A note on Erdős problem #1062

Asymptotic density, initial layers, and local structure

OpenAI

15 April 2026

Abstract

Let $f(n)$ denote the maximum size of a set $A \subseteq \{1, \ldots, n\}$ containing no three distinct elements $a, b, c \in A$ with $a \mid b$ and $a \mid c$. Equivalently, $f(n)$ is the extremal function for finite 2-primitive sets. This note consolidates the local-component approach to the problem.

First, using McNew's local-to-global theorem for divisor-graph statistics, we prove that there exists an effectively computable constant λ such that for every $\varepsilon > 0$,

$$f(n) = \lambda n + O_\varepsilon\Big(n \exp\big(-(1-\varepsilon)\sqrt{\log n \log\log n}\big)\Big).$$

In particular, $\lim_{n\to\infty} f(n)/n = \lambda$ exists. Second, we identify the first four layers of the McNew expansion:

$$w_1 = \frac{1}{2}, \qquad w_2 = \frac{1}{6}, \qquad w_3 = 0, \qquad w_4 = \frac{1}{270},$$

so that

$$\lambda = \frac{2}{3} + \frac{1}{270} + \sum_{i\geq 5} w_i.$$

Here the evaluation of the fourth layer is an exact finite verification.

Third, we record several structural facts about the local optimization problem. Every finite 2-primitive set has a rigid height-two form "tops plus leaves". The extremal quantity on a finite divisor component is the optimum of an explicit 0–1 linear program. Finally, the fifth layer already contains infinitely many rooted divisor-component types, and we give exact finite computations for two sample point slices along the family $d_a = 27 \cdot 2^a$.

1 Introduction

Let

$$f(n) := \max\{|A| : A \subseteq \{1, \ldots, n\},\ \text{there are no distinct } a, b, c \in A \text{ with } a \mid b \text{ and } a \mid c\}.$$

This is Problem B24 in Guy's book [1]. Following Vijay and Hegarty, we call such sets 2-*primitive*. The main point of the present note is that the asymptotic part of the problem is a direct instance of McNew's divisor-graph framework [2]: after passing from $f(n)$ to its local increments, one obtains a bounded statistic depending only on rooted divisor components.

We also collect several exact local results that emerge naturally from the same point of view. The first three layers of the local expansion are elementary. The fourth layer can be evaluated exactly by a finite computer-assisted check. We then record three structural facts that are useful for further local analysis:

(i) every finite 2-primitive set has height two;

(ii) the local extremal problem is an exact binary linear optimization problem;

(iii) the fifth layer is already infinite-state in a very concrete sense.

2 Local increments on divisor components

For a finite set $S \subseteq \mathbb{N}$, let $\mathcal{D}(S)$ denote its *divisor graph*: the vertex set is S, and two distinct vertices $x, y \in S$ are adjacent if either $x \mid y$ or $y \mid x$.

Definition 2.1. *A set $A \subseteq S$ is* 2-primitive *if there do not exist three distinct elements $a, b, c \in A$ such that $a \mid b$ and $a \mid c$.*

For finite $S \subseteq \mathbb{N}$, write

$$m(S) := \max\{|A| : A \subseteq S \text{ is 2-primitive}\}.$$

For integers $1 \le k \le n$, set

$$F(k, n) := m([k, n]),$$

and adopt the convention $F(n + 1, n) := 0$.

Lemma 2.2 (Additivity over connected components)**.** *Let $S \subseteq \mathbb{N}$ be finite, and let $S = S_1 \sqcup \cdots \sqcup S_r$ be the decomposition of S into connected components of $\mathcal{D}(S)$. Then*

$$m(S) = \sum_{j=1}^{r} m(S_j).$$

Proof. A forbidden triple a, b, c with $a \mid b$ and $a \mid c$ lies inside a single connected component of $\mathcal{D}(S)$, since divisibility gives edges ab and ac. Hence $A \subseteq S$ is 2-primitive if and only if each $A \cap S_j$ is 2-primitive. Taking maxima componentwise proves the claim. □

For $1 \le k \le n$, let $C_{k,n}$ be the connected component of k in $\mathcal{D}([k, n])$.

Lemma 2.3 (The local increment)**.** *For $1 \le k \le n$ define*

$$v(k, n) := F(k, n) - F(k + 1, n). \tag{1}$$

Then

$$v(k, n) = m(C_{k,n}) - m(C_{k,n} \setminus \{k\}),$$

so in particular

$$v(k, n) \in \{0, 1\}.$$

Furthermore, $v(k, n)$ depends only on the linear isomorphism type of the rooted component $(C_{k,n}, k)$.

Proof. Deleting the vertex k from $[k, n]$ leaves every connected component of $\mathcal{D}([k, n])$ unchanged except the component $C_{k,n}$; there this deletion simply replaces $C_{k,n}$ by $C_{k,n} \setminus \{k\}$. By Lemma 2.2,

$$F(k, n) = m(C_{k,n}) + \sum_{D \neq C_{k,n}} m(D)$$

and

$$F(k + 1, n) = m(C_{k,n} \setminus \{k\}) + \sum_{D \neq C_{k,n}} m(D),$$

whence the displayed formula for $v(k,n)$.

The bounds $0 \le v(k,n) \le 1$ are immediate: removing one vertex can lower the maximum by at most one, and adding a new available vertex cannot decrease the maximum. Thus $v(k,n) \in \{0,1\}$.

Finally, suppose $\phi : C_{k,n} \to C_{k',n'}$ is a linear isomorphism, so $\phi(x) = \ell x$ for some positive integer ℓ and all $x \in C_{k,n}$, with $\phi(k) = k'$. Such a map preserves divisibility, hence carries 2-primitive subsets of $C_{k,n}$ bijectively onto 2-primitive subsets of $C_{k',n'}$. Therefore

$$m(C_{k,n}) = m(C_{k',n'}) \qquad \text{and} \qquad m(C_{k,n} \setminus \{k\}) = m(C_{k',n'} \setminus \{k'\}),$$

so $v(k,n) = v(k',n')$. ☐

By telescoping,

$$f(n) = F(1,n) = \sum_{k=1}^{n} v(k,n). \tag{2}$$

Thus the problem is reduced to estimating a bounded local statistic on rooted divisor components.

3 The asymptotic density via McNew's theorem

We use the following specialization of McNew's theorem. McNew formulates the dependence condition in terms of rooted graph isomorphism types of divisor components; in footnote 1 he notes that the same proof remains valid if one weakens this to linear isomorphism types, which is the form needed here [2, Theorem 3 and footnote 1].

Theorem 3.1 (McNew, specialized)**.** *Let $g(k,n)$ be a bounded function, $|g(k,n)| \le A$, and suppose that $g(k,n)$ depends only on the linear isomorphism type of the rooted connected component of k in the divisor graph of $[k,n]$. Then for every $\varepsilon > 0$ there exists an effectively computable constant*

$$C_g = \sum_{i=1}^{\infty} \sum_{P^+(d) \le i} \sum_{t \in [id,(i+1)d)} \frac{g(d,t)}{t(t+1)} \prod_{p \le i} \frac{p-1}{p}$$

(with the convention $P^+(1) = 1$) such that

$$\sum_{k=1}^{n} g(k,n) = nC_g + O_\varepsilon\Big(An \exp\big(-(1-\varepsilon)\sqrt{\log n \log\log n}\big)\Big).$$

In particular, C_g is effectively computable from the local values $g(d,t)$.

Proof. This is exactly McNew's theorem in the linear-isomorphism variant noted in footnote 1 of [2]. ☐

Theorem 3.2. *There exists an effectively computable constant $\lambda \in [0,1]$ such that, for every $\varepsilon > 0$,*

$$f(n) = \lambda n + O_\varepsilon\Big(n \exp\big(-(1-\varepsilon)\sqrt{\log n \log\log n}\big)\Big).$$

In particular,

$$\lim_{n\to\infty} \frac{f(n)}{n} = \lambda.$$

Moreover,

$$\lambda = \sum_{i=1}^{\infty} \sum_{P^+(d) \le i} \sum_{t \in [id,(i+1)d)} \frac{v(d,t)}{t(t+1)} \prod_{p \le i} \frac{p-1}{p}.$$

Proof. Apply Theorem 3.1 to $g = v$. By Lemma 2.3, the function $v(k,n)$ is bounded by 1 and depends only on the linear isomorphism type of the rooted component of k in $\mathcal{D}([k,n])$. Hence

$$\sum_{k=1}^{n} v(k,n) = nC_v + O_\varepsilon\Big(n \exp(-(1-\varepsilon)\sqrt{\log n \log\log n})\Big).$$

Using the telescoping identity (2), this is exactly the claimed asymptotic formula with $\lambda := C_v$. □

For later use, write

$$\lambda = \sum_{i\geq 1} w_i, \qquad w_i := \rho_i \sum_{P^+(d)\leq i} \sum_{t=id}^{(i+1)d-1} \frac{v(d,t)}{t(t+1)}, \qquad \rho_i := \prod_{p\leq i}\Big(1-\frac{1}{p}\Big). \tag{3}$$

4 Structure of finite 2-primitive sets

The local optimization problem has a simple intrinsic shape.

Proposition 4.1 (height-two structure)**.** *Let $S \subseteq \mathbb{N}$ be finite, and let $A \subseteq S$ be 2-primitive. Write $T \subseteq A$ for the set of maximal elements of A under divisibility. Then:*

(i) A contains no divisibility chain $x \mid y \mid z$ of three distinct elements;

(ii) every element of $A \setminus T$ divides a unique *element of T;*

(iii) both T and $A \setminus T$ are antichains in the divisibility order.

Equivalently, every 2-primitive set has the form

$$A = T \sqcup L,$$

where T is an antichain of maximal vertices, L is an antichain of leaves, and each $x \in L$ divides exactly one element of T.

Proof. If $x \mid y \mid z$ are distinct elements of A, then x divides both y and z, contradicting 2-primitivity. This proves (i).

Let $x \in A \setminus T$. Since x is not maximal, there exists $y \in A$ with $x \mid y$ and $x < y$. By (i), such a y must already be maximal in A, because otherwise there would be some $z \in A$ with $y \mid z$ and hence a chain $x \mid y \mid z$. Thus x divides at least one element of T.

If x divided two distinct elements $y, z \in T$, then x would divide two distinct larger elements of A, again contradicting 2-primitivity. Hence the maximal element above x is unique. This proves (ii).

The set T is an antichain by maximality. If $x, x' \in A \setminus T$ with $x \mid x'$ and $x \neq x'$, let $y \in T$ be the unique maximal element above x'. Then $x \mid x' \mid y$ is a chain of length three in A, contradicting (i). So $A \setminus T$ is also an antichain. This proves (iii), and the equivalent formulation is just a restatement. □

Corollary 4.2. *Let $d \in \mathbb{N}$, let $t < 6d$, and let $A \subseteq C_{d,t}$ be 2-primitive with $d \in A$. Then A contains at most one of the four integer multiples*

$$2d,\ 3d,\ 4d,\ 5d.$$

Proof. The root d divides each of $2d, 3d, 4d, 5d$. Thus if two of these vertices belonged to A, the root would divide two distinct larger elements of A, contradicting 2-primitivity. □

Proposition 4.3 (exact binary optimization). *Let $S \subseteq \mathbb{N}$ be finite. Then $m(S)$ is the optimum of*

$$\max \sum_{x \in S} z_x$$

subject to the constraints

$$z_x + z_y + z_z \le 2$$

for every triple of distinct elements $x, y, z \in S$ with $x \mid y$ and $x \mid z$, together with the binary conditions $z_x \in \{0, 1\}$.

Consequently, for each pair (d, t), the local value $v(d, t)$ is the difference of two exact finite optimizations: the optimum on $C_{d,t}$ and the optimum on $C_{d,t}$ under the additional constraint $z_d = 0$.

Proof. A subset $A \subseteq S$ is 2-primitive if and only if no forbidden triple (x, y, z) with $x \mid y$ and $x \mid z$ is entirely contained in A. Writing $z_x = 1$ when $x \in A$ and $z_x = 0$ otherwise, this is exactly the family of linear inequalities

$$z_x + z_y + z_z \le 2.$$

Thus feasible binary points are in bijection with 2-primitive subsets of S, and the objective $\sum z_x$ is exactly $|A|$. This proves the formula for $m(S)$.

Applying this to $S = C_{d,t}$ gives the first optimum. Imposing $z_d = 0$ restricts the feasible subsets to those avoiding the root, which is exactly the optimization defining $m(C_{d,t} \setminus \{d\})$. □

5 The first four layers

The first few local layers in (3) can be described explicitly.

Proposition 5.1. *For every $d \in \mathbb{N}$ one has:*

(i) if $2d \le t < 3d$, then $v(d, t) = 1$;

(ii) if $3d \le t < 4d$, then $v(d, t) = 0$.

Proof. (i) In the interval $[d, t]$ with $t < 3d$, the only multiple of d larger than d is $2d$. Hence if $A \subseteq C_{d,t} \setminus \{d\}$ is 2-primitive, then $A \cup \{d\}$ is still 2-primitive: the new vertex d divides at most the single vertex $2d$, while no existing vertex can suddenly acquire a second larger multiple because there is no selected vertex below d. Therefore

$$m(C_{d,t}) = m(C_{d,t} \setminus \{d\}) + 1,$$

so $v(d, t) = 1$.

(ii) Now assume $3d \le t < 4d$, and let $A \subseteq C_{d,t}$ be a 2-primitive set containing d. Since $d \mid 2d$ and $d \mid 3d$, the set A contains at most one of $2d, 3d$. Define

$$B := (A \setminus \{d\}) \cup \{2d, 3d\}.$$

Then $|B| \ge |A|$. We claim that B is still 2-primitive.

First, d is removed, so there is no vertex of B dividing both $2d$ and $3d$: indeed any common divisor of $2d$ and $3d$ lying in $[d, t]$ must divide d, hence must equal d itself.

Second, no vertex $x > d$ divides $2d$, because every proper divisor of $2d$ is at most d.

Third, the only possible divisor of $3d$ larger than d is $3d/2$ (when d is even). But then the only multiple of $3d/2$ in $[d, t] \subseteq [d, 4d)$ is $3d$ itself, because $3 \cdot (3d/2) = 9d/2 > 4d > t$.

Finally, neither $2d$ nor $3d$ divides any larger vertex of $[d, t]$, since $2(2d) = 4d > t$ and $2(3d) = 6d > t$.

Thus B is 2-primitive. Starting from a maximum 2-primitive set A containing d, we obtain a 2-primitive set $B \subseteq C_{d,t} \setminus \{d\}$ with $|B| \geq |A|$. Hence

$$m(C_{d,t} \setminus \{d\}) \geq m(C_{d,t}),$$

and therefore $v(d, t) = 0$. □

Corollary 5.2. *The first three layers are*

$$w_1 = \frac{1}{2}, \qquad w_2 = \frac{1}{6}, \qquad w_3 = 0.$$

Equivalently,

$$\lambda = \frac{2}{3} + \sum_{i \geq 4} w_i.$$

Proof. For $i = 1$ the only term is $(d, t) = (1, 1)$, and $v(1, 1) = 1$, contributing $1/2$. For $i = 2$, Proposition 5.1(i) shows that every local value in the range $t \in [2d, 3d)$ equals 1, so the total $i = 2$ contribution is the full layer mass $1/(2 \cdot 3) = 1/6$. For $i = 3$, Proposition 5.1(ii) shows that every local value in the range $t \in [3d, 4d)$ equals 0, so the $i = 3$ contribution vanishes. □

Proposition 5.3 (computer-assisted classification of the fourth layer)**.** *For every 4-smooth integer d and every integer $t \in [4d, 5d)$ one has*

$$v(d, t) = 1 \iff 6 \mid d \quad \text{and} \quad t \geq \frac{9d}{2}.$$

Consequently,

$$w_4 = \frac{1}{270}.$$

Proof. For d 4-smooth and $t \in [4d, 5d)$, the rooted component of d in $[d, t]$ depends only on the scaled ratios $x/d \in [1, 5)$ that can be reached from 1 by multiplying or dividing by 2 and 3 while staying in the interval. This yields only finitely many possibilities.

For $t < \frac{9d}{2}$, every vertex of the rooted component is represented, after scaling by d, by a ratio in the finite set

$$\left\{1,\ \frac{4}{3},\ \frac{3}{2},\ 2,\ \frac{8}{3},\ 3,\ 4\right\},$$

with the obvious divisibility restrictions according to whether $2 \mid d$ and $3 \mid d$. For $t \geq \frac{9d}{2}$, additional ratios can occur, but only from the finite set

$$\left\{\frac{9}{8},\ \frac{27}{16},\ \frac{9}{4},\ \frac{27}{8},\ \frac{9}{2}\right\},$$

again subject only to the 2- and 3-adic divisibility of d. Thus only finitely many rooted component types occur in the whole layer $i = 4$.

For each of these finitely many rooted graphs, the local value $v(d, t)$ was computed exactly using Proposition 4.3. The outcome is precisely the displayed criterion.

Once the criterion is known, the value of w_4 follows immediately. Since $\rho_4 = (1-\frac{1}{2})(1-\frac{1}{3}) = \frac{1}{3}$, we obtain

$$\begin{aligned} w_4 &= \frac{1}{3} \sum_{\substack{P^+(d)\le 4 \\ 6|d}} \sum_{t=9d/2}^{5d-1} \frac{1}{t(t+1)} \\ &= \frac{1}{3} \sum_{\substack{P^+(d)\le 4 \\ 6|d}} \left(\frac{1}{9d/2} - \frac{1}{5d} \right) = \frac{1}{3} \cdot \frac{1}{45} \sum_{\substack{P^+(d)\le 4 \\ 6|d}} \frac{1}{d}. \end{aligned}$$

Every such d has the form $2^a 3^b$ with $a, b \ge 1$, so

$$\sum_{\substack{P^+(d)\le 4 \\ 6|d}} \frac{1}{d} = \sum_{a\ge 1} \frac{1}{2^a} \sum_{b\ge 1} \frac{1}{3^b} = 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

Therefore

$$w_4 = \frac{1}{3} \cdot \frac{1}{45} \cdot \frac{1}{2} = \frac{1}{270}.$$

□

Corollary 5.4. *The first four layers are*

$$w_1 = \frac{1}{2}, \qquad w_2 = \frac{1}{6}, \qquad w_3 = 0, \qquad w_4 = \frac{1}{270},$$

and hence

$$\lambda = \frac{2}{3} + \frac{1}{270} + \sum_{i\ge 5} w_i.$$

6 Local structure in the fifth layer

The fifth layer is already much richer than the first four.

Proposition 6.1. *For $m \ge 0$, let*

$$d_m := 2^m, \qquad t_m := 6 \cdot 2^m - 1,$$

and write $C_m := C_{d_m, t_m}$. Then the rooted connected components

$$(C_m, d_m)$$

are pairwise non-isomorphic. In particular, the fifth local layer contains infinitely many rooted component types.

Proof. Define a sequence of positive rationals $(q_n)_{n\ge 0}$ recursively by

$$q_0 := 1, \qquad q_{n+1} := \begin{cases} 3q_n, & q_n < 2, \\ q_n/2, & q_n \ge 2. \end{cases}$$

We first note that every q_n lies in $[1, 6)$. Indeed, if $q_n < 2$ then $q_{n+1} = 3q_n < 6$ and $q_{n+1} > 1$; if $q_n \ge 2$ then $q_{n+1} = q_n/2 \in [1, 3)$.

Write

$$q_n = \frac{3^{a_n}}{2^{b_n}}$$

with $a_n, b_n \in \mathbb{N}$. Since multiplication by 3 leaves the denominator exponent unchanged and division by 2 increases it by 1, the sequence (b_n) is nondecreasing and each increment is either 0 or 1.

The values q_n are all distinct. For if $q_n = q_{n'}$ with $n > n'$, then

$$3^{a_n - a_{n'}} = 2^{b_n - b_{n'}},$$

which forces $a_n = a_{n'}$ and $b_n = b_{n'}$, hence $n = n'$.

Now $b_n \to \infty$. Otherwise b_n would be bounded. Since $q_n \in [1, 6)$, the quantity

$$a_n \log 3 - b_n \log 2 = \log q_n$$

is bounded; if b_n were bounded then a_n would also be bounded, contradicting the fact that the pairs (a_n, b_n) are all distinct. Because $b_0 = 0$, the sequence (b_n) is nondecreasing, and its increments are at most 1, every nonnegative integer occurs as some b_n.

Fix m. Whenever $b_n \le m$, the rational number $d_m q_n = 2^m q_n$ is an integer. Since $1 \le q_n < 6$, it lies in $[d_m, 6d_m)$ and hence, being integral, in $[d_m, t_m]$. Consecutive terms of the q-orbit correspond to divisibility edges: if $q_n < 2$ then $q_{n+1} = 3q_n$, so $d_m q_n \mid d_m q_{n+1}$; if $q_n \ge 2$ then $q_{n+1} = q_n/2$, so $d_m q_{n+1} \mid d_m q_n$. Therefore all integers $d_m q_n$ with $b_n \le m$ lie in the connected component C_m.

Now scaling by 2 gives an injective map $x \mapsto 2x$ from C_m into C_{m+1}, because divisibility and the interval bounds are preserved. Hence $2C_m \subseteq C_{m+1}$, and every element of $2C_m$ is even.

On the other hand, since some index n satisfies $b_n = m + 1$, the integer

$$2^{m+1} q_n = 3^{a_n}$$

is an odd vertex of C_{m+1}. This odd vertex is not contained in $2C_m$. Therefore

$$|C_{m+1}| > |C_m| \qquad (m \ge 0).$$

In particular the rooted components (C_m, d_m) have strictly increasing numbers of vertices, so they are pairwise non-isomorphic. □

Remark 6.2. *Thus the fifth layer is not finite-state in any naive sense. Even along the single family* $(2^m, 6 \cdot 2^m - 1)$ *one sees infinitely many distinct rooted divisor components.*

The next proposition records two exact finite computations along the family $d_a = 27 \cdot 2^a$. These are simply local data points, included here because they show how varied the fifth-layer behaviour already is.

Proposition 6.3 (computer-assisted finite verification). *For a rational* $r \in [5, 6)$ *define*

$$u_a(r) := v(27 \cdot 2^a, \lfloor r\, 27 \cdot 2^a \rfloor).$$

Then the following exact values hold.

(i) If $r_1 = \frac{729}{128}$*, then for* $0 \le a \le 100$*,*

$$(u_a(r_1))_{a=0}^{100} = 00010001000100011101000\,1^{78}.$$

In particular,

$$u_a\left(\frac{729}{128}\right) = 1 \qquad (23 \le a \le 100).$$

(ii) If $r_2 = \frac{6075}{1024}$*, then for* $0 \le a \le 74$*,*

$$(u_a(r_2))_{a=0}^{74} = 0\,1^{11}\,0^3\,1^5\,0^3\,1^3\,0^7\,1\,0^7\,1\,(01)^{11}\,0^7\,1\,0^3,$$

while for $75 \le a \le 140$ *one has the exact parity pattern*

$$u_a\left(\frac{6075}{1024}\right) = \begin{cases} 1, & a \text{ odd}, \\ 0, & a \text{ even}. \end{cases}$$

All values were computed exactly from Proposition 4.3.

7 Concluding summary

The divisor-component viewpoint yields a compact package of results for Erdős problem #1062.

(1) The extremal function $f(n)$ has an asymptotic density λ, with an effective McNew-type error term.

(2) The first four local layers are explicit, giving

$$\lambda = \frac{2}{3} + \frac{1}{270} + \sum_{i \ge 5} w_i.$$

(3) The finite local optimization problem has a rigid height-two structure and an exact binary linear-program formulation.

(4) The fifth layer already contains infinitely many rooted component types, and concrete exact computations show varied behaviour even along a single dyadic tower.

These statements are independent of any further question about the arithmetic nature of λ.

References

[1] Richard K. Guy, *Unsolved Problems in Number Theory*, 3rd ed., Springer, New York, 2004.

[2] Nathan McNew, *Counting primitive subsets and other statistics of the divisor graph of* $\{1, 2, \ldots, n\}$, European Journal of Combinatorics **92** (2021), 103237.