# On Product Representations of Powers, I

P. ERDŐS, A. SÁRKÖZY AND V. T. SÓS

The solvability of the equation $a_1a_2 \cdots a_k = x^2$, $a_1, a_2, \ldots, a_k \in \mathcal{A}$ is studied for fixed $k$ and 'dense' sets $\mathcal{A}$ of positive integers. In particular, it is shown that if $k$ is even and $k \geq 4$, and $\mathcal{A}$ is of positive upper density, then this equation can be solved.



## 1. INTRODUCTION

Throughout this paper, we use the following notations. $\mathbb{N}$ denotes the set of the positive integers. If $f(n) = O(g(n))$, then we write $f(n) \ll g(n)$. $\pi(n)$ denotes the number of primes not exceeding $n$ so that, by the prime number theorem, we have $\pi(n) \sim n/\log n$. $\mu(n)$ denotes the Möbius function. Some further notations will be introduced in Sections 2 and 3.

A problem in number theory is said to be a hybrid problem if it involves both general sequences (characterized usually by density assumptions) and special sequences (squares, primes, etc.) of integers. In the last 15 years many problems of this type have been studied, and a survey of these results has been given in [11]. In particular, Lagarias, Odlyzko and Shearer [10] have studied the following problem: What density assumption is needed to ensure the solvability of the equation

$$a + a' = x^2, \qquad a, a' \in \mathcal{A}?$$

As the sequence $\mathcal{A} = \{1, 4, 7, \ldots, 3k + 1, \ldots\}$ shows, it is not enough to assume that $\mathcal{A}$ is of positive (lower) density. Examples of similar type show that it does not help to take more summands on the left-hand side; i.e. for all $k \in \mathbb{N}$ there is a set $\mathcal{A}$ of positive density such that

$$a_1 + a_2 + \cdots + a_k = x^2, \qquad a_1, a_2, \ldots, a_k \in \mathcal{A}$$

cannot be solved. In this paper we will study the multiplicative analogue of this problem by studying the solvability of the equation

$$a_1a_2 \cdots a_k = x^2, \quad a_1, a_2, \ldots, a_k \in \mathcal{A}, \qquad a_1 < a_2 < \cdots < a_k, \quad x \in \mathbb{N}. \tag{1.1}$$

It will turn out that the solvability of this equation strongly depends on the parity of $k$. If $k$ is even and $k \geq 4$ then, unlike the additive case, in order to ensure the solvability of (1.1) it suffices to assume that $\mathcal{A}$ is of positive (upper) density (indeed, a much weaker assumption is enough).

If $k \geq 2$ and $\mathcal{A}$ is a set of positive integers such that equation (1.1) cannot be solved, then $\mathcal{A}$ is said to have property $P_k$, and $\Gamma_k$ denotes the family of those subsets of $\mathbb{N}$ which have property $P_k$. We write

$$F_k(n) = \max_{\substack{\mathcal{A} \subset \{1, 2, \ldots, n\} \\ \mathcal{A} \in \Gamma_k}} |\mathcal{A}|. \tag{1.2}$$

(In other words, $t = F_k(n) + 1$ is the smallest positive integer such that, for every set $\mathscr{A}$ with $\mathscr{A} \subset \{1, \ldots, n\}$, $|\mathscr{A}| = t$, equation (1.1) can be solved.) Moreover, we write

$$L_k(n) = \max_{\substack{\mathscr{A} \subset \{1, 2, \ldots, n\} \\ \mathscr{A} \in \Gamma_k}} \sum_{a \in \mathscr{A}} \frac{1}{a}.$$

In this paper, our goal is to study the functions $F_k(n)$ and $L_k(n)$, while in Part II we will study the analogous problems with higher powers instead of squares in (1.1).

It will turn out that for fixed $k$ and $n \to +\infty$ we have $F_{2k+1}(n) \gg n$ for all $k$ and, on the other hand, $F_{2k}(n) = o(n)$ for $k \geqslant 2$. Moreover, the asymptotics for $F_{2k}(n)$ depends on the parity of $k$.

We will prove the following theorems:

THEOREM 1. *For all $n \in \mathbb{N}$, $F_2(n)$ is equal to the number of the square-free integers not exceeding $n$*:

$$F_2(n) = \sum_{i \leqslant n} \mu^2(i) \sim \frac{6}{\pi^2} n. \tag{1.3}$$

THEOREM 2. *For $\varepsilon > 0$, $n > n_0(\varepsilon)$, we have*

$$n - n(\log n)^{(e/2)\log 2 - 1 + \varepsilon} < F_3(n) < n - n(\log n)^{-1-\varepsilon}. \tag{1.4}$$

THEOREM 3. *There is a positive absolute constant c and, for all $\varepsilon > 0$, a number $n_0(\varepsilon)$ such that for $n > n_0(\varepsilon)$ we have*

$$(2^{\frac{1}{2}} - \varepsilon)n^{\frac{3}{4}}(\log n)^{-\frac{3}{2}} < F_4(n) - \pi(n) < cn^{\frac{3}{4}}(\log n)^{-\frac{3}{2}}. \tag{1.5}$$

THEOREM 4. *There is an absolute constant c and, for all $\varepsilon > 0$, a number $n_0(\varepsilon)$ such that for $n > n_0(\varepsilon)$ we have*

$$(2^{\frac{1}{2}} - \varepsilon)n^{\frac{2}{3}}(\log n)^{-\frac{4}{3}} < F_6(n) - (\pi(n) + \pi(n/2)) < cn^{\frac{7}{9}} \log n. \tag{1.6}$$

THEOREM 5. *There is a positive absolute constant c and, for all $k \in \mathbb{N}$, there exist absolute constants $c_k > 0$ and $n_0(k)$ such that for $n > n_0(k)$ we have*

$$c_k(n^{\frac{1}{2}}(\log n)^{-1})^{1+(4k+1)^{-1}} < F_{4k}(n) - \pi(n) < cn^{\frac{3}{4}}(\log n)^{-\frac{3}{2}}.$$

THEOREM 6. *There is a positive absolute constant c and, for all $k \in \mathbb{N}$, there exist absolute constants $c_k > 0$ and $n_0(k)$ such that for $n > n_0(k)$ we have*

$$c_k(n^{\frac{1}{2}}(\log n)^{-1})^{1+(4k+1)^{-1}} < F_{4k+2}(n) - (\pi(n) + \pi(n/2)) < cn^{\frac{7}{9}} \log n. \tag{1.7}$$

THEOREM 7. *For all $k \in \mathbb{N}$, $k < 1$ and $\varepsilon > 0$, there is a number $n_0(k, \varepsilon)$ such that for $n > n_0(k, \varepsilon)$ we have*

$$(\log 2 - \varepsilon)n < F_{2k+1}(n) < n - (1 - \varepsilon)n(\log n)^{-2}. \tag{1.8}$$

The lower bound in (1.8) could be improved slightly (see the remark following the proof of Theorem 7): however, this would take a lengthy computation, and since we have not been able to decide whether $F_{2k+1}(n) \sim n$, thus we have preferred to work out the simpler version in (1.8).

There is a considerable gap between the lower and upper bounds in (1.8) for $F_{2k+1}(n)$ that we have not been able to eliminate for $k \geq 2$. On the other hand, we will prove much more satisfactory estimates for $L_{2k+1}(n)$:

THEOREM 8. *If $k$ is a fixed positive integer and $n \to +\infty$, then we have*

$$L_{4k}(n) = (1 + o(1)) \log \log n, \tag{1.9}$$

$$L_{4k+2}(n) = (\tfrac{3}{2} + o(1)) \log \log n \tag{1.10}$$

*and*

$$L_{2k+1}(n) = 1 + (\tfrac{1}{2} + o(1)) \log n. \tag{1.11}$$

## 2. COMBINATORIAL LEMMAS

In the proofs we will use Turán type extremal graph theorems for cycles. In the following lemmas we give a list of these.

$G_n^e(V; E)$ will denote a graph with vertex set $V$ and edge set $E$, $|V| = n$ and $|E| = e$. The degree of the vertex $P$ will be denoted by $d(P)$. $G_{u,v}^e(U, V; E)$ will denote a bipartite graph with vertex set $U \cup V$ ($U \cap V = \emptyset$) and $|U| = u$, $|V| = v$ and $|E| = e$. $K_{u,v}$ will denote the complete bipartite graph. $C_l$ denotes the cycle of length $l$, and we also use $K_3$ instead of $C_3$.

We shall need the following well-known and nearly trivial fact.

For $k, n \in \mathbb{N}$, let $q_k(n)$ denote the smallest positive integer $q$ such that every graph of $n$ vertices and $q$ edges contains a $C_k$.

LEMMA 1. (a) *For $n \in \mathbb{N}$, $n \to +\infty$, we have*

$$q_4(n) = (\tfrac{1}{2} + o(1)) n^{\frac{3}{2}}. \tag{2.1}$$

(b) *There is a positive absolute constant $c$ such that, for all $n \in \mathbb{N}$, we have*

$$q_6(n) < cn^{\frac{4}{3}}. \tag{2.2}$$

(c) *For fixed $k \in \mathbb{N}$, and for $n \in \mathbb{N}$, $n \to +\infty$, we have*

$$q_{2k+1}(n) = (\tfrac{1}{4} + o(1)) n^2. \tag{2.3}$$

PROOF. (a) See [3] or [8].

(b) This is a special case of a result of Bondy and Simonovits [2] (see also [12, Corollary 6.13]).

(c) This is a special case of a result of Erdős and Simonovits [7] (see also [12, Theorem 3.1]).

LEMMA 2. *For all $\varepsilon > 0$, there is a number $n_0 = n_0(\varepsilon)$ such that if $n > n_0(\varepsilon)$, then there is a graph $G_n^e$ with*

$$e > (\tfrac{1}{2} - \varepsilon) n^{\frac{4}{3}}$$

*which contains no cycles $C_l$ with $3 \leq l \leq 6$.*

PROOF. See [1], [13] or [12, p. 184].

LEMMA 3. *If $k \geq 3$, then there is a positive constant $c_k$ such that, for every $n \geq 2$, there is a graph $G_n^e$ with*

$$e > c_k n^{1+(k-1)^{-1}}$$

*which does not contain a cycle $C_l$ with $3 \leq l \leq k$.*

PROOF. This is Corollary 8.3 in [12].

For $n \in \mathbb{N}$, let $r(n)$ denote the smallest positive integer $r$ such that if $G_{u,v}^r(U, V; E)$ is a bipartite graph of $r$ edges,

$$v \leq u \leq v^2 \qquad \text{and} \qquad uv \leq n, \tag{2.4}$$

then $G$ must contain a $C_6$. We conjecture that

$$r(n) < cn^{\frac{2}{3}}; \tag{2.5}$$

unfortunately, we have not been able to show this. We could prove only the following weaker result:

LEMMA 4. *There is an absolute constant $c$ such that, for all $n \in \mathbb{N}$, we have*

$$r(n) < cn^{\frac{7}{9}}.$$

PROOF. We have to show that if $c_1$ is large enough, $G = G_{u,v}^r$ is a bipartite graph which satisfies (2.4), and

$$r = |E| \geq c_1 n^{\frac{7}{9}}, \tag{2.6}$$

then $G_{u,v}^r$ contains a $C_6$.

Let $U = \{P_1, P_2, \ldots, P_u\}$, $V = \{Q_1, Q_2, \ldots, Q_v\}$. By (2.6), we have

$$\sum_{i=1}^{u} d(P_i) = |E| \geq c_1 n^{\frac{7}{9}}.$$

It follows that

$$\sum_{j=0}^{[u/v]-1} \sum_{i=1}^{v} d(P_{jv+i}) + \sum_{i=1}^{v} d(P_{u-v+i}) \geq \sum_{i=1}^{u} d(P_i) \geq c_1 n^{\frac{7}{9}}.$$

Thus there is an integer $m$ with $0 \leq m \leq u - v$ such that

$$\sum_{i=1}^{v} d(P_{m+i}) \geq \frac{c_1 n^{\frac{7}{9}}}{[u/v]+1} \geq \frac{c_1}{2} \frac{vn^{\frac{7}{9}}}{u}. \tag{2.7}$$

Let $G^*$ denote the subgraph of $G$ induced by the $2v$ vertices $U^* = \{P_{m+1}, P_{m+2}, \ldots, P_{m+v}\}$, $V^* = \{Q_1, Q_2, \ldots, Q_v\}$. By (2.4) and (2.7) we have

$$e^* \geq \frac{c_1}{2} \frac{vn^{\frac{7}{9}}}{u} \geq \frac{c_1}{2} v^{\frac{4}{3}}. \tag{2.8}$$

If $c_1 = 8c$, where $c$ is the constant in (2.2), then by (2.8) we have

$$e^* > c(2v)^{\frac{4}{3}},$$

and thus, by Lemma 1, $G^*$ contains a $C_6$. For $u, v, n \in \mathbb{N}$ and

$$uv \leq n, \tag{2.9}$$

let $s(u, v, n)$ denote the smallest positive integer $s$ such that if $G = G^s_{u,v}(U, V; E)$ is a bipartite graph of $s$ edges, then $G$ must contain a $C_6$.

We conjecture that for

$$v^2 < u \tag{2.10}$$

we have

$$s(u, v, n) < 2u + cn^{\frac{2}{3}}. \tag{2.11}$$

(Clearly, $s(u, v) > 2u$, as the following example shows: consider the graph obtained by joining each of the vertices in $\mathcal{U}$ to two fixed vertices in $\mathcal{V}$.)

Note that, of course, conjectures (2.5) and (2.11) can be combined: if $G = G^e_{(u,v)}(U, V; E)$ is a bipartite graph $v \leq u$, $uv \leq n$ and

$$e \geq 2u + cn^{\frac{2}{3}}, \tag{2.12}$$

then $G$ contains a $C_6$.

Unfortunately, we have not been able to prove (2.11). We could prove only the following weaker result:

Lemma 5. *If $u$, $v$ and $n$ satisfy* (2.9) *and* (2.10), *then we have*

$$s(u, v, n) \leq \begin{cases} 2u + v^3 & \text{for } v^3 \leq 8u, \\ 18vu^{\frac{2}{3}} & \text{for } v^3 > 8u. \end{cases}$$

Proof. We have to show that if $u, v$ *and* $n$ satisfy (2.9) and (2.10), and $G = G^e_{(u,v)}(U, V; E)$ is a bipartite graph with

$$e \geq \begin{cases} 2u + v^3 & \text{for } v^3 \leq 8u, \\ [18vu^{\frac{2}{3}}] & \text{for } v^3 > 8u, \end{cases} \tag{2.13}$$

then $G$ contains a $C_6$. Clearly, it suffices to show that there is a $K(3, 3)$ in $G$.

Assume that contrary to the assertion (2.13) holds; however, $G$ does not contain a $K(3, 3)$. Define the integer $l$ by

$$l = \begin{cases} 2 & \text{for } v^3 \leq 8u, \\ [vu^{-\frac{1}{3}}] & \text{for } v^3 > 8u, \end{cases} \tag{2.14}$$

and put

$$\mathcal{I}_1 = \{i : 1 \leq i \leq u,\ d(P_i) \leq l\}$$

and

$$\mathcal{I}_2 = \{i : 1 \leq i \leq u,\ d(P_i) > l\}$$

so that

$$\{1, 2, \ldots, u\} = \mathcal{I}_1 \cup \mathcal{I}_2, \qquad \mathcal{I}_1 \cap \mathcal{I}_2 = \emptyset. \tag{2.15}$$

Let $U = \{P_1, P_2, \ldots, P_u\}$, $V = \{Q_1, Q_2, \ldots, Q_v\}$. By the definition of $\mathcal{I}_1$, we have

$$\sum_{i \in \mathcal{I}_1} d(P_i) \leq lu. \tag{2.16}$$

For all $i \in \mathcal{I}_2$, there are $\binom{d(P_i)}{3}$ triples $Q_x, Q_y, Q_z$ $(1 \leq x < y < z \leq v)$ such that each of $Q_x, Q_y, Q_z$ is joined to $P_i$. On the other hand, since there is no $K(3, 3)$ in $G$, thus each of the $\binom{v}{3}$ triples $Q_x$, $Q_y$, $Q_z$ $(1 \le x < y < z \le v)$ can be joined to at most two distinct $P_i$'s. Thus we have

$$\sum_{i \in \mathcal{I}_2} \binom{d(P_i)}{3} \le 2\binom{v}{3} < \tfrac{1}{3}v^3. \tag{2.17}$$

For $i \in \mathcal{I}_2$ we have $d(P_i) \ge l + 1$, and thus

$$\binom{d(P_i)}{3} \ge \tfrac{1}{6}d(P_i)l(l-1) \ge \tfrac{1}{12}d(P_i)l^2$$

(since $l \ge 2$). Thus it follows from (2.17) that

$$\sum_{i \in \mathcal{I}_2} d(P_i) < 4v^3 l^{-2}. \tag{2.18}$$

We obtain from (2.15), (2.16) and (2.18) that

$$e(G) = \sum_{i=1}^{u} d(P_i) = \sum_{i \in \mathcal{I}_1} d(P_i) + \sum_{i \in \mathcal{I}_2} d(P_i) < lu + 4v^3 l^{-2}$$

so that, in view of (2.14),

$$e(G) < 2u + v^3 \qquad \text{for } v^3 \le 8u$$

and

$$e(G) \le (vu^{-\frac{1}{3}})u + 4v^3(vu^{-\frac{1}{3}}/2)^{-2} = 17vu^{\frac{2}{3}} \le 18vu^{\frac{2}{3}} - 1 < [18vu^{\frac{2}{3}}] \qquad \text{for } v^3 > 8u,$$

which contradicts (2.13), and thus completes the proof of the lemma.

Lemma 6. *Let $k$, $t$ and $n$ be positive integers with $k \ge 3$, let $S = \{s_1, s_2, \dots, s_n\}$, and let $S_1, S_2, \dots, S_t$ be distinct subsets of $S$. For $j = 1, 2, \dots, n$, $R_1 \subset S, \dots, R_l \subset S$, write*

$$f_j(R_1, \dots, R_l) = |\{x : 1 \le x \le l,\ s_j \in R_x\}|.$$

*If*

$$t \ge 2^{n-1} + k - 1, \tag{2.19}$$

*then there are subsets $S_{i_1}, S_{i_2}, \dots, S_{i_k}$ (with $1 \le i_1 < i_2 < \cdots < i_k \le t$) such that*

$$f_j(S_{i_1}, S_{i_2}, \dots, S_{i_k}) \text{ is even for } j = 1, 2, \dots, n. \tag{2.20}$$

Note that (2.19) is best possible for $k = 3$, as the following example shows: if $n$ is odd and $S_1, S_2, \dots, S_{2^{n-1}}$ are those subsets of $S$ the cardinality of which is odd and $S_{2^{n-1}+1} = \emptyset$, then there is no triple $S_{i_1}, S_{i_2}, S_{i_3}$ (with $i_1 < i_2 < i_3$) satisfying (2.20).

For fixed $k$ and $n$, let $\varphi_k(n)$ denote the smallest integer $t$ for which the conclusion of the lemma holds. Then, by (2.19) and the above example, we have

$$\varphi_3(2l+1) = 2^{2l} + 2.$$

Moreover, if $n \in \mathbb{N}$, $k$ is odd and $S_1, S_2, \dots, S_{2^{n-1}}$ are the subsets containing $s_1$, then there is no $k$-tuple of them satisfying (2.20), which shows that

$$2^{n-1} + 1 \le \varphi_{2l+1}(n) \qquad (\le 2^{n-1} + 2l).$$

If $k$ is even, then the situation is different. We will study this case in a subsequent paper.

Proof. Assume to the contrary that (2.19) holds; however, there are no subsets $S_{i_1}, S_{i_2}, \dots, S_{i_k}$ satisfying (2.20).

By $S_{k-2} \neq S_{k-1}$, at least one of the following statements holds:

$$\begin{aligned} &\text{there is a } j \text{ such that } 1 \leq j \leq n \text{ and } f_j(S_1, S_2, \ldots, S_{k-3}, S_{k-2}) \text{ is odd;} \\ &\text{there is a } j \text{ such that } 1 \leq j \leq n \text{ and } f_j(S_1, S_2, \ldots, S_{k-3}, S_{k-1}) \text{ is odd.} \end{aligned} \tag{2.21}$$

Without loss of generality, we may assume that (2.21) holds. Then, for each of $u = k-1, k, \ldots, t$, there is a uniquely determined subset $T_u$ of $S$ such that

$$f_i(S_1, S_2, \ldots, S_{k-2}, S_u, T_u) \text{ is even for } j = 1, 2, \ldots, n. \tag{2.22}$$

Then, clearly,

$$T_u \neq T_v \qquad \text{for } k-1 \leq u < v \leq t \tag{2.23}$$

and, by (2.21) and (2.22),

$$S_u \neq T_u \qquad \text{for } k-1 \leq u \leq t. \tag{2.24}$$

Let $U$ denote the set of the integers $u$ such that

$$k-1 \leq u \leq t \tag{2.25}$$

and

$$T_u \neq S_1, \ldots, T_u \neq S_{k-2}. \tag{2.26}$$

There are $t-k+2$ values of $u$ satisfying (2.25) and, in view of (2.23), with at most $k-2$ exceptions all these $u$'s also satisfy (2.26), so that we have

$$|U| \geq (t-k+2) - (k-2) = t - 2k + 4. \tag{2.27}$$

By (2.24) and (2.26), for all $u \in U$ the subsets $S_1, S_2, \ldots, S_{k-2}, S_u, T_u$ are pairwise distinct. Thus, by our indirect assumption, (2.22) implies that $T_u$ is different from each of $S_1, S_2, \ldots, S_t$. Then, in view of (2.23), $S_1, S_2, \ldots, S_t$ and the $T_u'$s with $u \in U$ are pairwise distinct subsets of $S$. On the other hand, by (2.19) and (2.27), their total number is

$$t + |U| \geq 2t - 2k + 4 \geq 2(2^{n-1} + k - 1) - 2k + 4 = 2^n + 2$$

which is greater than the total number of the distinct subsets of $S$, and this contradiction completes the proof of Lemma 6.

## 3. Arithmetic Lemmas

Lemma 7. *Let $\mathcal{P}$ be a set of $t$ prime numbers $p_1 < p_2 < \cdots < p_t$, and let $\mathcal{A}$ be a set of positive integers all the elements $a$ of which can be represented in the form $a = p_i p_j$, with $i \neq j$. Define the graph $G[\mathcal{A}]$ on the $t$ vertices $P_1, \ldots, P_t$ so that $P_iP_j \in E(G)$ iff $p_ip_j \in \mathcal{A}$. Then $\mathcal{A} \notin \Gamma_k$, i.e. (1.1) can be solved iff the graph $G[\mathcal{A}]$ contains a subgraph $H^k$ of $k$ edges such that the degree of every vertex of it is a positive even integer.*

Proof. This follows easily from the fundamental theorem of arithmetics. Assume that $a_1, \ldots, a_k \in \mathcal{A}$, $a_1 < \cdots < a_k$, and, for $i = 1, 2, \ldots, k$, let $a_i = p_{j_i} p_{l_i}$ (where $p_{j_i}$, $p_{l_i} \in \mathcal{P}$, $j_i \neq l_i$). Then consider the subgraph $H^k$ the $k$ edges of which are $(P_{j_1}, P_{l_1})$, $(P_{j_2}, P_{l_2}), \ldots; (P_{j_k}, P_{l_k})$ (and the vertices of which are the end vertices of these edges). $a_1 a_2 \cdots a_k$ is a square iff the degree of every vertex of $H^k$ is a positive even integer, and this completes the proof.

Lemma 8. *Let $G = G(V; E)$ be a graph, with $V = \{P_1, P_2, \ldots, P_t\}$. Assume that two mappings $f\colon V \to \mathbb{N}$ and $g\colon E \to \mathbb{N}$ are given, with the following properties:*
(i) *if $1 \le i < j \le t$, then $f(P_i) \ne f(P_j)$;*
(ii) *if $e \in E$ and the end vertices of $e$ are $P_i$ and $P_j$, then $g(e) = f(P_i)f(P_j)$;*
(iii) *if $e \in E$, $e' \in E$ and $e \ne e'$, then $g(e) \ne g(e')$.*
*Denote the range of $E$ by $\mathcal{A}$: $\mathcal{A} = \{g(e)\colon e \in E\}$. Then, if $G$ contains a $C_k$, $\mathcal{A} \notin \Gamma_k$.*

Proof. If $G$ contains the cycle of length $k$ the edges of which are $e_1, e_2, \ldots, e_k$ and the vertices of which are $P_{i_1}, P_{i_2}, \ldots, P_{i_k}$, then

$$g(e_1)g(e_2)\cdots g(e_k) = (f(P_{i_1})f(P_{i_2})\cdots f(P_{i_k}))^2.$$

Lemma 9. *Using the same notations as in Lemma 8, assume that* (i) *and* (ii) *in Lemma 8 hold, but replace* (iii) *by the following*:
(iii′) *if $e \in E$, $e' \in E$, $e \ne e'$ and $e$ and $e'$ are adjacent edges, then $g(e) \ne g(e')$. Then, if $G$ contains a triangle, $\mathcal{A} \notin \Gamma_3$.*

Proof. If $G$ contains a triangle the edges of which are $e_1$, $e_2$ and $e_3$, and the vertices of which are $P_{i_1}$, $P_{i_2}$ and $P_{i_3}$, then $g(e_1)$, $g(e_2)$ and $g(e_3)$ are distinct integers. Moreover, we have

$$g(e_1)g(e_2)g(e_3) = (f(P_{i_1})f(P_{i_2})f(P_{i_3}))^2.$$

Lemma 10. *Let $\mathcal{P} = \{p_1, p_2, \ldots, p_t\}$ be a finite set of distinct primes, and let $\mathcal{A}$ be a set of distinct integers all the elements of which are of the form $p_ip_j$, with $p_i \in \mathcal{P}$, $p_j \in \mathcal{P}$, $i \ne j$. Let $G[\mathcal{A}]$ denote the graph with $V = \{P_1, \ldots, P_t\}$ and $E = \{(P_i, P_j)\colon p_ip_j \in \mathcal{A}\}$. Then*:
(i) *$\mathcal{A} \in \Gamma_3$ iff $G[\mathcal{A}]$ does not contain a triangle*;
(ii) *$\mathcal{A} \in \Gamma_4$ iff $G[\mathcal{A}]$ does not contain a $C_4$*;
(iii) *if $k \in \mathbb{N}$, $k \ge 3$, and $G[\mathcal{A}]$ does not contain a $C_l$ with $3 \le l \le k$, then $\mathcal{A} \in \Gamma_k$.*

Proof. By Lemma 7, $\mathcal{A} \notin \Gamma_k$ iff $G[\mathcal{A}]$ contains a subgraph $H^k$ of $k$ edges such that the degree of every vertex of it is a positive even integer. For $k = 3$ and 4, the only graphs $H^k$ with these properties are $K_3$, resp. $C_4$ which proves (i) and (ii). Moreover, if $\mathcal{A} \notin \Gamma_k$, then since the degree $d(P_i)$ of every vertex $P_i$ of $H^k$ is a positive even integer, thus $d(P_i) \ge 2$ for every vertex $P_i$. $H^k$ contains a $C_l$ with $3 \le l \le k$, and this completes the proof of (iii).

The number of distinct prime factors of $n$ will be denoted by $\omega(n)$, and $\Omega(n)$ will denote the number of prime factors of $n$ counted with multiplicity. Moreover, for $0 < x$, we write $\phi(x) = 1 + x\log x - x$.

Lemma 11. *If $0 < y \le 1$ and $\varepsilon > 0$, then for $x > x_0(\varepsilon)$ we have*

$$\frac{x}{(\log x)^{\phi(y)+\varepsilon}} < |\{m\colon m \le x,\ \Omega(m) \le y\log\log x\}| < \frac{x}{(\log x)^{\phi(y)-\varepsilon}}.$$

Proof. This follows from a result of Hardy and Ramanujan [9].

Lemma 12. *If $1 < y < 2$ and $\varepsilon > 0$, then for $x > x_0(\varepsilon)$ we have*

$$|\{m\colon m \le x,\ \Omega(m) \ge y\log\log x\}| < \frac{x}{(\log x)^{\phi(y)-\varepsilon}}.$$

PROOF. See Corollary 2 in [6].

LEMMA 13. *Let $\mathscr{B}$ denote the set of the integers $b$ such that $b \leq n$ and $b$ can be represented in the form*

$$b = uv \tag{3.1}$$

*with integers $u$ and $v$ such that*

$$n^{\frac{1}{2}}(\log n)^{-5} < u, v < n^{\frac{1}{2}}(\log n)^5. \tag{3.2}$$

*Then for $\varepsilon > 0$, $n > n_0(\varepsilon)$, we have*

$$|\mathscr{B}| < n(\log n)^{(e/2)\log 2 - 1 + \varepsilon}. \tag{3.3}$$

PROOF. Let $\mathscr{B}_1$ denote the set of the integers $b$ such that $b \leq n$ and $b$ can be represented in the form (3.1) with $u$ and $v$ such that

$$\min(\Omega(u), \Omega(v)) > \frac{e}{4}\log\log n \tag{3.4}$$

and let $B_2$ denote the set of the integers $b$ such that $b \leq n$ and $b$ can be represented in the form (3.1) with $u$ and $v$ such that (3.2) holds and

$$\Omega(v) \leq \frac{e}{4}\log\log n. \tag{3.5}$$

Then, clearly, $\mathscr{B} \subset \mathscr{B}_1 \cup \mathscr{B}_2$; whence

$$|\mathscr{B}| \leq |\mathscr{B}_1| + |\mathscr{B}_2|. \tag{3.6}$$

If $b \in \mathscr{B}_1$, then, by (3.1) and (3.4), we have

$$\Omega(b) = \Omega(uv) = \Omega(u) + \Omega(v) > \frac{e}{2}\log\log n. \tag{3.7}$$

By Lemma 12, it follows from $b \leq n$ and (3.7) that

$$|\mathscr{B}_1| < \frac{n}{(\log n)^{\phi(e/2)-\varepsilon}} = n(\log n)^{(e/2)\log 2 - 1 + e}. \tag{3.8}$$

Write

$$U = \{u : u \in \mathbb{N}, n^{\frac{1}{2}}(\log n)^{-5} < u < n^{\frac{1}{2}}(\log n)^5\},$$

and for $u \in U$ let

$$\mathscr{V}_n = \left\{v : v \in \mathbb{N}, v \leq n/u, \Omega(v) \leq \frac{e}{4}\log\log n\right\}$$

and

$$\mathscr{V}^* = \left\{v : v \in \mathbb{N}, v \leq n/u, \Omega(v) \leq \left(\frac{e}{4} + \frac{1}{\log\log n}\right)\log\log\frac{n}{u}\right\}.$$

Then, clearly, we have

$$|\mathscr{B}_2| \leq \sum_{u \in U} |\mathscr{V}_u|. \tag{3.9}$$

A simple computation shows that, for $u \in U$, we have

$$\log\log n - \log\log(n/u) = \log 2 + o(1)$$

and thus, for $n > n_0$ (and all $u \in U$), we have

$$\mathcal{V}_u \subseteq \mathcal{V}_u^*. \tag{3.10}$$

Moreover, by Lemma 11, for all $u \in U$ and $n > n_0(\varepsilon)$, we have

$$|\mathcal{V}_n^*| < \frac{n}{u}\left(\log\frac{n}{u}\right)^{-\phi(e/4)+\varepsilon/2}. \tag{3.11}$$

It follows from (3.9), (3.10) and (3.11) that, for $n > n_0(\varepsilon)$, we have

$$\begin{aligned}|\mathcal{B}_2| &\leq \sum_{u\in U}|\mathcal{V}_u| < \sum_{u\in U}\frac{n}{u}\left(\log\frac{n}{u}\right)^{-\phi(e/4)+\varepsilon/2} \\ &\ll n(\log n)^{-\phi(e/4)+\varepsilon/2}\sum_{u\in U}\frac{1}{u} \ll n(\log n)^{-\phi(e/4)+\varepsilon/2}\log\log n \\ &\ll n(\log n)^{-\phi(e/4)+\varepsilon} n(\log n)^{e/2\log 2-1+\varepsilon}.\end{aligned} \tag{3.12}$$

(3.3) follows from (3.6), (3.8) and (3.12), and this completes the proof of the lemma.

Lemma 14. *Assume that $a_1 < \cdots < a_k \leq n$ is a sequence of positive integers for which the products $a_i a_j$, with $i < j$, are all distinct. Then, for $n \geq 2$ and for some positive absolute constants $c_1$ and $c_2$, we have*

$$\pi(n) + c_1 n^{\frac{3}{4}}(\log n)^{-\frac{3}{2}} < \max k < \pi(n) + c_2 n^{\frac{3}{4}}(\log n)^{-\frac{3}{2}}.$$

Proof. This is a result of Erdős [5]. Note that in [5], the products $a_i a_j$ with $i = j$ are not excluded explicitly; however, the proof also gives the result in this slightly sharper form.

Lemma 15. *Every $n \in \mathbb{N}$ can be written in the form*

$$n = xy, \qquad x \geq y$$

*where either $x$ is a prime greater than $n^{\frac{2}{3}}$ or $x \leq n^{\frac{2}{3}}$.*

Proof. This lemma is due to Erdős [4]. For the sake of completeness, we sketch the proof. If $n = 1$, then $x = y = 1$ can be chosen. If $n > 1$, then let $p$ denote the greatest prime factor of $n$, and write $n = pn_1$. If $p \geq n_1$, then we may choose $x = p$ and $y = n_1$. If $p \leq n_1$ and $n_1 \leq n^{\frac{2}{3}}$, then $x = n_1$ and $y = p$ can be chosen. Finally, if $p < n_1$ and $n_1 > n^{\frac{2}{3}}$, then we have $p = n/n_1 < n^{1/3}$. Let $n_1 = p_1 p_2 \cdots p_k$, where $p_1, p_2, \ldots, p_k$ are primes and $n^{\frac{1}{3}} > p \geq p_1 \geq p_2 \geq \cdots \geq p_k$. Define $i$ by $p_1 p_2 \cdots p_i \leq n^{\frac{2}{3}} < p_1 p_2 \cdots p_{i+1}$. Then

$$x = \max\left(p_1 p_2 \cdots p_i, \frac{n}{p_1 p_2 \cdots p_i}\right), \qquad y = \min\left(p_1 p_2 \cdots p_i, \frac{n}{p_1, p_2, \cdots p_i}\right)$$

can be chosen.

Lemma 16. *For every positive integer $k \geq 2$, we have*

$$F_{k+4}(n) \leq \max(F_k(n) + 4, F_4(n)). \tag{3.13}$$

Proof. Assume that

$$|\mathcal{A}| > F_4(n) \tag{3.14}$$

and

$$|\mathcal{A}| > F_k(n) + 4. \tag{3.15}$$

Then, by (3.14), there are integers $a_1, a_2, a_3, a_4$ and $x$ such that

$$a_1a_2a_3a_4 = x^2, \qquad a_1, a_2, a_3, a_4 \in \mathscr{A}, \quad a_1 < a_2 < a_3 < a_4.$$

Write $\mathscr{A}^* = \mathscr{A}\setminus\{a_1, a_2, a_3, a_4\}$ so that, by (3.15),

$$|\mathscr{A}^*| = |\mathscr{A}| - 4 > F_k(n).$$

Thus, by the definition of $F_k(n)$, there are integers $a_5, \ldots, a_{k+4}$ and $y$ such that

$$a_5 \cdots a_{k+4} = y^2, \qquad a_5, \ldots, a_{k+4} \in \mathscr{A}^* \subset \mathscr{A}, \quad a_5 < \cdots < a_{k+4}.$$

Then we have

$$a_1 \cdots a_{k+4} = (xy)^2, \qquad a_1, \ldots, a_{k+4} \in \mathscr{A}, \quad a_i \neq a_j \quad \text{for } 1 \leq i < j \leq k+4,$$

so that

$$\mathscr{A} \notin \Gamma_{k+4}$$

for all $\mathscr{A}$ satisfying (3.14) and (3.15), which implies (3.13).

LEMMA 17. *For $n \to +\infty$, we have*

$$\sum_{i \leq n} 2^{\omega(i)} = \left(\frac{6}{\pi^2} + o(1)\right) n \log n.$$

PROOF. Clearly, we have

$$\sum_{i \leq n} 2^{\omega(i)} = \sum_{i \leq n} \sum_{\substack{d \mid i \\ |\mu(d)|=1}} 1 = \sum_{\substack{d \leq n \\ |\mu(d)|=1}} \sum_{\substack{i \leq n \\ d \mid i}} 1 = \sum_{\substack{d \leq n \\ |\mu(d)|=1}} \left[\frac{n}{d}\right]$$

$$= n \sum_{\substack{d \leq n \\ |\mu(d)|=1}} \frac{1}{d} + O(n) = n \sum_{d \leq n} \left(\sum_{m^2 \mid d} \mu(m)\right) \frac{1}{d} + O(n)$$

$$= n \sum_{m^2 \leq n} \frac{\mu(m)}{m^2} \sum_{i \leq n/m^2} \frac{1}{i} + O(n)$$

$$= n\left(\sum_{m=1}^{+\infty} \frac{\mu(m)}{m^2} + o(1)\right) \log n + O(n)$$

$$= n \log n \prod_p \left(1 - \frac{1}{p^2}\right) + o(n \log n)$$

$$= \left(\frac{1}{\xi(2)} + o(1)\right) n \log n = \left(\frac{6}{\pi^2} + o(1)\right) n \log n.$$

## 4. PROOFS OF THE THEOREMS

PROOF OF THEOREM 1. The set $\mathscr{A}$ of the square-free integers not exceeding $n$ has property $P_2$, *whence it follows that*

$$F_2(n) \geq |\mathscr{A}| = \sum_{i \leq n} \mu^2(i). \tag{4.1}$$

Now assume that $\mathscr{B} \subset \{1, 2, \ldots, n\}$ and

$$|\mathscr{B}| > \sum_{i \leq n} \mu^2(i).$$

Then there are distinct integers $b_1 \in \mathscr{B}$ and $b_2 \in \mathscr{B}$ the square-free parts of which are the same:

$$b_1 = r^2 t, \qquad b_2 = s^2 t, \qquad |\mu(t)| = 1.$$

Then we have

$$b_1 b_2 = (rst)^2,$$

so that $\mathscr{B}$ does not have property $P_2$. It follows that

$$F_2(n) \leqslant \sum_{i \leqslant n} \mu^2(i). \tag{4.2}$$

(1.3) follows from (4.1) and (4.2), and this completes the proof of Theorem 1.

Proof of Theorem 2. First we will prove the lower bound in (1.4).

Let $\mathscr{A}_n$ denote the set of the integers $a$ such that:

(i) $n(\log n)^{-1} < a \leqslant n$;

(ii) there is no positive integer $b$ such that $b > \log n$ and $b^2 \mid a$;

(iii) $a$ cannot be represented in the form

$$a = uv$$

with integers $u$ and $v$ such that

$$n^{\frac{1}{2}}(\log n)^{-5} < u, v < n^{\frac{1}{2}}(\log n)^5.$$

First, we will give a lower bound for $|\mathscr{A}_n|$. For $n \geqslant 3$, the number of the integers $a$ satisfying (i) and (ii) is

$$\begin{aligned}
&\geqslant |\{a : a \in \mathbb{N}, n(\log n)^{-1} < a \leqslant n\}| \\
&\quad - \sum_{\log n < b \leqslant n} |\{a : a \in \mathbb{N}, b^2 \mid a, n(\log n)^{-1} < a \leqslant n\}| \\
&\geqslant n - n(\log n)^{-1} - \sum_{\log n < b} \frac{n}{b^2} \\
&> n - n(\log n)^{-1} - n \sum_{\log n < b} \frac{1}{(b-1)b} = n - n(\log n)^{-1} - n[\log n]^{-1} \\
&> n - 3n(\log n)^{-1}.
\end{aligned}$$

By Lemma 13, for $n > n_0(\varepsilon)$ all but $n(\log n)^{(e/2)\log 2 - 1 + \varepsilon/2}$ of these integers $a$ also satisfy (iii). Thus, for large $n$, we have

$$|\mathscr{A}_n| > n - 3n(\log n)^{-1} - n(\log n)^{(e/2)\log 2 - 1 + \varepsilon/2} > n - n(\log n)^{(e/2)\log 2 - 1 + \varepsilon}. \tag{4.3}$$

Next we will show that $\mathscr{A} \in \Gamma_3$. Assume to the contrary that there are $a_1, a_2, a_3 \in \mathscr{A}$, $x \in \mathbb{N}$ such that $a_1 < a_2 < a_3$ and

$$a_1 a_2 a_3 = x^2. \tag{4.4}$$

Write $a_1$, $a_2$ and $a_3$ as the product of a square and a square-free number:

$$a_1 = b_1^2 q_1, \quad a_2 = b_2^2 q_2, \quad a_3 = b_3^2 q_3 \qquad (q_1, q_2 \text{ and } q_3 \text{ are square-free}). \tag{4.5}$$

By (ii), here we have

$$b_1, b_2, b_3 \leqslant \log n. \tag{4.6}$$

By (4.4) and (4.5) we have

$$a_1 a_2 a_3 = (b_1 b_2 b_3)^2 q_1 q_2 q_3 = x^2. \tag{4.7}$$

It follows that $q_1q_2q_3$ is a square:

$$q_1q_2q_3 = y^2. \tag{4.8}$$

Since $q_1$, $q_2$ and $q_3$ are square-free, this implies that any prime factor of $q_1q_2q_3$ divides exactly two of the numbers $q_1$, $q_2$ and $q_3$. Thus, writing $(q_1, q_2) = d_3$, $(q_1, q_3) = d_2$ and $(q_2, q_3) = d_1$, we have

$$q_1 = d_2d_3, \qquad q_2 = d_1d_3 \qquad \text{and} \qquad q_3 = d_1d_2. \tag{4.9}$$

It follows from (i), (4.5) and (4.6) that

$$\begin{aligned} d_1 &= \frac{d_1d_2d_3}{d_2d_3} = \frac{(q_1q_2q_3)^{\frac{1}{2}}}{q_1} \\ &= \frac{(a_1a_2a_3(b_1b_2b_3)^{-2})^{\frac{1}{2}}}{a_1b_1^{-2}} \geq \frac{(n^3(\log n)^{-3}(\log n)^{-6})^{\frac{1}{2}}}{n \cdot 1} = n^{\frac{1}{2}}(\log n)^{-\frac{9}{2}} \end{aligned}$$

and

$$\begin{aligned} d_1 &= \frac{(a_1a_2a_3(b_1b_2b_3)^{-2})^{\frac{1}{2}}}{a_1b_1^{-2}} \leq \frac{(n^3 \cdot 1)^{\frac{1}{2}}}{n(\log n)^{-1}(\log n)^{-2}} \\ &= n^{\frac{1}{2}}(\log n)^3, \end{aligned}$$

and, in the same way, we have

$$n^{\frac{1}{2}}(\log n)^{\frac{9}{2}} \leq d_2, d_3 \leq n^{\frac{1}{2}}(\log n)^3. \tag{4.10}$$

Write $u_1 = b_1d_2$ and $v_1 = b_1d_3$. Then, by (4.5) and (4.9), we have

$$u_1v_1 = (b_1d_2)(b_1d_3) = b_1^2q_1 = a_1. \tag{4.11}$$

Moreover, by (4.6) and (4.10), we have

$$u_1 = b_1d_2 \geq d_2 \geq n^{\frac{1}{2}}(\log n)^{-\frac{9}{2}} \tag{4.12}$$

and

$$u_1 = b_1d_2 \leq (\log n)n^{\frac{1}{2}}(\log n)^3 = n^{\frac{1}{2}}(\log n)^4, \tag{4.13}$$

and, in the same way,

$$n^{\frac{1}{2}}(\log n)^{-\frac{9}{2}} \leq v_1 \leq n^{\frac{1}{2}}(\log n)^4. \tag{4.14}$$

By (iii) in the definition of $\mathscr{A}_n$, it follows from (4.11), (4.12), (4.13) and (4.14) that $a_1 \in \mathscr{A}_n$ cannot hold, and this contradiction proves that indeed we have

$$\mathscr{A}_n \in \Gamma_3. \tag{4.15}$$

It follows from (4.3) and (4.15) that

$$F_3(n) > n - (\log n)^{(e/2)\log 2 - 1 + \varepsilon} \qquad (\text{for } n > n_0(\varepsilon)).$$

Now we will prove the upper bound in (1.4). Assume that $\mathscr{A} \subset \{1, 2, \ldots, n\}$, $\mathscr{A} \in \Gamma_3$. Let $\mathscr{D} = \{d_1, d_2, \ldots, d_t\}$ denote the set of the integers $d$ such that $d \leq n^{\frac{1}{2}}$ and $\Omega(d) \leq \frac{1}{2}\log\log n$. Then, by Lemma 11, for large $n$ we have

$$t = |\mathscr{D}| > n^{\frac{1}{2}}(\log n)^{-\phi(\frac{1}{2}) - (\varepsilon/3)}. \tag{4.16}$$

Define the graph $G$ on the $t$ vertices $P_1, P_2, \ldots, P_t$ so that $P_i$ and $P_j$ $(i \neq j)$ are joined iff $d_id_j \in \mathscr{A}$. Let $\mathscr{M}$ denote the set of the integers $m$ such that $m \leq n$ and $m$ has a representation in the form

$$m = d_id_j \qquad \text{with} \qquad 1 \leq i < j \leq t \quad d_id_j \notin \mathscr{A}. \tag{4.17}$$

Then $m \in M$ implies that $m \notin \mathscr{A}$, so that

$$|\mathscr{A}| \leq n - |\mathscr{M}|. \tag{4.18}$$

For fixed $m$, let $h(m)$ denote the number of pairs $d_i$, $d_j$ satisfying (4.17), and write

$$H = \max_{m \in M} h(m).$$

If $m = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r} \in \mathscr{M}$, then by (4.17) and the definition of $\mathscr{D}$ we have

$$\Omega(m) = \Omega(d_i) + \Omega(d_j) \leq \log\log n$$

so tha, clearly,

$$h(m) \leq \tau(m) = \prod_{i=1}^{r} (k_i + 1) \leq \prod_{i=1}^{r} 2^{k_i} = 2^{\Omega(m)} \leq (\log n)^{\log 2}$$

and thus

$$H \leq (\log n)^{\log 2}. \tag{4.19}$$

By Lemma 9, it follows from $\mathscr{A} \in \Gamma_3$ that the graph $G$ does not contain a triangle. Thus, by (2.3) in Lemma 1, for $n > n_0(\varepsilon)$ there are at least $(1 - \varepsilon)(t^2/4)$ pairs $i, j$ (with $1 \leq i < j \leq t$) such that $P_i$ and $P_j$ are not joined in $G$. For all these pairs $i, j$, the number $m$ defined by (4.17) belongs to $M$. Each of these numbers $m$ can be represented at most $H$ times in form (4.17), so that by (4.16) and (4.19) for large $n$ we have

$$|\mathscr{M}| \geq (1 - \varepsilon)\frac{t^2}{4} H^{-1} > n(\log n)^{-2\phi(\frac{1}{2}) - \log 2 - \varepsilon} = n(\log n)^{-1-\varepsilon}$$

and, by (4.18), this proves the upper bound in (1.4).

Proof of Theorem 3. First we will prove the lower bound in (1.5). Let $\mathscr{P} = \{p_1, p_2, \ldots, p_t\}$ denote the set of the primes not exceeding $n^{\frac{1}{2}}$ so that, by the prime number theorem,

$$t = \pi(n^{\frac{1}{2}}) \sim 2\frac{n^{\frac{1}{2}}}{\log n}. \tag{4.20}$$

By (2.1) in Lemma 1, for $\varepsilon > 0$, $n > n_0(\varepsilon)$ there is a graph $G_t$ on $t$ vertices $P_1, \ldots, P_t$ such that

$$e(G_t) > \left(\frac{1}{2} - \frac{\varepsilon}{5}\right) t^{\frac{3}{2}} \tag{4.21}$$

and it does not contain a $C_4$. Let $\mathscr{B}$ denote the set of all the integers of the form $p_i p_j$, where $p_i, p_j \in \mathscr{P}$, $1 \leq i < j \leq t$, and $P_i$ and $P_j$ are joined in $G_t$. Then, by Lemma 9(ii), we have

$$\mathscr{B} \in \Gamma_4. \tag{4.22}$$

Let $\mathscr{Q}$ denote the set of primes with $n^{\frac{1}{2}} < p \leq n$, so that

$$|\mathscr{Q}| = \pi(n) - \pi(n^{\frac{1}{2}}) = \pi(n) - t, \tag{4.23}$$

and write

$$\mathscr{A} = \mathscr{Q} \cup \mathscr{B}. \tag{4.24}$$

Then,

$$|\mathscr{A}| = |\mathscr{Q}| + |\mathscr{B}| = (\pi(n) - t) + e(G_t) > \pi(n) + \left(\frac{1}{2} - \frac{\varepsilon}{4}\right) t^{\frac{3}{2}} > \pi(n) + (2^{\frac{1}{2}} - \varepsilon) n^{\frac{3}{4}} (\log n)^{-\frac{3}{2}}. \tag{4.25}$$

Now assume that $a_1, a_2, a_3, a_4 \in \mathscr{A}$ and $a_i \neq a_j$ for $i \neq j$. If one of $a_1, a_2, a_3$ and $a_4$, say $a_1$, belongs to $Q$, then $a_1$ is a prime greater, than $n^{\frac{1}{2}}$, and none of $a_2, a_3$ and $a_4$ is divisible by this prime, so that the product $a_1a_2a_3a_4$ cannot be a square. If none of $a_1$, $a_2$, $a_3$ and $a_4$ belongs to $Q$, then each of them belongs to $\mathscr{B}$; thus, by (4.22), again their product cannot be a square. Thus we have

$$\mathscr{A} \in \Gamma_4. \tag{4.26}$$

The lower bound in (1.5) follows from (4.25) and (4.26).

To prove the upper bound in (1.5), assume that $\mathscr{A} \subset \{1, 2, \dots, n\}$ and

$$|\mathscr{A}| > \pi(n) + c_2 n^{\frac{3}{4}}(\log n)^{-\frac{3}{2}},$$

where $c_2$ is the constant defined in Lemma 14. Then, by Lemma 14, there are four distinct integers $a_1, a_2, a_3, a_4 \in \mathscr{A}$ such that

$$a_1a_2 = a_3a_4,$$

so that their product

$$a_1a_2a_3a_4 = (a_1a_2)^2$$

and thus $\mathscr{A} \notin \Gamma_4$. This implies the upper bound in (1.5).

Proof of Theorem 4. First we will prove the lower bounds in (1.6) in Theorem 4 and in (1.7) in Theorem 6 simultaneously. Let

$$\mathscr{D} = \{p : n^{\frac{1}{2}} < p \leq n, p \text{ prime}\} \cup \{2p : n^{\frac{1}{2}} < p \leq n/2, p \text{ prime}\}.$$

Then, by the prime number theorem, for $n > n_0(\varepsilon)$ we have

$$|\mathscr{D}| = \pi(n) + \pi(n/2) - O(n^{\frac{1}{2}}(\log n)^{-1}). \tag{4.27}$$

Let $\mathscr{P} = \{p_1, p_2, \dots, p_t\}$ denote the set of the primes $p$ with $2 < p \leq n^{\frac{1}{2}}$, so that by the prime number theorem $t \sim 2n^{\frac{1}{2}}(\log n)^{-1}$. Let $G_t$ denote a graph on the $t$ vertices $P_1, P_2, \dots, P_t$ with the maximal number of edges, so that it does not contain a cycle of length $l$ with $3 \leq l \leq 4k+2$. Then let $\mathscr{E}$ denote the set the elements of which are of the form $p_ip_j$ $(1 \leq i < j \leq t)$, and $p_ip_j \in \mathscr{E}$ (where $1 \leq i < j \leq t$) iff the vertices $P_i$ and $P_j$ are joined in $G_t$. By Lemmas 2 and 3 we have

$$|\mathscr{E}| = e(G(t,k)) > \begin{cases} \left(\frac{1}{2} - \frac{\varepsilon}{8}\right)t^{\frac{4}{3}} > \left(2^{\frac{1}{3}} - \frac{\varepsilon}{2}\right)n^{\frac{2}{3}}(\log n)^{-\frac{4}{3}} & \text{for } k = 1, \quad n > n_0(\varepsilon), \\ c_{4k+2}t^{1+(4k+1)^{-1}} > c_k'(n^{\frac{1}{2}}(\log n)^{-1})^{1+(4k+1)^{-1}} & \text{for } k > 1. \end{cases} \tag{4.28}$$

Thus, writing $\mathscr{A} = \mathscr{D} \cup \mathscr{E}$, it follows from (4.27) and (4.28) that

$$\begin{aligned} |\mathscr{A}| = |\mathscr{D}| + |\mathscr{E}| &> \pi(n) + \pi(n/2) \\ &+ \begin{cases} (2^{\frac{1}{3}} - \varepsilon)n^{\frac{2}{3}}(\log n)^{-\frac{4}{3}} & \text{for } k = 1, \quad n > n_0(\varepsilon), \\ c_k''(n^{\frac{1}{2}}(\log n)^{-1})^{1+(4k+1)^{-1}} & \text{for } k > 1. \end{cases} \end{aligned} \tag{4.29}$$

Now we will prove that

$$\mathscr{A} \in \Gamma_{4k+2}. \tag{4.30}$$

Assume that, contrary to (4.30), we have

$$a_1 \cdots a_{4k+2} = x^2, \qquad a_1, \dots, a_{4k+2} \in \mathscr{A}, \quad a_1 < \cdots < a_{4k+2}. \tag{4.31}$$

Assume that $q_1$ is a prime, with $q_1 > n^{\frac{1}{2}}$ and $q_1 \mid x^2$. This implies that $q_1^2 \mid x^2$. By the construction of the set $\mathscr{A}$, it follows that one of the numbers $a_1, \dots, a_{4k+2}$ is equal to

$q_1$, and another one is equal to $2q_1$. In this way, we obtain that the left-hand side of (4.31) is of the form

$$a_1 \cdots a_{4k+2} = \left(\prod_{i=1}^{l} q_i \cdot 2q_i\right) \prod_{j=1}^{4k+2-2l} e_j = 2^l \left(\prod_{i=1}^{l} q_i\right)^2 \prod_{j=1}^{4k+2-2l} e_j, \tag{4.32}$$

where the $q_i$'s are distinct primes greater than $n^{\frac{1}{2}}$, the $e_j$'s are distinct elements of $\mathscr{E}$ so that if $p$ is a prime with $p \mid \prod_{j=1}^{4k+2-2l} e_j$, then $2 < p \leqslant n^{\frac{1}{2}}$, and it may occur that $l = 0$ or $l = 2k+1$ (so that there are no $q_i$'s or $e_j$'s). Thus it follows from (4.31) and (4.32) that both $2^l$ and $\prod_{j=1}^{4k+2-2l} e_j$ are squares:

$$2^l = y^2, \qquad \prod_{j=1}^{4k+2-2l} e_j = z^2. \tag{4.33, 4.34}$$

By Lemma 10(iii), it follows from the construction of the set $\mathscr{E}$ that (4.34) cannot hold unless the product on the left-hand side is empty, i.e. $l = 2k+1$. But then, by (4.33), we have

$$2^{2k+1} = y^2$$

and this is impossible. This contradiction completes the proof of (4.30).

The lower bound in both (1.6) and (1.7) follows from (4.29) and (4.30).

To prove the upper bound in (1.6), we have to show that, assuming $\mathscr{A} \subset \{1, 2, \ldots, n\}$ and

$$\mathscr{A} \in \Gamma_6, \tag{4.35}$$

we have

$$|\mathscr{A}| < \pi(n) + \pi(n/2) + n^{\frac{7}{9}} \log n. \tag{4.36}$$

Let $\mathscr{D}$ denote the set of the numbers $a \in \mathscr{A}$ that are of the form

$$a = py, \qquad p > n^{\frac{2}{3}} \tag{4.37}$$

and let

$$\mathscr{E} = \mathscr{A} \backslash \mathscr{D}$$

so that we have

$$|\mathscr{A}| \leqslant |\mathscr{D}| + |\mathscr{E}|. \tag{4.38}$$

Define the integer $k_0$ by

$$n 2^{-(k_0+1)} < n^{\frac{2}{3}} \leqslant n 2^{-k_0},$$

so that

$$k_0 = \left[\frac{1}{3} \frac{\log n}{\log 2}\right].$$

For $k = 0, 1, 2, \ldots, k_0$, let $\mathscr{D}_k$ denote the set of the numbers $a \in \mathscr{D}$ for which in (4.37) we have

$$n/2^{k+1} < p \leqslant n/2^k, \tag{4.39}$$

so that

$$\mathscr{D} = \bigcup_{k=0}^{k_0} \mathscr{D}_k. \tag{4.40}$$

It follows from $a \leqslant n$, (4.37) and (4.39) that

$$y < 2^{k+1}. \tag{4.41}$$

Thus, for $k = 0$ we have $y = 1$, so that $a = p$, whence

$$|\mathscr{D}_0| \leqslant |\{p : n/2 < p \leqslant n, p \text{ prime}\}| = \pi(n) - \pi(n/2). \tag{4.42}$$

Now assume that $1 \leq k \leq k_0$. Denote the primes $p$ satisfying (4.39) by $p_1, p_2, \ldots, p_{u_k}$, so that

$$u_k = \pi(n/2^k) - \pi(n/2^{k+1})$$

and, by the prime number theorem,

$$2^{k+1}u_k \leq 2^{k+1}\pi(n/2^k) < 2^{k+1} \cdot 2\frac{n}{2^k \log n^{\frac{1}{2}}} = 8\frac{n}{\log n}. \tag{4.43}$$

Define the bipartite graph $G[k] = G(U, V)$ on the vertices $U = \{P_1, P_2, \ldots, P_{u_k}\}$, $V = \{Q_1, Q_2, \ldots, Q_{2^{k+1}}\}$ so that the vertices $P_i$ and $Q_j$ are joined iff $p_i p_j \in \mathscr{D}_k$. By Lemma 8, it follows from (4.35) and $\mathscr{D} \subset \mathscr{A}$ that $G[k]$ cannot contain a $C_6$. Thus we have

$$|\mathscr{D}_k| = e(G[k]) < s(u_k, 2^{k+1}, 2^{k+1}u_k) \overset{\text{def}}{=} S_k \tag{4.44}$$

(where $s(u, v, n)$ is the function defined in Section 2).

Let

$$\mathscr{K}_1 = \{k: 1 \leq k \leq k_0,\ 2^{3k} \leq u_k\},$$
$$\mathscr{K}_2 = \{k: 1 \leq k \leq k_0,\ 2^{2(k+1)} < u_k < 2^{3k}\},$$
$$\mathscr{K}_3 = \{k: 1 \leq k \leq k_0,\ u_k \leq 2^{2(k+1)}\}.$$

Then, by (4.40), (4.42) and (4.44), and by using Lemmas 4 and 5, we have

$$\begin{aligned}|\mathscr{D}| &= |\mathscr{D}_0| + \sum_{k=1}^{k_0} |\mathscr{D}_k| < (\pi(n) - \pi(n/2)) + \sum_{k=1}^{k_0} S_k \\ &= (\pi(n) - \pi(n/2)) + \sum_{k \in \mathscr{K}_1} S_k + \sum_{k \in \mathscr{K}_2} S_k + \sum_{k \in \mathscr{K}_3} S_k \\ &< (\pi(n) - \pi(n/2)) + 2\sum_{k=1}^{k_0} u_k + \sum_{k \in \mathscr{K}_1} 2^{3(k+1)} + 18 \sum_{k \in \mathscr{K}_2} 2^{k+1}u_k^{\frac{2}{3}} \\ &\quad + \sum_{k \in \mathscr{K}_3} c(2^{k+1}u_k)^{\frac{7}{9}}.\end{aligned} \tag{4.45}$$

Here we have

$$\begin{aligned}&(\pi(n) - \pi(n/2)) + 2\sum_{k=1}^{k_0} u_k \\ &\quad = (\pi(n) - \pi(n/2)) + 2\sum_{k=1}^{k_0} (\pi(n/2^k) - \pi(n/2^{k+1})) < \pi(n) + \pi(n/2).\end{aligned} \tag{4.46}$$

Moreover, in view of (4.43), we obtain by a simple computation that, writing $\max_{k \in \mathscr{K}_1} k = K_1$ and $\max_{k \in \mathscr{K}_2} k = K_2$, we have

$$\sum_{k \in \mathscr{K}_1} 2^{3(k+1)} \ll 2^{3K_1} \ll n^{\frac{3}{4}}(\log n)^{-\frac{3}{4}}, \tag{4.47}$$

$$\begin{aligned}\sum_{k \in \mathscr{K}_2} 2^{k+1}u_k^{\frac{2}{3}} &\ll \sum_{k \in \mathscr{K}_2} 2^{k/3}(2^k u_k)^{\frac{2}{3}} \\ &\ll n^{\frac{2}{3}}(\log n)^{-\frac{2}{3}} 2^{K_2/3} \ll n^{\frac{2}{3}}(\log n)^{-\frac{2}{3}} n^{\frac{1}{9}}(\log n)^{-\frac{1}{9}} \\ &= n^{\frac{7}{9}}(\log n)^{-\frac{7}{9}}\end{aligned} \tag{4.48}$$

and

$$\sum_{k \in \mathscr{K}_3} (2^{k+1}u_k)^{\frac{7}{9}} \ll n^{\frac{7}{9}}(\log n)^{-\frac{7}{9}} |\mathscr{K}_3| \ll n^{\frac{7}{9}}(\log n)^{-\frac{7}{9}} \log\log n. \tag{4.49}$$

It follows from (4.45), (4.46), (4.47), (4.48) and (4.49) that

$$|\mathscr{D}|<\pi(n)+\pi(n/2)+n^{7/9}(\log n)^{-7/9}\log\log n. \tag{4.50}$$

It remains to estimate $|\mathscr{E}|$. If $a\in\mathscr{E}$, then $a\in\mathscr{A}$, $a\notin\mathscr{D}$, so that, by Lemma 15, $a$ can be written in the form

$$a=xy,\qquad x\leqslant n^{2/3},\quad y\leqslant x. \tag{4.51}$$

To each $a\in\mathscr{A}$, assign a unique pair $x=x(a)$ and $y=y(a)$ satisfying (4.51) (e.g. consider that pair $x, y$ for which (4.51) holds and $x$ is maximal). Define the integer $k_1$ by

$$n^{2/3}2^{-(k_1+1)}<n^{1/2}\leqslant n^{2/3}2^{-k_1} \tag{4.52}$$

so that $k_1=[(\log n)/(6\log 2)]$. For $k=0,1,\ldots,k_1-1$, let $\mathscr{E}_k$ denote the set of the numbers $a\in\mathscr{E}$ such that

$$n^{2/3}2^{-(k+1)}<x(a)\leqslant n^{2/3}2^{-k}, \tag{4.53}$$

and let $\mathscr{E}_{k_1}$ denote the set of the numbers $a\in\mathscr{E}$ such that

$$x(a)\leqslant n^{2/3}2^{-k_1}.$$

Clearly, we have

$$\mathscr{E}=\bigcup_{k=1}^{k_1}\mathscr{E}_k.$$

It follows from (4.51) and (4.53) that for $a\in\mathscr{O}_k$, we have

$$y(a)\leqslant\frac{n}{x(a)}<2^{k+1}n^{1/3}$$

and, in view of (4.51) and (4.53), for $a\in\mathscr{E}_{k_1}$ we have

$$y(a)\leqslant x(a)\leqslant n^{2/3}2^{-k_1}=(n^{1/6}2^{-(k_1+1)})^2\cdot 2^{k_1+2}n^{1/3}<2^{k_1+2}n^{1/3}.$$

Now assume that $1\leqslant k\leqslant k_1$. Write $u_k=[\,n^{2/3}2^{-k}]$ and $v_k=[2^{k+2}n^{1/3}]$ so that

$$u_k\leqslant v_k^2,\qquad v_k\leqslant 5u_k\qquad\text{and}\qquad u_kv_k\leqslant 4n. \tag{4.54}$$

Define the bipartite graph $G[k]$ on the vertices $U=\{P_1,P_2,\ldots,P_{u_k}\}$; $V=\{Q_1,Q_2,\ldots,Q_{v_k}\}$ so that the vertices $P_i$ and $Q_j$ are joined iff there is an $a\in\mathscr{E}_k$ such that $x(a)=i$ and $y(a)=j$. By Lemma 8, it follows from (4.35) and $\mathscr{E}_k\subset\mathscr{A}$ that $G[k]$ cannot contain a $C_6$. Thus, by Lemma 4, we have

$$|\mathscr{E}_k|=e(G[k])<r(4n)<cn^{7/9}$$

for some positive constant $c$ (where $r(n)$ is the function defined in Section 2). It follows that

$$|\mathscr{E}|=\sum_{k=1}^{k_1}|\mathscr{E}_k|\ll k_1n^{7/9}\ll n^{7/9}\log n. \tag{4.55}$$

(4.36) follows from (4.38), (4.50) and (4.55), and this completes the proof of Theorem 4.

Proof of Theorem 5. The lower bound can be proved in the same way as the lower bound in Theorem 3, except that (2.1) in Lemma 1 has to be replaced by Lemma 3 and Lemma 10(iii). The upper bound follows from Lemma 16 and the upper bound in Theorem 3.

Proof of Theorem 6. The lower bound was proved in the proof of Theorem 4. The upper bound follows from Lemma 16 and the upper bound in Theorem 4.

Proof of Theorem 7. Let $\mathscr{A}$ denote the set of the integers $a$ such that $a \leq n$ and $a$ has a prime factor greater than $n^{\frac{1}{2}}$. Then, by

$$\sum_{p<x} \frac{1}{p} = \log\log x + c + o(1),$$

we have

$$|\mathscr{A}| = \sum_{n^{\frac{1}{2}}<p\leq n} \left[\frac{n}{p}\right] = n \log 2 + o(n),$$

so that

$$|\mathscr{A}| > (\log 2 - \varepsilon)n \qquad \text{for } n > n_0(\varepsilon). \tag{4.56}$$

Moreover, if $a_1, \ldots, a_{2k+1} \in \mathscr{A}$, then each of the $a_i$'s has exactly one prime factor greater than $n^{\frac{1}{2}}$. It follows that there is a prime $p > n^{\frac{1}{2}}$ such that defining $r = r(p)$ by $p^r \mid a_1 \cdots a_{2k+1}$, $p^{r+1} \nmid a_1 \cdots a_{2k+1}$, $r$ is odd. Thus $a_1 \cdots a_{2k+1}$ cannot be a square, so that

$$\mathscr{A} \in \Gamma_{2k+1}. \tag{4.57}$$

The lower bound in (1.8) follows from (4.56) and (4.57).

To prove the upper bound, assume that $\mathscr{A} \subset \{1, 2, \ldots, n\}$, $\mathscr{A} \in \Gamma_{2k+1}$. Write $t = \pi(n^{\frac{1}{2}}) = (2 + o(1))n^{\frac{1}{2}}(\log n)^{-1}$. Let $\mathscr{B}$ denote the set of the integers $b$ such that $b \in \mathscr{A}$ and $b$ can be represented in the form $b = p_i p_j$ with $1 \leq i < j \leq t$. Define the graph $G[B]$ on the $t$ vertices $P_1, P_2, \ldots, P_t$ so that $P_i$ and $P_j$ $(i \neq j)$ are joined iff $p_i p_j \in \mathscr{B}$. Since $\mathscr{B} \subset \mathscr{A} \in \Gamma_{2k+1}$, the graph $G[\mathscr{B}]$ cannot contain a $C_{2k+1}$. Thus, by Lemma 1(c),

$$e(G[\mathscr{B}]) \leq \left(\frac{1}{4} + \frac{\varepsilon}{3}\right)t^2.$$

Thus the number of pairs of vertices not joined in $G[\mathscr{B}]$, i.e. the number of integers $p_i p_j$ $(\leq n)$ missing from $\mathscr{A}$, in at least

$$\binom{t}{2} - \left(\frac{1}{4} + \frac{\varepsilon}{6}\right)t^2 > (1 - \varepsilon)n(\log n)^{-2},$$

which completes the proof of the theorem.

Remark. The constant factor $\log 2$ in (1.8) could be improved slightly. In fact, if $u$ is a fixed real number with $0 < u < 1$, then let $\mathscr{A}_u$ denote the set of integers $a$ such that $a \leq n$ and the number of the primes $p$ with $n^u < p \leq n$, $p \mid a$ is odd. Let $c(u)$ denote the greatest positive number such that for all $\varepsilon > 0$, $n > n_0(\varepsilon)$ we have

$$|\mathscr{A}_u| > (c(u) - \varepsilon)n.$$

Then, clearly, $\mathscr{A}_u \in \Gamma_{2k+1}$, so that $(c(u) - \varepsilon)n < F_{2k+1}(n)$ for all $0 < u < 1$.

It could be shown that there is a number $0 < u_0 < 1/2$ such that $c(u)$ is increasing on the left of $u_0$ and it is decreasing in $[u_0, 1/2]$. Then the best lower bound obtained in this way is

$$(c(u_0) - \varepsilon)n < F_{k+2}(n).$$

However, it would need a lengthy computation to compute or just to estimate these numbers $u_0$ and $c(u_0)$.

PROOF OF THEOREM 8. (1.9) follows from $\{p: p \text{ prime}, p \leqslant n\} \in \Gamma_{4k}$ and the upper bound in Theorem 5.

(1.10) follows from $(\{p: p \text{ prime}, 2 < p \leqslant n\} \cup \{2p: p \text{ prime}, 4 < 2p \leqslant n\}) \in \Gamma_{4k+2}$ and the upper bound in Theorem 6.

To give a lower bound for $L_{2k+1}(n)$, consider the set $\mathscr{A} = \{a: a \leqslant n, \Omega(a) \text{ is odd}\}$. Then, clearly, $\mathscr{A} \in \Gamma_{2k+1}$. Moreover, it is well known that the prime number theorem implies

$$\sum_{n \leqslant x} \lambda(n) = o(x),$$

where $\lambda(n) = (-1)^{\Omega(n)}$ is the Liouville function. It follows by partial summation that for $\varepsilon > 0$, $n > n_0(\varepsilon)$ we have

$$\sum_{a \in \mathscr{A}} \frac{1}{a} > (\tfrac{1}{2} - \varepsilon) \log n,$$

whence

$$L_{2k+1}(n) > (\tfrac{1}{2} - \varepsilon) \log n \qquad (\text{for } n > n_0(\varepsilon)).$$

It remains to show that for $\varepsilon > 0$, $n > n_1(\varepsilon, k)$ we have

$$L_{2k+1}(n) < (\tfrac{1}{2} + \varepsilon) \log n \qquad (\text{for } n > n_1(\varepsilon, k)).$$

In other words, we have to show that if $\varepsilon > 0$, $n > n_1(\varepsilon, k)$, $\mathscr{A} \subset \{1, 2, \ldots, n\}$ and

$$\sum_{a \in \mathscr{A}} \frac{1}{a} \geqslant (\tfrac{1}{2} + \varepsilon) \log n, \tag{4.58}$$

then

$$\mathscr{A} \notin \Gamma_{2k+1}. \tag{4.59}$$

To show this, write every $a \in \mathscr{A}$ as the product of a square and a square-free number:

$$a = (u(a))^2 v(a), \qquad \text{where } |\mu(v(a))| = 1.$$

Then we have

$$\sum_{a \in \mathscr{A}} \frac{1}{a} = \sum_{a \in \mathscr{A}} \frac{1}{(u(a))^2 v(a)} = \sum_{u^2 \leqslant n} \frac{1}{u^2} \sum_{\substack{a \in \mathscr{A} \\ u(a) = u}} \frac{1}{v(a)}$$

$$\leqslant \left( \max_{u^2 \leqslant n} \sum_{\substack{a \in \mathscr{A} \\ u(a) = u}} \frac{1}{v(a)} \right) \sum_{u=1}^{+\infty} \frac{1}{u^2} = \frac{\pi^2}{6} \max_{u^2 \leqslant n} \sum_{\substack{a \in \mathscr{A} \\ u(a) = u}} \frac{1}{v(a)}. \tag{4.60}$$

It follows from (4.58) and (4.60) that

$$\max_{u^2 \leqslant n} \sum_{\substack{a \in \mathscr{A} \\ u(a) = u}} \frac{1}{v(a)} \geqslant \frac{6}{\pi^2} (\tfrac{1}{2} + \varepsilon) \log n. \tag{4.61}$$

Assume that here the maximum is attained for, say, $u = u_0$, and let $\mathscr{V}$ denote the set of the integers $v$ such that there is an $a \in \mathscr{A}$ with $u(a) = u_0$ and $v(a) = v$. Then

$$v \leqslant n \text{ and } v \text{ is square-free for all } v \in \mathscr{V}, \tag{4.62}$$

and, by (4.61), we have

$$\sum_{v \in \mathscr{V}} \frac{1}{v} \geqslant \frac{6}{\pi^2} (\tfrac{1}{2} + \varepsilon) \log n. \tag{4.63}$$

For a positive integer $i$, let $d_{\mathscr{V}}(i)$ denote the number of integers $v$ with $v \mid i$, $v \in \mathscr{V}$. By (4.62), $d_{\mathscr{V}}(i)$ does not exceed the number of the square-free divisors of $i$, so that

$$d_{\mathscr{V}}(i) \leqslant 2^{\omega(i)} \qquad \text{for all } i. \tag{4.64}$$

By (4.63), for sufficiently large $n$ we have

$$\begin{aligned}\sum_{i\leqslant n} d_{\mathscr{V}}(i) &= \sum_{i\leqslant n} \sum_{\substack{v \mid i \\ v\in\mathscr{V}}} 1 \\ &= \sum_{v\in\mathscr{V}} \sum_{\substack{i<n \\ v \mid i}} 1 = \sum_{v\in\mathscr{V}} \left[\frac{n}{v}\right] \\ &\geqslant n \sum_{n\in\mathscr{V}} \frac{1}{v} - |\mathscr{V}| \geqslant \frac{6}{\pi^2}(\tfrac{1}{2}+\varepsilon)n\log n - n \\ &> \frac{6}{\pi^2}\left(\frac{1}{2}+\frac{\varepsilon}{2}\right)n\log n.\end{aligned} \tag{4.65}$$

Now we will show that there is an integer $j$ such that $j \leqslant n$,

$$2^{\omega(j)} > \frac{\varepsilon}{10}\log n \tag{4.66}$$

and

$$d_{\mathscr{V}}(j) > \left(\frac{1}{2}+\frac{\varepsilon}{10}\right)2^{\omega(j)}. \tag{4.67}$$

In fact, assume that, contrary to this statement, for all $i \leqslant n$ either

$$2^{\omega(j)} \leqslant \frac{\varepsilon}{10}\log n \tag{4.68}$$

or

$$d_{\mathscr{V}}(i) \leqslant \left(\frac{1}{2}+\frac{\varepsilon}{10}\right)2^{\omega(i)} \tag{4.69}$$

holds. Let $\mathscr{N}$ denote the set of the integers $i$ satisfying $i \leqslant n$ and (4.68). Then, by (4.64) and Lemma 17, for sufficiently large $n$ we have

$$\begin{aligned}\sum_{i\leqslant n} d_{\mathscr{V}} &\leqslant \sum_{i\in\mathscr{N}} 2^{\omega(i)} + \sum_{\substack{i\leqslant n \\ i\notin\mathscr{N}}} d_{\mathscr{V}}(i) \\ &\leqslant \frac{\varepsilon}{10}\log n \sum_{i\in\mathscr{N}} 1 + \sum_{i\leqslant n}\left(\frac{1}{2}+\frac{\varepsilon}{10}\right)2^{\omega(i)} \\ &\leqslant \frac{\varepsilon}{10}n\log n + \left(\frac{1}{2}+\frac{\varepsilon}{10}\right)\sum_{i\leqslant n} 2^{\omega(i)} \\ &< \frac{\varepsilon}{10}n\log n + \left(\frac{1}{2}+\frac{\varepsilon}{9}\right)\frac{6}{\pi^2}n\log n \\ &< \frac{6}{\pi^2}\left(\frac{1}{2}+\frac{\varepsilon}{2}\right)n\log n,\end{aligned}$$

which contradicts (4.65), and this proves the existence of a $j$ satisfying (4.66) and (4.67).

Write $S = \{p : p \text{ prime}, p \mid j\}$ so that

$$|S| = \omega(j), \tag{4.70}$$

and let $v_1, \dots, v_t$ (with $v_1 < \cdots < v_t$) denote the elements of $\mathcal{V}$ that divide $j$ so that, by (4.67) and (4.70),

$$t = d_{\mathcal{V}}(j) > \left(\frac{1}{2} + \frac{\varepsilon}{10}\right) 2^{\omega(j)} = \left(\frac{1}{2} + \frac{\varepsilon}{10}\right) 2^{|S'|}. \tag{4.71}$$

For $i = 1, 2, \dots, t$, write $S_i = \{p : p \text{ prime}, p \mid v_i\}$. By (4.62), the sets $S_1, S_2, \dots, S_t$ are distinct subsets of $S$, and their number, $t$, satisfies (4.71) which, by (4.66) and (4.70), for sufficiently large $n$ implies that

$$t > 2^{|S|-1} + 2k,$$

so that (2.19) in Lemma 6 holds with $2k+1$ in place of $k$, and thus the lemma can be applied (with $2k+1$ in place of $k$). We obtain that there are subsets $S_{i_1}, S_{i_2}, \dots, S_{i_{2k+1}}$ such that each $p \in S$ is contained in an even number of these subsets. Then all the prime factors of the product $v_{i_1} v_{i_2} \cdots v_{i_{2k+1}}$ belong to $S$, and this product is divisible by an even power of each of these primes. Thus this product is a square: $v_{i_1} v_{i_2} \cdots v_{i_{2k+1}} = x^2$. Then $u_0^2 v_{i_1}, u_0^2 v_{i_2}, \dots, u_0^2 v_{i_{2k+1}}$ are distinct elements of $\mathcal{A}$, and their product is a square:

$$(u_0^2 v_{i_1})(u_0^2 v_{i_2}) \cdots (u_0^2 v_{i_{2k+1}}) = (u_0^{2k+1} x)^2,$$

so that (4.59) holds, which completes the proof of the theorem.

## Acknowledgements

The research of A. Sárközy and V. T. Sós was partially supported by Hungarian National Foundation for Scientific Research, under Grant No. 1901.

## References

1. C. T. Benson, Minimal regular graphs of girths eight and twelve, *Can. J. Math.* **18** (1966), 1091–1094.
2. J. A. Bondy and M. Simonovits, Cycles of even length in graphs, *J. Combin. Theory, Ser. B*, **16** (1974), 97–105.
3. W. G. Brown, On graphs that do not contain a Thomsen graph, *Can. Math. Bull.* **9** (1966), 281–285.
4. P. Erdős, On sequences of integers no one of which divides the product of two others and on some related problems. *Tomsk. Gos. Univ. Ucen Zap.*, **2** (1938), 74–82.
5. P. Erdős, On some applications of graph theory to number theoretic problems, *Publ. Ramanujan Inst.*, **1** (1969), 131–136.
6. P. Erdős and A. Sárközy, On the number of prime factors of integers, *Acta Sci. Math. Szeged*, **42** (1980), 237–246.
7. P. Erdős and M. Simonovits, A limit theorem in graph theory, *Stud. Sci. Math. Hung.*, **1** (1966), 51–57.
8. P. Erdős, A Rényi and V. T. Sós, On a problem of graph theory, *Stud. Sci. Math. Hung.*, **1** (1966), 215–235.
9. G. H. Hardy and S. Ramanujan, The normal number of prime factors of a number *n*, *Q. J. Math.* **48** (1920), 76–92.
10. J. C. Lagarias, A. M. Odlyzko and J. B. Shearer, On the density of sequences of integers the sum of no two of which is a square, I and II, *J. Combin. Theory*, **33** (1982), 167–185; **34** (1983), 123–139.
11. A. Sárközy, Hybrid problems in number theory, in: *Number Theory, New York 1985–88*, Lecture Notes in Mathematics 1383, Springer-Verlag, Berlin, 1989, pp. 146–169.
12. M. Simonovits, Extremal graph theory, in: *Selected Topics in Graph Theory 2*, L. W. Beineke and R. J. Wilson (eds), Academic Press, London, 1983, pp. 161–200.
13. R. R. Singleton, On minimal graphs of maximum even girth, *J. Combin. Theory*, **1** (1966), 306–332.

*Received 9 September 1993 and accepted 25 July 1994*

P. Erdős, A. Sárközy and V. T. Sós
*Mathematical Institute of the Hungarian Academy of Sciences,*
*H-1053 Budapest, Reáltanoda u. 13–15, Hungary*