sequence of G. Let $G_i = G[v_i,...,v_1]$, for every $i \in [n]$. Define the set sequence $S_G := \{S_n, S_{n-1},...,S_2\}$, where $S_i = N_G(v_i, G_{i-1})$. Then $S_i \subseteq \{v_{i-1},...,v_1\}$, and $|S_i| = d_i$. By the construction, $[Y_G, D_G, S_G]$ uniquely determines the graph G and so we build a certificate $[Y_G, D_G, S_G]$ for G. Therefore, instead of counting graphs, it is equivalent to estimate the number of their certificates.

For a graph $G \in \widehat{\mathcal{G}}_n$, its certificate has some important properties which would help us to achieve the desired bound. First, by Corollary 2.3, its min-degree ordering $D_G = \{d_n, d_{n-1},..., d_1\}$ satisfying $d_i \leq 2\sqrt{n}$. Let f_i be the number of 4-cycles in G_i containing vertex v_i. Since each 4-cycle contributes exactly to one of f_i's, we have $\sum_{i=1}^{n} f_i \leq n^2/a$. We call v_i a *heavy* vertex if $f_i > n/g$; otherwise, v_i is a *light* vertex. Another crucial fact about graphs in $\widehat{\mathcal{G}}_n$ is that the number of heavy vertices is at most

$$\frac{\sum_{v_i \in V_h} f_i}{n/g} \leq \frac{n^2/a}{n/g} = \frac{n}{\log n}. \tag{10}$$

Now we start to estimate the number of certificates which would generate graphs in $\widehat{\mathcal{G}}_n$. By the above discussion, we first observe that the number of ways to choose the min-degree orderings and the min-degree sequences is at most

$$n!(2\sqrt{n})^n. \tag{11}$$

Then we fix a min-degree ordering $Y^* = v_n < v_{n-1} < \cdots < v_1$, and a min-degree sequence $D^* = \{d_n,...,d_1\}$. Next, we fix the positions of heavy vertices and by (10) the number of ways is at most

$$\sum_{i \leq \frac{n}{\log n}} \binom{n}{i}. \tag{12}$$

A major part of the proof is to count set sequences $S = \{S_n, S_{n-1},...,S_2\}$, where $S_i \subseteq \{v_{i-1},...,v_1\}$ and $|S_i| = d_i$, such that the graph reconstructed by $[Y^*, D^*, S]$, denoted by G_S, are in $\widehat{\mathcal{G}}$. For every $2 \leq i \leq n$, let M_i be the number of choices for S_i with fixed sets $S_{i-1}, ..., S_2$. Define

$$\mathcal{I}_1 = \{i : v_i \text{ is a heavy vertex}\}, \quad \mathcal{I}_2 = \left\{i : d_i < \frac{\sqrt{n}}{\log n}\right\},$$

and

$$\mathcal{I}_3 = \left\{i : v_i \text{ is a light vertex and } d_i \geq \frac{\sqrt{n}}{\log n}\right\}.$$

For every $i \in \mathcal{I}_1$, since $|S_i| = d_i \le 2\sqrt{n}$, we have a trivial upper bound

$$M_i \le \binom{i-1}{d_i} \le \binom{n}{2\sqrt{n}} \le n^{2\sqrt{n}} = 2^{2\sqrt{n}\log n}. \tag{13}$$

Similarly, for every $i \in \mathcal{I}_2$, we have

$$M_i \le \binom{i-1}{d_i} \le \binom{n}{\sqrt{n}/\log n} \le n^{\sqrt{n}/\log n} = 2^{\sqrt{n}}. \tag{14}$$

It remains to estimate M_i for $i \in \mathcal{I}_3$. With fixed sets $S_{i-1}, ..., S_2$, the graph $G_{i-1} = G_S[v_{i-1},...,v_1]$ is uniquely determined. Since $G_{i-1} \subseteq G_S$ and $G_S \in \widehat{\mathcal{G}}$, for every $u, v \in V(G_{i-1})$, we have $|N_{G_{i-1}}(u) \cap N_{G_{i-1}}(v)| \le \sqrt{n}/b$. Applying Lemma 2.6 on G_{i-1}, we obtain that every eligible S_i contains a subset T of size at most $2\sqrt{n}/\log n$, which determines a set $C(T) \supseteq S_i$ of size at most $5n/d_i$. Since the number of choices for T is at most

$$\sum_{0 \le j \le 2\sqrt{n}/\log n} \binom{i-1}{j} \le 2\binom{i-1}{2\sqrt{n}/\log n} \le 2\binom{n}{2\sqrt{n}/\log n} \le 2^{2\sqrt{n}},$$

we then have

$$M_i \le \sum_T \binom{C(T)}{d_i} \le \sum_T \binom{5n/d_i}{d_i} \le \sum_T 2^{\frac{2}{2^{1/\ln 2}\ln 2}\sqrt{5en}} \le \sum_T 2^{4\sqrt{n}} \le 2^{6\sqrt{n}} \tag{15}$$

for every $i \in \mathcal{I}_3$, where the third inequality is given by Lemma 2.4.

Combining (13), (14), and (15), we obtain that the number of choices for S is

$$\prod_{i=2}^{n} M_i \le \prod_{i \in \mathcal{I}_1} M_i \prod_{i \in \mathcal{I}_2} M_i \prod_{i \in \mathcal{I}_3} M_i \le (2^{2\sqrt{n}\log n})^{\frac{n}{\log n}}(2^{\sqrt{n}})^n(2^{6\sqrt{n}})^n \le 2^{9n^{3/2}}.$$

Finally, together with (11) and (12), the total number of certificates is at most

$$n!(2\sqrt{n})^n \sum_{i \le \frac{n}{\log n}} \binom{n}{i} \prod_{i=2}^{n} M_i \le n!(2\sqrt{n})^n 2^n 2^{9n^{3/2}} \le 2^{10n^{3/2}},$$

which leads to $|\widehat{\mathcal{G}}_n| \le 2^{10n^{3/2}}$.

Proof of Theorem 1.4. Lemma 2.8 and Theorem 2.9 imply Theorem 1.4. □

3 | THE NUMBER OF GRAPHS WITH SPARSE SHORT CYCLES

In the previous section, we estimated the number of graphs containing a few 4-cycles. Unfortunately, we are not ready to provide a similar result for longer cycles due to the failure of getting an appropriate counting lemma, like Lemma 2.5. However, this method still works

when the target graph has a sparse structure on short cycles. More specifically, for $\ell \geq 4$, we are going to consider the family of graphs such that each of its edges is contained in only $O(1)$ cycles of length at most 2ℓ. Following the idea from [15], we construct a proper auxiliary graph and provide a suitable counting lemma on it.

3.1 | Expansion properties of graphs with sparse short cycles

Given a graph G, a vertex $v \in V(G)$, and an integer $k \geq 1$, let $\Gamma_k(v)$ be the set of vertices of G at distance exactly k from v. Recall that for an edge $uv \in E(G)$, $c_k(u, v; G)$ is the number of k-cycles in G containing edge uv.

Lemma 3.1. *For integers $\ell \leq m$ and a constant $L > 0$, let F be an m-vertex graph such that for every $uv \in E(F)$ and $3 \leq i \leq 2\ell$, $c_i(u, v) \leq L$. Then for every $1 \leq k \leq \ell - 1$ and $v \in V(F)$, we have*

$$d(u, \Gamma_k(v)) \leq Lk$$

for all $u \in \Gamma_k(v)$.

Proof. Suppose there exists a vertex $u \in \Gamma_k(v)$ such that $d(u, \Gamma_k(v)) \geq Lk + 1$. Since $u \in \Gamma_k(v)$, there exists a (u, v)-path P_u of length k. Let u' be the neighbor of u in P_u. Similarly, for every vertex $w \in N(u, \Gamma_k(v))$, there is a (w, v)-path P_w of length k. Note that every $P_u + P_w + \{uw\}$ forms a closed walk of length $2k + 1$, which contains an odd cycle of length at most $2k + 1$ containing edges uu' and uw. Since $d(u, \Gamma_k(v)) \geq Lk + 1$, we have at least $Lk + 1$ distinct odd cycles of length at most $2k + 1$ containing uu'. However, since $c_h(u, u') \leq L$ for every odd $h \leq 2k + 1$, there are at most Lk odd cycles of length at most $2k + 1$ containing uu', which is a contradiction. $\square$

Lemma 3.2. *For integers $\ell \leq m$ and a constant $L > 0$, let F be an m-vertex graph such that for every $uv \in E(F)$ and $3 \leq i \leq 2\ell$, $c_i(u, v) \leq L$. Then for every $2 \leq k \leq \ell$ and $v \in V(F)$, we have*

$$d(u, \Gamma_{k-1}(v)) \leq L(k - 1) + 1$$

for all $u \in \Gamma_k(v)$.

Proof. Suppose there exists a vertex $u \in \Gamma_k(v)$ such that $d(u, \Gamma_{k-1}(v)) \geq L(k - 1) + 2$. Let u' be a vertex in $N(u, \Gamma_{k-1}(v))$. Since $u' \in \Gamma_{k-1}(v)$, there exists a (u', v)-path $P_{u'}$ of length $k - 1$. Similarly, for every vertex $w \in N(u, \Gamma_{k-1}(v)) \setminus \{u'\}$, there is a (w, v)-path P_w of length $k - 1$. Note that every $P_{u'} + P_w + \{uu'\} + \{uw\}$ forms a closed walk of length $2k$, which contains an even cycle of length at most $2k$ containing edges uu' and uw. Since $|N(u, \Gamma_{k-1}(v)) \setminus \{u'\}| \geq L(k - 1) + 1$, we have at least $L(k - 1) + 1$ distinct even cycles of length at most $2k$ containing uu'. However, since $c_h(u, u') \leq L$ for every even $4 \leq h \leq 2k$, there are at most $L(k - 1)$ even cycles of length at most $2k$ containing uu', which is a contradiction. $\square$

Now, we give a lemma on the expansion of graphs with sparse short cycles. This lemma can be viewed as a generalization of Lemma 11 in [15].

Lemma 3.3. *For integers ℓ, $d \leq m$ and a constant $L \ll d$, let F be an m-vertex graph with minimum degree at least $d - 1$, such that for every $uv \in E(F)$ and $3 \leq i \leq 2\ell$, $c_i(u, v) \leq L$. Suppose v is a vertex in F with degree $d(v)$. Then for every $1 \leq k \leq \ell$, we have*

$$|\Gamma_k(v)| \geq \frac{d(v)d^{k-1}}{g_k(L)}$$

for some constants $g_k(L)$ which only depend on k and L.

Proof. The case $k = 1$ is trivially true with $g_1(L) = 1$. Suppose that the lemma is true for $k < \ell$, that is, $|\Gamma_k(L)| \geq d(v)d^{k-1}/g_k(L)$ for some constant $g_k(L)$.

For every vertex $u \in \Gamma_k(v)$, neighbors of u only appear in $\Gamma_{k-1}(v)$, $\Gamma_k(v)$ and $\Gamma_{k+1}(v)$. By Lemmas 3.1 and 3.2, we have

$$d(u, \Gamma_{k+1}(v)) \geq (d - 1) - d(u, \Gamma_{k-1}(v)) - d(u, \Gamma_k(v)) \geq d - 2(Lk + 1) + L \geq \frac{d}{2} \tag{16}$$

for all $u \in \Gamma_k(v)$ and this gives

$$e(\Gamma_k(v), \Gamma_{k+1}(v)) \geq \frac{d|\Gamma_k(v)|}{2}.$$

Again by Lemma 3.2, we know that for every $u \in \Gamma_{k+1}(v)$, $d(u, \Gamma_k(v)) \leq Lk + 1$. Therefore, we have

$$|\Gamma_{k+1}(v)| \geq \frac{e(\Gamma_k(v), \Gamma_{k+1}(v))}{Lk + 1} \geq \frac{d|\Gamma_k(v)|}{2(Lk + 1)} \geq \frac{d(v)d^k}{2(Lk + 1)g_k(L)} = \frac{d(v)d^k}{g_{k+1}(L)}$$

for $g_{k+1}(L) = 2(Lk + 1)g_k(L)$ and the lemma follows by induction. □

Lemma 3.3 gives an upper bound on the maximum degree of the graph with sparse short cycles.

Corollary 3.4. *For integers ℓ, $d \leq m$ and a constant $L \ll d$, let F be an m-vertex graph with minimum degree $d - 1$, such that for every $uv \in E(F)$ and $3 \leq i \leq 2\ell$, $c_i(u, v) \leq L$. Then*

$$\Delta(F) \leq \frac{m}{d^{\ell-1}} \cdot g_\ell(L),$$

where $g_\ell(L)$ is the constant defined in Lemma 3.3.

Proof. By Lemma 3.3, for every $v \in V(F)$, we have

$$|\Gamma_\ell(v)| \geq \frac{d(v)d^{\ell-1}}{g_\ell(L)},$$

which gives

$$d(v) \leq \frac{|\Gamma_\ell(v)|}{d^{\ell-1}}g_\ell(L) \leq \frac{m}{d^{\ell-1}}g_\ell(L),$$

This implies the corollary. □

3.2 | Construction of the auxiliary graph

In this section, we aim to give a generalization of Lemma 2.5 for longer cycles. We use a definition of *composed walk* from [15]. For every integer $k \geq 1$, call a $2k$-walk $x_0x_1...x_{2k}$ a *composed walk* if $x_0...x_k$ and $x_k...x_{2k}$ are two shortest paths and they are different but not necessarily vertex-disjoint or edge-disjoint. A composed walk is said to be *closed* if its endpoints are the same.

Lemma 3.5. *For integers ℓ, $\Delta \leq m$ and a constant $L \ll \Delta$, let F be an m-vertex graph with maximum degree Δ, such that for every $uv \in E(F)$ and $3 \leq k \leq 2\ell$, $c_k(u, v) \leq L$. Then for every vertex $u \in V(F)$ and every integer $2 \leq s \leq \ell - 1$, the number of closed composed walks of length $2s$ with endpoints u is at most*

$$\Delta^{s-1}\alpha_s(L)$$

for some constants $\alpha_s(L)$ which only depends on s and L.

Proof. For every vertex $u \in V(F)$ and every integer $2 \leq s \leq \ell - 1$, let $\mathcal{W}_s(u)$ be the set of closed composed walks of length $2s$ with endpoints u. For the case $s = 2$, the lemma is true with $\alpha_2(L) = L$. This is because that a closed composed walk of length 4 with endpoint u is exactly a 4-cycle containing u and then we have $|\mathcal{W}_2(u)| \leq \sum_{v \in N(u)} c_4(u, v) \leq \Delta L$.

Suppose for $s - 1 < \ell - 1$, the lemma is true for all integers $k \leq s - 1$, that is, for every $v \in V(F)$, $|\mathcal{W}_k(v)| \leq \Delta^{k-1}\alpha_k(L)$ with some constants $\alpha_k(L)$. Fix an arbitrary vertex $u \in V(F)$, and let

$$\mathcal{W}_s^i(u) = \{ux_1x_2...x_{2s-1}u \in \mathcal{W}_s(u)|i \text{ is the first integer such that } x_i = x_{2s-i}\}$$

for every $1 \leq i \leq s$. Then we have $\mathcal{W}_s(u) = \bigcup_{i=1}^s \mathcal{W}_s^i(u)$.

First, every composed walk $W \in \mathcal{W}_s^1(u)$ consists of an edge ux_1 and a closed composed walk of length $2s - 2$ with endpoints x_1. Therefore, we have

$$|\mathcal{W}_s^1(u)| \leq \sum_{x_i \in N(u)} |\mathcal{W}_{s-1}(x_1)| \leq \Delta^{s-1}\alpha_{s-1}(L).$$

Let $2 \leq i \leq s-1$. For every composed walk

$$W = ux_1x_2...x_{2s-1}u \in \mathcal{W}_s^i(u),$$

$\{ux_1...x_i\, x_{2s-(i-1)}...x_{2s-1}u\}$ forms a cycle C of length $2i$ containing u. Since for every $x_1 \in N(u)$, $c_{2i}(u, x_1) \leq L$, then the number of choices for C is at most ΔL. For a fixed C and $x_i \in C$, $W - C$ forms a path of length $(s-i)$ with endpoints x_i or a closed composed walks of length $2(s-i)$ with endpoints x_i. In the first case there are at most Δ^{s-i} choices, while in the later case there are at most $|\mathcal{W}_{s-i}(x_i)|$ choices. Therefore, we have

$$\begin{aligned} |\mathcal{W}_s^i(u)| &\leq \Delta L \cdot (\Delta^{s-i} + |\mathcal{W}_{s-i}(x_i)|) \\ &\leq \Delta L \cdot (\Delta^{s-i} + \Delta^{s-i-1}\alpha_{s-i}(L)) \\ &\leq 2\Delta^{s-i+1}L \leq 2\Delta^{s-1}L. \end{aligned}$$

Finally, every composed walk $W \in \mathcal{W}_s^s(u)$ is a cycle of length $2s$ containing u, and then we have

$$|\mathcal{W}_s^s(u)| \leq \sum_{v \in N(u)} c_{2s}(u, v) \leq \Delta L.$$

Hence, we have

$$|\mathcal{W}_s(u)| = \bigcup_{i=1}^{s} |\mathcal{W}_s^i(u)| \leq \Delta^{s-1}\alpha_{s-1}(L) + 2(s-2)\Delta^{s-1}L + \Delta L \leq \Delta^{s-1}\alpha_s(L)$$

for $\alpha_s(L) = \alpha_{s-1}(L) + 2(s-2)L + 1$, and the lemma follows by induction. $\qquad\square$

For an integer $\ell \geq 3$ and a graph F, denote by F^ℓ the multigraph defined on $V(F)$ such that for every distinct $u, v \in V(F^\ell)$, the multiplicity of uv in F^ℓ is the number of composed (u, v)-walks of length $2(\ell - 1)$ in F.

Lemma 3.6. *For an integer $\ell \geq 3$ and a constant $L > 0$, let n be a sufficiently large integer. Let m and d be the integers satisfying $m \leq n$ and $d \geq n^{1/\ell}/\log n$. Suppose F is an m-vertex graph with minimum degree $d - 1$, such that for every $uv \in E(F)$ and $3 \leq k \leq 2\ell$, $c_k(u, v) \leq L$. Then for every set $J \subseteq V(F)$ of size at least $2^\ell n/d^{\ell-1}$, we have*

$$e(F^\ell[J]) \geq \frac{d^{2\ell-2}|J|^2}{2^{2\ell+1}n}.$$

Proof. Let $\mathcal{W}$ be the set of composed walks of length $2(\ell - 1)$ with endpoints in J, and $\mathcal{W}_c$ be the set of closed composed walks of length $2(\ell - 1)$ with endpoint in J. By the definition of F^ℓ, we have

$$e(F^\ell[J]) = |\mathcal{W}| - |\mathcal{W}_c|.$$

By Lemma 3.5, we know that

$$|\mathcal{W}_c| \leq \Delta^{\ell-2}\alpha_{\ell-1}(L) \cdot |J|,$$

where Δ is the maximum degree of F, which, by Corollary 3.4, satisfies

$$\Delta \leq \frac{m}{d^{\ell-1}} \cdot g_\ell(L) \leq \frac{n}{d^{\ell-1}} \cdot g_\ell(L) \leq \frac{d^\ell \log^\ell n}{d^{\ell-1}} \cdot g_\ell(L) = d \log^\ell n \cdot g_\ell(L). \tag{17}$$

Now, it remains to estimate the lower bound of $\mathcal{W}$. For every $v \in J$, let a_v be the number of shortest paths of length $\ell - 1$ such that v is one of the endpoints. For every $u \in V(F)$, let $\mathcal{P}_u$ be the set of shortest paths of length $\ell - 1$ such that one endpoint is u and another endpoint is in J. Let $b_u = |\mathcal{P}_u|$ and then we have $\sum_{u \in V(F)} b_u = \sum_{v \in J} a_v$. By (16), we have

$$\sum_{u \in V(F)} b_u = \sum_{v \in J} a_v \geq \sum_{v \in J} (d/2)^{\ell-1} = \frac{d^{\ell-1}|J|}{2^{\ell-1}}.$$

Note that for every vertex $u \in V(F)$ and $P_1, P_2 \in \mathcal{P}_u$, $P_1 + P_2$ forms a composed walk in $\mathcal{W}$ and vice versa. Therefore, we have

$$|\mathcal{W}| = \sum_{u \in V(F)} \binom{b_u}{2} \geq m \binom{\frac{\sum_u b_u}{m}}{2} \geq m \binom{\frac{d^{\ell-1}|J|}{2^{\ell-1} \cdot m}}{2} \geq \frac{d^{2\ell-2}|J|^2}{2^{2\ell}m} \geq \frac{d^{2\ell-2}|J|^2}{2^{2\ell}n}$$

for $|J| \geq 2^\ell n/d^{\ell-1} \geq 2^\ell m/d^{\ell-1}$. Note that

$$\frac{|\mathcal{W}_c|}{|\mathcal{W}|} \leq \frac{\Delta^{\ell-2}\alpha_{\ell-1}(L) \cdot |J|}{\frac{d^{2\ell-2}|J|^2}{2^{2\ell}n}} \leq \frac{d^{\ell-2}\log^{\ell(\ell-2)}n}{d^{2\ell-2}} \cdot \frac{n}{|J|} \cdot g_\ell^{\ell-2}(L)\alpha_{\ell-1}(L)2^{2\ell}$$

$$\leq \frac{d^{\ell-2}\log^{\ell(\ell-2)}n}{d^{2\ell-2}} \cdot \frac{d^{\ell-1}}{2^\ell} \cdot g_\ell^{\ell-2}(L)\alpha_{\ell-1}(L)2^{2\ell}$$

$$\leq \frac{\log^{\ell(\ell-2)}n}{d} \cdot g_\ell^{\ell-2}(L)\alpha_{\ell-1}(L)2^\ell \ll 1,$$

when n is sufficiently large. Hence, we have

$$e(F^\ell[J]) = |\mathcal{W}| - |\mathcal{W}_c| \geq \frac{1}{2}|\mathcal{W}| \geq \frac{d^{2\ell-2}|J|^2}{2^{2\ell+1}n}. \qquad \square$$

Now, we start to define the auxiliary graph, which will be used in Lemma 3.10 in the next section. For every integer $k \geq 1$, call a path $x_0 x_1 ... x_{2k}$ a *composed path* if $x_0 ... x_k$ and $x_k ... x_{2k}$ are both shortest paths of length k. For an integer $\ell \geq 3$ and a graph F, denote by F_*^ℓ the simple graph defined on $V(F)$ such that for every distinct $u, v \in V(F_*^\ell)$, $uv \in E(F_*^\ell)$ if there is a composed (u, v)-path of length at most $2(\ell - 1)$ in F. To estimate the number of edges in F_*^ℓ, we need the following lemma.

Lemma 3.7. *For integers ℓ, $\Delta \leq m$ and a constant $L \ll \Delta$, let F be an m-vertex graph with maximum degree Δ, such that for every $uv \in E(F)$ and $3 \leq k \leq 2\ell$, $c_k(u, v) \leq L$. For every $1 \leq s \leq \ell - 1$ and every distinct $u, v \in V(F)$, the number of composed paths of length $2s$ with endpoints u, v is at most*

$$\Delta^{s-1}((sL+1)^s+1).$$

Proof. Let $\mathcal{P}$ be the set of composed paths of length $2s$ with endpoints u, v. For given vertices $a_1, \dots, a_{s-1}$, let

$$\mathcal{P}(a_1,\dots,a_{s-1}) = \{ux_1\dots x_{2s-1}v \in \mathcal{P} | x_1 = a_1,\dots,x_{s-1} = a_{s-1}\}.$$

Note that the number of nonempty $\mathcal{P}(a_1,\dots,a_{s-1})$ is at most Δ^{s-1}, since $ua_1\dots a_{s-1}$ is a path.

Suppose that $P_0 = ua_1\dots a_{2s-1}v$ is a composed path in $\mathcal{P}(a_1,\dots,a_{s-1})$. For every composed path $P = ua_1\dots a_{s-1}x_s\dots x_{2s-1}v \in \mathcal{P}(a_1,\dots,a_{s-1})\setminus\{P_0\}$, $a_{s-1}\dots a_{2s-1}v$ and $a_{s-1}x_s\dots x_{2s-1}v$ form a closed walk W of length $2(s+1)$. For every $s \le i \le 2s-1$, if $x_i = a_i$, the number of choices for x_i is 1. Otherwise, W contains an even cycle of length at most $2(s+1)$, which contains the edge $a_{i-1}a_i$ and vertex x_i. Since $c_{2k}(a_{i-1}, a_i) \le L$ for every $2 \le k \le s+1$, the number of choices for $x_i \ne a_i$ is at most sL. Therefore, we have

$$|\mathcal{P}(a_1,\dots,a_{s-1})| \le (sL+1)^s + 1.$$

Finally, we obtain

$$|\mathcal{P}| = \sum_{a_1,\dots,a_{s-1}} |\mathcal{P}(a_1,\dots,a_{s-1})| \le \Delta^{s-1}((sL+1)^s + 1). \qquad \square$$

Now, we give an upper bound on the multiplicity of F^ℓ.

Lemma 3.8. *For integers ℓ, $\Delta \le m$ and a constant $L \ll \Delta$, let F be an m-vertex graph with maximum degree Δ, such that for every $uv \in E(F)$ and $3 \le k \le 2\ell$, $c_k(u, v) \le L$. For every distinct u, $v \in V(F)$, the number of composed walks of length $2(\ell - 1)$ with endpoints u, v is at most*

$$\Delta^{\ell-2}\beta_\ell(L)$$

for a constant $\beta_\ell(L)$ which only depends on ℓ and L.

Proof. Let $\mathcal{W}$ be the number of composed walks of length $2(\ell - 1)$ in F with endpoints u, v. For every $1 \le i \le \ell - 1$, let

$$\mathcal{W}_i = \{ux_1\dots x_{2(\ell-1)-1}v \in \mathcal{W} | i \text{ is the first integer such that } x_i = x_{2(\ell-1)-i}\},$$

and then we have $\mathcal{W} = \bigcup_{i=1}^{\ell-1} \mathcal{W}_i$.

Let $1 \le i \le \ell - 2$. For every composed walk

$$W = ux_1\dots x_{\ell-1}\dots x_{2(\ell-1)-1}v \in \mathcal{W}_i,$$

$\{ux_1\dots x_i x_{2(\ell-1)-(i-1)}\dots x_{2(\ell-1)-1}v\}$ forms a composed path P of length $2i$. By Lemma 3.7, there are at most $\Delta^{i-1}[(iL+1)^i + 1]$ choices for P. For a fixed P, $W - P$ forms a path of

length $\ell - i - 1$ with endpoint x_i or a close composed walk of length $2(\ell - i - 1)$ with endpoint x_i. In the first case, there are at most $\Delta^{\ell-i-1}$ choices, while in the later case, by Lemma 3.5, there are at most $\Delta^{\ell-i-2}\alpha_{\ell-i-1}(L)$ choices. Therefore, we have

$$|\mathcal{W}_i| \leq \Delta^{i-1}((iL+1)^i+1)\cdot(\Delta^{\ell-i-1}+\Delta^{\ell-i-2}\alpha_{\ell-i-1}(L)) \leq 2\Delta^{\ell-2}((iL+1)^i+1).$$

Moreover, every walk $W \in \mathcal{W}_{\ell-1}$ is a composed path of length $2(\ell-1)$ with endpoints u and v. By Lemma 3.7, we have

$$|\mathcal{W}_{\ell-1}| \leq \Delta^{\ell-2}((\ell L - L + 1)^{\ell-1}+1).$$

Hence, we have

$$|\mathcal{W}| = \sum_{i=1}^{\ell-1}|\mathcal{W}_i| \leq \sum_{i=1}^{\ell-2}2\Delta^{\ell-2}((iL+1)^i+1) + \Delta^{\ell-2}((\ell L - L + 1)^{\ell-1}+1) = \Delta^{\ell-2}\beta_\ell(L)$$

for $\beta_\ell(L) = \sum_{i=1}^{\ell-2}2((iL+1)^i+1) + ((\ell L - L + 1)^{\ell-1}+1)$. $\qquad\square$

We have all the ingredients to give a lower bound on the number of edges in auxiliary graph F_*^l. This lemma will play the same role as Lemma 2.5 in the case of 4-cycles.

Lemma 3.9. *For an integer $\ell \geq 3$ and a constant $L > 0$, let n be a sufficiently large integer. Let m and d be the integers satisfying $m \leq n$ and $d \geq n^{1/\ell}/\log n$. Suppose F is an m-vertex graph with minimum degree $d - 1$, such that for every $uv \in E(F)$ and $3 \leq k \leq 2\ell$, $c_k(u, v) \leq L$. Then for every set $J \subseteq V(F)$ of size at least $2^\ell n/d^{\ell-1}$, we have*

$$e(F_*^\ell[J]) \geq \frac{d^\ell|J|^2}{n\log^{\ell(\ell-2)}n}f_\ell(L)$$

for a constant $f_\ell(L)$ which only depends on ℓ and L.

Proof. Note that every composed walk of length $2(\ell-1)$ with endpoints in J contains a composed path of length at most $2(\ell-1)$ with endpoints in J. Therefore, by Lemma 3.8, we have

$$e(F_*^\ell[J]) \geq \frac{e(F^\ell[J])}{\Delta^{\ell-2}\beta_\ell(L)},$$

where Δ is the maximum degree of F, which by (17), satisfies

$$\Delta \leq d\log^\ell n\cdot g_\ell(L).$$

Hence, we have

$$e(F_*^\ell[J]) \geq \frac{e(F^\ell[J])}{d^{\ell-2}\log^{\ell(\ell-2)}n\cdot g_\ell^{\ell-2}(L)\beta_\ell(L)} \geq \frac{d^\ell|J|^2}{n\log^{\ell(\ell-2)}n}f_\ell(L),$$

where $f_\ell(L) = \frac{1}{2^{2\ell+1}g_\ell^{\ell-2}(L)}\beta_\ell(L)$. $\qquad\square$

3.3 | Certificate lemma

In this section, we give our second main lemma, which will be used to build certificates for graphs with sparse short cycles. This lemma is a generalization of Lemma 2.6 for longer cycle, although the condition is slightly different. The idea of proof is also similar to Lemma 2.6, which originally comes from Kleitman and Winston [14] and Kohayakawa, Kreuter, and Steger [15].

Lemma 3.10. *For an integer $\ell \geq 3$ and constants L, $\alpha > 0$, let n be a sufficiently large integer. Let m and d be the integers satisfying $m \leq n$ and $(n^{1/\ell}/\log n) \leq d \leq \alpha n^{1/\ell}$. Suppose F is an m-vertex graph with minimum degree $d - 1$, such that for every $uv \in E(F)$ and $3 \leq k \leq 2\ell$, $c_k(u, v) \leq L$. Let $H = F_*^\ell$. Then for every set $I \subseteq V(F)$ of size d, such that $d_H(v, I) \leq (\ell - 1)L$ for all $v \in I$, there exist a set T and a set $C(T)$ depending only on T, not on I, such that*

  (i) $T \subseteq I \subseteq C(T)$,

 (ii) $|T| \leq n^{1/\ell}/\log n$,

(iii) $|C(T)| \leq (2^\ell + 1)n/d^{\ell-1}$.

Proof. This proof is similar to the proof of Lemma 2.6. We will describe a deterministic algorithm that associates to the set I a pair of sets T and $C(T)$.

We start the algorithm with sets $A_0 = V(H)$, $T_0 = \emptyset$ and a function $t_0(v) = 0$, for every $v \in V(H)$. In the ith iteration step, we pick a vertex $u_i \in A_i$ of maximum degree in $H[A_i]$. If $u_i \in I$, we define

$$t_{i+1}(v) = \begin{cases} t_i(v) + d_H(v, u_i) & \text{if } v \in A_i, \\ t_i(v) & \text{if } v \notin A_i, \end{cases}$$

and $Q = \{v | t_{i+1}(v) > (\ell - 1)L\}$, and let $T_{i+1} = T_i + u_i$, $A_{i+1} = A_i - u_i - Q$. Otherwise, let $T_{i+1} = T_i$, $A_{i+1} = A_i - u_i$ and $t_{i+1}(v) = t_i(v)$, for every $v \in V(H)$. The algorithm terminates at step K when we get a set A_K of size at most $2^\ell n/d^{\ell-1}$. We also assume that $u_{K-1} \in T_K$ as otherwise we can continue the algorithm until it is satisfied.

The algorithm outputs a vertex sequence $\{u_1, u_2, \ldots, u_{K-1}\}$, a set of 'representative' vertices T_K and a strictly decreasing set sequence $\{A_0, A_1, A_2, A_3, \ldots A_K\}$. Let $T = T_K$ and $C(T) = A_K \cup T$. From the algorithm, we have $T \subseteq I$. Furthermore, if a vertex v satisfies $t_i(v) > (\ell - 1)L$ for some i, then we have $d_H(v, I) \geq t_i(v) > (\ell - 1)L$, which implies $v \notin I$. Therefore, we maintain $I \subseteq A_i \cup T_i$ for every $i \leq K$ and especially get $I \subseteq A_K \cup T_K = C(T)$. Hence, Condition (i) is satisfied. Similarly as in Lemma 2.6, the set $C(T)$ only depends on T, not on I.

To finish the proof, it is sufficient to show that $|T_K| \leq n^{1/\ell}/\log n$. Once we prove it, we immediately obtain

$$|T| = |T_K| \leq \frac{n^{1/\ell}}{\log n},$$

and

$$|C(T)| = |A_K| + |T| \leq \frac{2^\ell n}{d^{\ell-1}} + \frac{n^{1/\ell}}{\log n} \leq \frac{(2^\ell + 1)n}{d^{\ell-1}},$$

which completes the proof.

In the rest of proof, we apply the same technique used in the proof of Lemma 2.6. We repeat the process as follows. Denote q the integer such that $n/2^q \le |A_K| < n/2^{q-1}$. By the choice of A_K, we have $q \le \log n$. For every integer $1 \le l \le q$, define A^l to be the first A-set satisfying

$$\frac{n}{2^l} \le |A^l| < \frac{n}{2^{l-1}}$$

if it exists, and let T^l be the corresponding T-set and $t^l(v)$ be the corresponding t-function of A^l. Note that A^l may not exist for every l, but A^q always exists and it could be that $A^q = A_K$. Suppose

$$A^{l_1} \supset A^{l_2} \supset \cdots \supset A^{l_p}$$

are all the defined A^l, where $p \le q$. By the above definition, we have $A^{l_1} = A_0$, $T^{l_1} = T_0$, and $l_p = q$. Define $A^{l_{p+1}} = A_K$, $T^{l_{p+1}} = T_K$. Now, we have

$$T_K = \bigcup_{j=2}^{p+1} (T^{l_j} - T^{l_{j-1}}). \tag{18}$$

To achieve our goal, we are going to estimate the size of $T^{l_j} - T^{l_{j-1}}$, for every $2 \le j \le p + 1$.

From the algorithm, we have $t^{l_j}(v) \le (\ell - 1)L$, for every $v \in A^{l_j} \cup T^{l_j}$. Moreover, for $v \in A^{l_{j-1}} - A^{l_j} - T^{l_j}$, suppose v is removed in step i, then we have

$$t^{l_j}(v) = t_{i-1}(v) + d_H(v, u_{i-1}) \le (\ell - 1)L + 1,$$

where u_i is the selected vertex in step i. Therefore, we obtain

$$\sum_{v \in A^{l_{j-1}}} t^{l_j}(v) \le ((\ell - 1)L + 1)|A^{l_{j-1}}| \le ((\ell - 1)L + 1)\frac{n}{2^{l_{j-1}-1}}. \tag{19}$$

Let $2 \le j \le p$. For every $u_i \in T^{l_j} - T^{l_{j-1}}$, u_i is chosen of maximum degree in $H[A_i]$, where A_i is a set between $A^{l_{j-1}}$ and A^{l_j}. By the choice of A^{l_j}, we have $|A_i| \ge n/2^{l_{j-1}}$. From Lemma 3.9, we obtain that

$$d_H(u_i, A_i) \ge \frac{d^\ell |A_i|}{n \log^{\ell(\ell-2)} n} f_\ell(L) \ge \frac{d^\ell}{2^{l_{j-1}} \log^{\ell(\ell-2)} n} f_\ell(L).$$

Note that $d_H(u_i, A_i)$ only contributes to $t^{l_j}(v)$, for $v \in A_i \subseteq A^{l_{j-1}}$. Then we obtain

$$|T^{l_j} - T^{l_{j-1}}| \frac{d^\ell}{2^{l_{j-1}} \log^{\ell(\ell-2)} n} f_\ell(L) \le \sum_{u_i \in T^{l_j} - T^{l_{j-1}}} d_H(u_i, A_i) \le \sum_{v \in A^{l_{j-1}}} t^{l_j}(v). \tag{20}$$

Combining (19) and (20), we have

$$|T^{l_j} - T^{l_{j-1}}| \frac{d^\ell}{2^{l_{j-1}} \log^{\ell(\ell-2)} n} f_\ell(L) \le ((\ell - 1)L + 1)\frac{n}{2^{l_{j-1}-1}},$$

which implies

$$|T^{l_j} - T^{l_{j-1}}| \leq \frac{2(\ell-1)L+1)}{f_\ell(L)} \cdot \frac{n \log^{\ell(\ell-2)} n}{d^\ell} \leq \frac{2((\ell-1)L+1)}{f_\ell(L)} \log^{\ell(\ell-1)} n \leq \frac{n^{1/\ell}}{\log^2 n}$$

for $2 \leq j \leq p$. For $j = p+1$, since we have $(n/2^q) \leq |A^{l_{p+1}}| \leq |A^{l_p}| \leq (n/2^{q-1})$, by a similar argument, we obtain that

$$|T^{l_{p+1}} - T^{l_p}| \frac{d^\ell |A_{l_{p+1}}|}{n \log^{\ell(\ell-2)} n} f_\ell(L) \leq \sum_{u_i \in T^{l_{p+1}-T^{l_p}}} d(u_i, A_i) \leq \sum_{v \in A^{l_p}} t^{l_{p+1}}(v)$$

$$\leq ((\ell-1)L+1)|A^{l_p}|,$$

which gives

$$|T^{l_{p+1}} - T^{l_p}| \leq \frac{2((\ell-1)L+1)}{f_\ell(L)} \cdot \frac{n \log^{\ell(\ell-2)} n}{d^\ell} \leq \frac{n^{1/\ell}}{\log^2 n}.$$

Finally, by (18), we have

$$|T_K| = \bigcup_{j=2}^{p+1} |T^{l_j} - T^{l_{j-1}}| \leq p \cdot \frac{n^{1/\ell}}{\log^2 n} \leq q \cdot \frac{n^{1/\ell}}{\log^2 n} \leq \frac{n^{1/\ell}}{\log n}. \qquad \square$$

3.4 | Proof of Theorem 1.5

This section is entirely devoted to the proof of Theorem 1.5. The idea is the same as the proof of Theorem 2.9: we will build a certificate for each graph in $\mathcal{G}_n(2\ell, L)$ and estimate the number of such certificates. Before we proceed, we first need the supersaturation result for $C_{2\ell}$ to give a bound on the min-degree sequence of graphs in $\mathcal{G}_n(2\ell, L)$. It was mentioned in [8] that Simonovits first proved the supersaturation for the even cycles, but the proof has not been published yet and it might appear in an upcoming paper of Faudree and Simonovits [9]. Morris and Saxton [20] recently provided a stronger version of supersaturation for even cycles. Very recently, Jiang and Yepremyan [12] give a supersaturation result of even linear cycles in linear hypergraphs, which includes the graph case. We use the graph version of their result and rephrase it in terms of the average degree.

Theorem 3.11 (Jiang and Yepremyan [12]). *For an integer $\ell \geq 2$, there exist constants C, c such that if G is an n-vertex graph with the average degree $d \geq 2Cn^{1/\ell}$, then G contains at least $c(d/2)^{2\ell}$ copies of $C_{2\ell}$.*

Corollary 3.12. *Let G be a n-vertex graph in $\mathcal{G}_n(2\ell, L)$, and $d_n, ..., d_1$ be the min-degree sequence of G. Then for every $i \in [n]$, we have*

$$d_i \leq \alpha n^{1/\ell}$$

for some constant $\alpha = \max\{2C, 2(L/2c)^{1/2\ell}\}$, where C, c are constants given in Theorem 3.11.

Proof. Suppose that there exists $k \in [n]$, such that $d_k > \alpha n^{1/\ell}$. Then by Theorem 3.11, the number of $C_{2\ell}$'s in G_k is at least

$$c\left(\frac{d_k}{2}\right)^{2\ell} > c\left(\frac{\alpha n^{1/\ell}}{2}\right)^{2\ell} \geq c\frac{L}{2c}n^2 \geq L\binom{k}{2},$$

which contradicts the fact that $G \in \mathcal{G}_n(2\ell, L)$. $\qquad\square$

Proof of Theorem 1.5. The way to construct the certificate is exactly same with in the proof of Theorem 2.9. Here we restate the process. For a graph $G \in \mathcal{G}_n(2\ell, L)$, let $Y_G := v_n < v_{n-1} < \cdots < v_1$ be the min-degree ordering of G and $D_G := \{d_n, d_{n-1},...,d_1\}$ be the min-degree sequence of G. Note that by Corollary 3.12, there exists a constant α such that $d_i \leq \alpha n^{1/\ell}$, for every $i \in [n]$. For every $i \in [n]$, let $G_i = G[v_i,...,v_1]$. Define the set sequence $S_G := \{S_n, S_{n-1},...,S_2\}$, where $S_i = N_G(v_i, G_{i-1})$. Note that $S_i \subseteq \{v_{i-1},...,v_1\}$ and $|S_i| = d_i$. By the construction, $[Y_G, D_G, S_G]$ uniquely determines the graph G and so we build a certificate $[Y_G, D_G, S_G]$ for G. To complete the proof, it is sufficient to estimate the number of such certificates.

We first choose a min-degree ordering $Y^* = v_n < v_{n-1} < \cdots < v_1$, and a min-degree sequence $D^* = \{d_n,...,d_1\}$; the number of options is at most

$$n!(\alpha n^{1/\ell})^n. \tag{21}$$

Next, we count set sequences $S = \{S_n, S_{n-1},...,S_2\}$, where $S_i \subseteq \{v_{i-1},...,v_1\}$ and $|S_i| = d_i$, such that the graph reconstructed by $[Y^*, D^*, S]$, denoted by G_S, are in $\mathcal{G}_n(2\ell, L)$. For every $2 \leq i \leq n$, let M_i be the number of choices for S_i with fixed sets $S_{i-1}, ...,S_2$. Define

$$\mathcal{I}_1 = \{i : d_i < n^{1/\ell}/\log n\}, \qquad \mathcal{I}_2 = \{i : d_i \geq n^{1/\ell}/\log n\}.$$

For every $i \in \mathcal{I}_1$, since $|S_i| = d_i < n^{1/\ell}/\log n$, we have a trivial bound

$$M_i \leq \binom{i-1}{d_i} \leq \binom{n}{n^{1/\ell}/\log n} \leq n^{n^{1/\ell}/\log n} = 2^{n^{1/\ell}}. \tag{22}$$

It remains to consider the upper bound on M_i for $i \in \mathcal{I}_2$. With fixed sets $S_{i-1}, ...,S_2$, the graph $G_{i-1} = G_S[v_{i-1},...,v_1]$ is uniquely determined. Since $G_{i-1} \subseteq G_S$ and $G_S \in \mathcal{G}_n(2\ell, L)$, for every $uv \in E(G_{i-1})$ and every $3 \leq k \leq 2\ell$, we know that $c_k(u, v; G_{i-1}) \leq L$. Note that every eligible S_i should satisfy $d_H(u, S_i) \leq (l-1)L$ for all $u \in S_i$, where $H = (G_{i-1})^l_*$. Otherwise, there exists a vertex $u \in S_i$, such that $\sum_{k=2}^{\ell} c_{2k}(v_i, u; G_i) \geq d_H(u, S_i) > (l-1)L$, which is a contradiction. Applying Lemma 3.10 on G_{i-1}, we obtain that every eligible S_i contains a subset T of size at most $n^{1/\ell}/\log n$, which uniquely determines a set $C(T) \supseteq S_i$ of size at most $(2^\ell + 1)n/d_i^{\ell-1}$. Since the number of choices for T is at most

$$\sum_{0 \leq j \leq n^{1/\ell}/\log n} \binom{i-1}{j} \leq 2\binom{i-1}{n^{1/\ell}/\log n} \leq 2\binom{n}{n^{1/\ell}/\log n} \leq 2^{n^{1/\ell}},$$

we then have

$$M_i \leq \sum_T \binom{C(T)}{d_i} \leq \sum_T \binom{\frac{(2^\ell + 1)n}{d_i^{\ell-1}}}{d_i} \leq \sum_T 2^{\frac{\ell}{1.88}((2^\ell+1)en)^{1/\ell}} \leq \sum_T 2^{3\ell n^{1/\ell}} \leq 2^{(3\ell+1)n^{1/\ell}} \quad (23)$$

for every $i \in I_2$, where the third inequality is given by Lemma 2.4.

Combining (22) and (23), we obtain that the number of choices for S is

$$\prod_{i=2}^n M_i \leq \prod_{i \in I_1} M_i \prod_{i \in I_2} M_i \leq 2^{n^{1+1/\ell}} 2^{(3\ell+1)n^{1+1/\ell}} \leq 2^{(3\ell+2)n^{1+1/\ell}}.$$

Hence, together with (21), the total number of certificates is at most

$$n!(\alpha n^{1/\ell})^n \prod_{i=2}^n M_i \leq 2^{n^{1+1/\ell}} 2^{(3\ell+2)n^{1+1/\ell}} \leq 2^{3(\ell+1)n^{1+1/\ell}}$$

for n sufficiently large, which leads to $|\mathcal{G}_n(2\ell, L)| \leq 2^{3(\ell+1)n^{1+1/\ell}}$. $\qquad\square$

4 | HYPERGRAPH ENUMERATION

In this section, we study the enumeration problems of r-graphs with given girth and r-graphs without C_4^r's. To prove them, we need a result on the linear Turán number of linear cycles given by Collier-Cartaino, Graber, and Jiang [6].

Theorem 4.1 (Collier-Cartaino et al [6]). *For every $r, \ell \geq 3$, there exists a constant $\alpha_{r,\ell} > 0$, depending on r and ℓ, such that*

$$\mathrm{ex}_L(n, C_\ell^r) \leq \alpha_{r,\ell} n^{1 + \frac{1}{\lceil \ell/2 \rceil}}.$$

4.1 | Proof of Theorem 1.3

Once we have Theorems 1.4 and 1.5, it is natural to think about reducing the hypergraph problems to problems on graphs and then apply our graph counting theorems.

Definition 4.2 (Shadow graph). Given a hypergraph H, the shadow graph of H, denoted by $\partial_2(H)$, is defined as

$$\partial_2(H) = \{D: |D| = 2, \exists e \in H, D \subseteq e\}.$$

Proposition 4.3. *Let $r \geq 3$, $\ell \geq 2$ and $H \in \mathrm{Forb}_L(n, r, 2\ell)$. For every r-element subset $S \in V(H)$, S forms an r-clique in $\partial_2(H)$ if and only if S is a hyperedge in H.*

Proof. Assume that there exists a r-clique with vertex set S in $\partial_2(H)$ and two edges e_1, e_2 such that e_1, e_2 lie on two different hyperedges f_1, f_2. Without loss of generality, we can

assume that e_1 and e_2 share a common vertex, as otherwise, we let $e_1 = ab$ and $e_2 = cd$ and one of the edge pairs $\{ab, ac\}$ or $\{ac, cd\}$ is contained in different hyperedges.

Let $e_1 = ab \subset f_1$ and $e_2 = ac \subset f_2$. Note that $c \notin f_1$ and $b \notin f_2$, as otherwise we have $f_1 = f_2$ by the linearity of H. Let f_3 be the hyperedge which includes bc. Then f_1, f_2, f_3 are distinct, and form a C_3^r by the linearity of H. This contradicts the fact that $H \in \mathrm{Forb}_L(n, r, 2\ell)$. $\qquad\square$

We also need the following short lemma on 4-cycles of the shadow graphs of hypergraphs in $\mathrm{Forb}_L(n, r, 4)$.

Lemma 4.4. *For every $r \geq 3$, there exists a constant $\beta = \beta(r)$ such that for every $H \in \mathrm{Forb}_L(n, r, 4)$, the shadow graph $\partial_2(H)$ contains at most $\beta n^{3/2}$ 4-cycles.*

Proof. Let $G = \partial_2(H)$. Since the girth of H is larger than 4, every 4-cycle in G must be contained in a hyperedge of H. By Theorem 4.1, we have $e(H) \leq \alpha_{r,4} n^{3/2}$. Hence, the number of 4-cycles in G is at most

$$\binom{r}{4} e(H) \leq \binom{r}{4} \alpha_{r,4} n^{3/2} = \beta n^{3/2}$$

for $\beta = \binom{r}{4} \alpha_{r,4}$. $\qquad\square$

Proof of Theorem 1.3 for $\ell = 2$. Define a map $\varphi\colon \mathrm{Forb}_L(n, r, 4) \to \mathcal{G} = \{\partial_2(H)\colon H \in \mathrm{Forb}_L(n, r, 4)\}$ given by $\varphi(H) = \partial_2(H)$. By Proposition 4.3, φ is a bijection. Note that by Lemma 4.4, every graph in $\mathcal{G}$ has at most $\beta n^{3/2}$ 4-cycles, where β is a constant depending on r. Applying Theorem 1.4, when n is sufficiently large, we have

$$|\mathcal{G}| \leq 2^{11 n^{3/2}}.$$

Hence, we obtain that $|\mathrm{Forb}_L(n, r, 4)| = |\mathcal{G}| \leq 2^{11 n^{3/2}}$ for n sufficiently large, which completes the proof.

Proof of Theorem 1.3 for $\ell \geq 3$. Define a map $\varphi\colon \mathrm{Forb}_L(n, r, 2\ell) \to \mathcal{G} = \{\partial_2(H)\colon H \in \mathrm{Forb}_L(n, r, 2\ell)\}$ given by $\varphi(H) = \partial_2(H)$. By Proposition 4.3, φ is a bijection. For a graph $G = \partial_2(H) \in \mathcal{G}$ and an edge $uv \in E(G)$, since the girth of H is larger than 2ℓ, each k-cycle in G, which contains edge uv, must be contained in a hyperedge of H, for all $3 \leq k \leq 2\ell$. Indeed, this hyperedge is unique by the linearity of H. Therefore, we have

$$c_k(u, v; G) \leq \binom{r - 2}{k - 2}$$

for all $3 \leq k \leq 2\ell$. Applying Theorem 1.5, when n is sufficiently large, we have

$$|\mathcal{G}| \leq 2^{3(\ell+1) n^{1+1/\ell}}.$$

Hence, we obtain that $|\mathrm{Forb}_L(n, r, 2\ell)| = |\mathcal{G}| \leq 2^{3(\ell+1) n^{1+1/\ell}}$ for n sufficiently large, which completes the proof.

4.2 | Proof of Theorem 1.1

We now estimate the number of r-graphs without C_4^r. The main idea is the same as in the previous section: we convert the hypergraph enumeration problem to a graph enumeration problem and then apply Theorem 1.4. However, because of the existence of C_3^r's, some facts we used before is no longer trivial and even not true. The first difficulty is to give an upper bound on the number of 4-cycles in shadow graphs, and we need the following lemma on the number of C_3^r's.

Lemma 4.5. *Let $r \geq 3$. For every $H \in \mathrm{Forb}_L(n, C_4^r)$ and every edge $e \in E(H)$, the number of C_3^r's in H containing e as an edge is at most*

$$\binom{r}{2}(4r^2 - 10r + 7).$$

Proof. For every distinct $u, v \in e$, let

$$C_{u,v} = \{\{e, f_i, g_i\} \subseteq H : e \cap f_i = \{u\}, e \cap g_i = \{v\}, |f_i \cap g_i| = 1\}.$$

Suppose $C_{u,v}$ is nonempty, and fix a $C_0 = \{e, f_0, g_0\} \in C_{u,v}$. For every $C = \{e, f_i, g_i\} \in C_{u,v} \setminus \{C_0\}$, we know that

$$(f_0 \cup g_0) \cap (f_i \cup g_i) - \{u, v\} \neq \emptyset,$$

otherwise, $\{f_0, g_0, f_i, g_i\}$ would form a C_4^r. Let w be a vertex in $(f_0 \cup g_0) \cap (f_i \cup g_i) - \{u, v\}$. Since $w \in f_0 \cup g_0 - \{u, v\}$, there are at most $2r - 3$ choices for w. By linearity of H, the number of linear 3-cycles in $C_{u,v}$ containing w is at most $2(r - 1)$. Therefore, we get

$$|C_{u,v}| \leq 1 + 2(r - 1)(2r - 3) = 4r^2 - 10r + 7. \tag{24}$$

Hence, the number of C_3^r's in H containing e as an edge is equal to

$$\sum_{u,v \in e} |C_{u,v}| \leq \binom{r}{2}(4r^2 - 10r + 7). \qquad \square$$

Lemma 4.6. *For every $r \geq 3$, there exists a constant $\beta = \beta(r)$ such that for every $H \in \mathrm{Forb}_L(n, C_4^r)$, the shadow graph $\partial_2(H)$ contains at most $\beta n^{3/2}$ 4-cycles.*

Proof. Since H is linear and contains no C_4^r, it is easy to check that for every 4-cycle of $\partial_2(H)$, its vertex set is contained in either a hyperedge or the vertex set of a C_3^r of H. By Lemma 4.5, there are at most

$$\frac{1}{3}\binom{r}{2}(4r^2 - 10r + 7)e(H)$$

C_3^r's in H, where $e(H) \leq \alpha_{r,4} n^{3/2}$ given by Theorem 4.1. Hence, the number of 4-cycles in $\partial_2(H)$ is at most

$$3\binom{r}{4}e(H) + 3\binom{3r-3}{4} \cdot \frac{1}{3}\binom{r}{2}(4r^2 - 10r + 7)e(H) \leq \beta n^{3/2}$$

for

$$\beta = \left[3\binom{r}{4} + \binom{3r-3}{4}\binom{r}{2}(4r^2 - 10r + 7)\right]\alpha_{r,4},$$

where $\alpha_{r,4}$ is a constant defined in Theorem 4.1. $\qquad\square$

Another difficulty is that the map we defined in the proof of Theorem 1.3 might be no longer injective. To overcome it, we have the following lemma to measure how far the map is from the injection.

Lemma 4.7. *For every $r \geq 3$, there exists a constant $\alpha = \alpha(r)$ such that for every $H \in \mathrm{Forb}_L(n, C_4^r)$, there are at most $\alpha n^{3/2}$ r-cliques in $\partial_2(H)$.*

Proof. Let $\mathcal{F}$ be the set of r-cliques in $\partial_2(H)$. For every $e \in E(H)$, let

$$\mathcal{F}_e = \{F \in \mathcal{F}: |F \cap e| = \max_{f \in E(H)} |F \cap f|\}.$$

Then we have $\mathcal{F} = \bigcup_{e \in H} \mathcal{F}_e$. Fix an arbitrary hyperedge $e \in H$. For every $2 \leq q \leq r$, let

$$\mathcal{R}_q = \{F \in \mathcal{F}_e: |F \cap e| = q\},$$

then we have $\mathcal{F}_e = \bigcup_{q=2}^r \mathcal{R}_q$.

First, it is trivial to get $|\mathcal{R}_r| = 1$. Let $2 \leq q \leq r-1$ and F be an r-clique in $\mathcal{R}_q$. Since $|F \cap e| = q$, the number of choices for $F \cap e$ is at most $\binom{r}{q}$. Given $F \cap e$, let u, v be two distinct vertices in $F \cap e$. For every $w \in F - e$, by the definition of the shadow graph and the linearity of H, there exist hyperedges f, g such that $\{e, f, g\}$ forms a C_3^r with $e \cap f = u$, $e \cap g = v$, and $f \cap g = w$. By (24), the number of such C_3^r's is at most $4r^2 - 10r + 7$. Therefore, the choices of w is at most $4r^2 - 10r + 7$. Hence, we have

$$\mathcal{R}_q \leq \binom{r}{q}(4r^2 - 10r + 7)^{r-q}.$$

Then, we obtain

$$|\mathcal{F}_e| = \sum_{q=2}^r |\mathcal{R}_q| \leq \sum_{r=2}^{r-1}\binom{r}{q}(4r^2 - 10r + 7)^{r-q} + 1 \leq 2^r(4r^2)^r.$$

Finally, we get

$$|\mathcal{F}| = \sum_{e \in E(H)} |\mathcal{F}_e| \leq 2^r(4r^2)^r e(\mathcal{H}) \leq \alpha n^{3/2}$$

for $\alpha = 2^r(4r^2)^r \alpha_{r,4}$, where $\alpha_{r,4}$ is the constant defined in Theorem 4.1. $\qquad\square$

Proof of Theorem 1.1. Define a map $\varphi: \mathrm{Forb}_L(n, C_4^r) \to \mathcal{G} = \{\partial_2(H): H \in \mathrm{Forb}_L(n, C_4^r)\}$ given by $\varphi(H) = \partial_2(H)$. By Lemma 4.6, every graph $G \in \mathcal{G}$ has at most $\beta n^{3/2}$ 4-cycles, where β is a constant depending on r. By Theorem 1.4, when n is sufficiently large, we have

$$|\mathcal{G}| \leq 2^{11n^{3/2}}.$$

By Lemma 4.7, for every $G \in \mathcal{G}$, the number of r-cliques in G is at most $\alpha n^{3/2}$, where α is a constant depending on r. Since every hyperedge corresponds to an r-clique in its shadow graph, we have

$$|\varphi^{-1}(G)| \leq 2^{\alpha n^{3/2}}.$$

Finally, we obtain

$$|\mathrm{Forb}_L(n, C_4^r)| \leq \sum_{G \in \mathcal{G}} |\varphi^{-1}(G)| \leq |\mathcal{G}| 2^{\alpha n^{3/2}} \leq 2^{(11+\alpha)n^{3/2}}$$

for n sufficiently large, which completes the proof.

ACKNOWLEDGMENTS

ORCID

Lina Li http://orcid.org/0000-0001-5740-1612

REFERENCES

1. J. Balogh and L. Li, *On the number of generalized Sidon sets,* available at https://arxiv.org/abs/1803.00659, preprint, 2018.
2. J. Balogh, R. Morris, and W. Samotij, *Independent sets in hypergraphs*, J. Amer. Math. Soc. **28** (2015), no. 3, 669–709.
3. J. Balogh, B. Narayanan, and J. Skokan, *The number of hypergraphs without linear cycles*, J. Combin. Theory Ser. B (2018).
4. F. A. Behrend, *On sets of integers which contain no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. USA **32** (1946), 331.
5. J. A. Bondy and M. Simonovits, *Cycles of even length in graphs*, J. Combin. Theory Ser. B **16** (1974), 97–105.
6. C. Collier-Cartaino, N. Graber, and T. Jiang, *Linear Turán numbers of linear cycles and cycle-complete ramsey numbers*, Combin. Probab. Comput. **27** (2018), 358–386.
7. P. Erdös, P. Frankl, and V. Rödl, *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs Combin. **2** (1986), 113–121.
8. P. Erdös and M. Simonovits, *Cube-supersaturated graphs and related problems*, Progress in Graph Theory, Academic Press, Toronto, ON, 1984, 203–218.

9. R. Faudree and M. Simonovits, *Cycle-supersaturated graphs*, in preparation.

10. Z. Füredi and T. Jiang, *Hypergraph Turán numbers of linear cycles*, J. Combin. Theory Ser. A **123** (2014), 252–270.

11. J. Han and Y. Kohayakawa, *On hypergraphs without loose cycles*, Discrete Math. **341** (2018), 946–949.

12. T. Jiang and L. Yepremyan, *Supersaturation of even linear cycles in linear hypergraphs*, available at https://arxiv.org/abs/1707.03091, preprint, 2017.

13. D. J. Kleitman and D. Wilson, On the number of graphs which lack small cycles (1996). unpublished manuscript.

14. D. J. Kleitman and K. J. Winston, *On the number of graphs without 4-cycles*, Discrete Math. **41** (1982), 167–172.

15. Y. Kohayakawa, B. Kreuter, and A. Steger, *An extremal problem for random graphs and the number of graphs with large even-girth*, Combinatorica **18** (1998), 101–120.

16. A. Kostochka, D. Mubayi, and J. Verstraëte, Personal communication.

17. A. Kostochka, D. Mubayi, and J. Verstraëte, *Turán problems and shadows I: Paths and cycles*, J. Combin. Theory Ser. A **129** (2015), 57–79.

18. B. Kreuter, Extremale und asymptotische graphentheorie für verbotene bipartite untergraphen, Ph.D. Thesis, Diplomarbeit, Forschungsinstitut für Diskrete Mathematik, Universität Bonn, 1994.

19. F. Lazebnik and J. Verstraëte, *On hypergraphs of girth five*, Electron. J. Combin. **10** (2003), 25.

20. R. Morris and D. Saxton, *The number of $C_{2\ell}$-free graphs*, Adv. Math. **298** (2016), 534–580.

21. D. Mubayi and L. Wang, *The number of triple systems without even cycles*, available at https://arxiv.org/abs/1701.00269, preprint, 2017.

22. C. Palmer et al., *Turán numbers for Berge-hypergraphs and related extremal problems*, available at https://arxiv.org/abs/1706.04249, preprint, 2017.

23. I. Z. Ruzsa and E. Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics **18** (1978), 939–945.

24. D. Saxton and A. Thomason, *Hypergraph containers*, Invent. Math. **201** (2015), 925–992.