A discrete-divergence proof of the asymptotic frontier theorem for Erdős problem #858

Abstract

We give a self-contained rewrite of the streamlined asymptotic proof for Erdős problem #858 in the language of a discrete divergence theorem, in the same spirit as Tao's recent flow proof of problem #1196. The Bellman recursion disappears: for a natural reciprocal flow on the rooted tree attached to the relation

$$a \preceq b \iff b = a \text{ or } b = at \text{ for some } t > 1 \text{ with } P^-(t) > a,$$

the reciprocal weight of any admissible antichain is exactly the total divergence of the upset generated by that antichain. The extremal problem is then reduced to understanding the sign of the divergence. This gives a short proof that

$$M(N) := \max_A \sum_{n \in A} \frac{1}{n} = (c_2 + o(1)) \log N,$$

where the maximum is over all admissible $A \subseteq \{1, \ldots, N\}$, and where c_2 is the same constant as in the recent streamlined note.

1 Introduction

Let $P^-(m)$ denote the least prime factor of $m > 1$. For $N \geq 1$, let $M(N)$ be the maximum of

$$\sum_{n \in A} \frac{1}{n}$$

over all sets $A \subseteq \{1, \ldots, N\}$ containing no pair a, b with

$$b = at, \qquad P^-(t) > a.$$

Equivalently, if we define a partial order on the positive integers by

$$a \preceq b \iff b = a \text{ or } b = at \text{ for some } t > 1 \text{ with } P^-(t) > a,$$

then the admissible sets are exactly the antichains for $\preceq$.

This note is a divergence-theorem rewrite of the streamlined asymptotic argument in the recent note on problem #858. The starting point is Tao's observation, in his proof of problem #1196, that one can work directly with a flow and its divergence rather than with explicit Markov chains. Here the relevant flow is much simpler: it is just the reciprocal flow $1/n$ down the parent tree.

Our main result is the following.

Theorem 1. *For $1/4 \le u \le 1/2$, define*

$$\Phi(u) := \log \frac{1-u}{u} + \int_u^{(1-u)/2} \frac{1}{v} \log \frac{1-u-v}{v}\, dv,$$

where the integral is interpreted as 0 *when* $u \ge 1/3$*. Let* $\alpha_2 \in (1/4, 1/3)$ *be the unique solution of*

$$\Phi(\alpha_2) = 1,$$

and set

$$c_2 := \frac{1}{2} + \int_{\alpha_2}^{1/2} (1 - \Phi(u))\, du.$$

Then

$$M(N) = (c_2 + o(1)) \log N \qquad (N \to \infty).$$

More precisely, for all sufficiently large N *there exists an integer* K_N *such that the frontier*

$$A_N(K_N) := \{n \le N : \pi(n) \le K_N < n\}$$

is an optimal admissible set and

$$K_N = N^{\alpha_2 + o(1)}.$$

The key point is that the optimal antichain is obtained by keeping precisely the positive-divergence region of the flow.

2 The rooted tree and the divergence identity

We begin with the structure of the order $\preceq$.

Lemma 2 (ordered-factorization criterion)**.** *Let*

$$n = p_1 p_2 \cdots p_r \qquad (p_1 \le p_2 \le \cdots \le p_r)$$

be the ordered prime factorization of n*, and write*

$$P_k := p_1 \cdots p_k \qquad (0 \le k \le r),$$

with $P_0 := 1$*. Then the proper ancestors of* n *are exactly the prefix products* P_k *with* $0 \le k < r$ *and*

$$p_{k+1} > P_k.$$

Hence every $n > 1$ *has a unique parent* $\pi(n)$*, namely the largest proper ancestor of* n*. Joining each* $n > 1$ *to* $\pi(n)$ *makes* $\{1, \ldots, N\}$ *into a rooted tree with root* 1*.*

Proof. Suppose first that $a \prec n$. Then $n = at$ with $t > 1$ and $P^-(t) > a$. Every prime factor of t is therefore $> a$, so every prime factor of n that is at most a must already lie in a with exactly the same multiplicity. Since the prime factors of n are listed in nondecreasing order, this forces $a = P_k$ for some $k < r$. The least prime factor of n/a is then p_{k+1}, so the condition $a \prec n$ is exactly $p_{k+1} > a = P_k$.

Conversely, if $a = P_k$ and $p_{k+1} > P_k$, then all prime factors of n/a are at least $p_{k+1} > a$, so indeed $a \prec n$.

Thus the proper ancestors are exactly the claimed prefix products. In particular the largest such prefix product is uniquely determined, so every $n > 1$ has a unique parent. Since $1 = P_0$ is always an ancestor and parents are strictly smaller than children, the resulting graph is a rooted tree with root 1. □

For $a \le N$, write

$$\mathrm{ch}_N(a) := \{n \le N : \pi(n) = a\}$$

for the children of a, and

$$T_N(a) := \{n \le N : a \preceq n\}$$

for the descendant subtree rooted at a.

We now put a flow on this tree. Add one extra sink vertex 0, and extend the parent map by the convention

$$\pi(1) := 0.$$

For each $1 \le n \le N$, place one directed edge

$$n \to \pi(n)$$

of weight $1/n$. Thus the outflow from every $a \in \{1, \dots, N\}$ is exactly $1/a$.

For $a \le N$ define

$$C_N(a) := \sum_{\pi(n)=a} \frac{1}{n} = \sum_{n \in \mathrm{ch}_N(a)} \frac{1}{n}, \qquad d_N(a) := \frac{1}{a} - C_N(a).$$

Then $C_N(a)$ is the inflow at a, and $d_N(a)$ is its divergence.

We shall use the following elementary form of the discrete divergence theorem.

Lemma 3 (discrete divergence theorem)**.** *Let G be a finite directed weighted graph. If $\mathrm{div}(v)$ denotes the outflow from v minus the inflow into v, then for every finite set Ω of vertices,*

$$\sum_{v \in \Omega} \mathrm{div}(v) = \mathrm{out}(\Omega) - \mathrm{in}(\Omega),$$

where $\mathrm{out}(\Omega)$ is the total flow from Ω to its complement and $\mathrm{in}(\Omega)$ is the total flow from the complement into Ω.

Proof. Sum, over $v \in \Omega$, the outflow from v minus the inflow into v. Any edge with both endpoints in Ω contributes once positively and once negatively and therefore cancels. The only surviving terms are exactly the edges leaving Ω and the edges entering Ω. □

The next observation is the combinatorial heart of the argument.

Lemma 4 (antichains and upsets)**.** *Let $A \subseteq \{1, \dots, N\}$ be an antichain, and define its generated upset by*

$$\Omega(A) := \bigcup_{a \in A} T_N(a).$$

Then the minimal elements of $\Omega(A)$ are exactly the elements of A.

Conversely, if $\Omega \subseteq \{1, \dots, N\}$ is an upset, then the set $\min(\Omega)$ of minimal elements of Ω is an antichain and

$$\Omega = \bigcup_{a \in \min(\Omega)} T_N(a).$$

Hence the map $A \mapsto \Omega(A)$ is a bijection from antichains to upsets, with inverse $\Omega \mapsto \min(\Omega)$.

Proof. If $n \in T_N(a_1) \cap T_N(a_2)$, then both a_1 and a_2 are ancestors of n. By Lemma 2, the ancestors of a fixed vertex form a chain, so one of a_1, a_2 is below the other. Since A is an antichain, this forces $a_1 = a_2$. Thus the subtrees $T_N(a)$ with $a \in A$ are pairwise disjoint, and their minimal elements are exactly the roots $a \in A$.

Now let Ω be an upset. Clearly $\min(\Omega)$ is an antichain. Fix $n \in \Omega$. Starting from n, move downward along the unique parent chain for as long as the current vertex remains in Ω. Because the tree is finite, this process terminates at some vertex $a \in \Omega$ whose parent is not in Ω (or $a = 1$). Such a vertex is minimal in Ω, and of course $a \preceq n$. Hence $n \in T_N(a)$ for some $a \in \min(\Omega)$, so

$$\Omega \subseteq \bigcup_{a \in \min(\Omega)} T_N(a).$$

The reverse inclusion holds because Ω is an upset. Therefore

$$\Omega = \bigcup_{a \in \min(\Omega)} T_N(a),$$

as claimed. ☐

Proposition 5 (divergence representation of antichain weight)**.** *For every antichain $A \subseteq \{1, \ldots, N\}$,*

$$\sum_{a \in A} \frac{1}{a} = \sum_{n \in \Omega(A)} d_N(n).$$

Equivalently, for every upset $\Omega \subseteq \{1, \ldots, N\}$,

$$\sum_{a \in \min(\Omega)} \frac{1}{a} = \sum_{n \in \Omega} d_N(n).$$

Proof. Take an upset Ω. No edge enters Ω: if $x \notin \Omega$ and $x \to y$ with $y \in \Omega$, then $y = \pi(x)$ and hence $y \preceq x$, so the upset property would force $x \in \Omega$, a contradiction.

The edges leaving Ω are exactly the edges

$$a \to \pi(a) \qquad (a \in \min(\Omega)),$$

one from each minimal element, each of weight $1/a$. Therefore Lemma 3 gives

$$\sum_{n \in \Omega} d_N(n) = \mathrm{out}(\Omega) = \sum_{a \in \min(\Omega)} \frac{1}{a}.$$

Applying this to $\Omega = \Omega(A)$ and using Lemma 4 yields the first formula. ☐

A particularly important family of upsets is obtained by cutting at a numerical height. For $0 \le K \le N$, define

$$\Omega_K := \{n \in \{1, \ldots, N\} : n > K\}$$

and the corresponding frontier

$$A_N(K) := \min(\Omega_K) = \{n \le N : \pi(n) \le K < n\}.$$

Its reciprocal weight is

$$S_N(K) := \sum_{n \in A_N(K)} \frac{1}{n}.$$

Since descendants are always larger than their ancestors, Ω_K is indeed an upset.

Corollary 6 (frontier divergence formula)**.** *For every* $0 \le K \le N$*,*

$$S_N(K) = \sum_{n>K} d_N(n).$$

In particular, for $1 \le K \le N$*,*

$$S_N(K) - S_N(K-1) = C_N(K) - \frac{1}{K}.$$

Proof. Apply Proposition 5 with $\Omega = \Omega_K$. Then

$$S_N(K) = \sum_{n \in \Omega_K} d_N(n) = \sum_{n>K} d_N(n).$$

Subtracting the identities for K and $K-1$ gives

$$S_N(K) - S_N(K-1) = -d_N(K) = C_N(K) - \frac{1}{K}. \qquad \square$$

Thus the extremal problem has been rewritten as a problem about the sign of the divergence d_N.

3 The sign pattern and the optimal frontier

We first isolate the children of a vertex.

Lemma 7 (prime child lemma)**.** *If* $p > a$ *is prime and* $ap \le N$*, then* $\pi(ap) = a$*.*

Proof. Certainly $a \prec ap$. Suppose there were a proper ancestor b with $a < b < ap$. Then $b = au$ where every prime factor of u exceeds a. Since every prime factor of a is at most $a < p$, we have $\gcd(a, u) = 1$. Also $b \mid ap$, so $u \mid p$, hence $u = p$. Therefore $b = ap$, contradicting $b < ap$. $\square$

For $a \le N$ define

$$P_N(a) := \sum_{a<p\le N/a} \frac{1}{p}, \qquad Q_N(a) := \sum_{\substack{a<p\le q \\ apq\le N}} \frac{1}{pq},$$

where all sums are over primes.

Lemma 8 (prime-semiprime children above the quarter-power layer)**.** *If* $a > N^{1/4}$*, then every child of* a *is of exactly one of the two forms*

$$ap \qquad \text{or} \qquad apq,$$

with primes $a < p \le q$ *and* $apq \le N$*. Consequently,*

$$aC_N(a) = P_N(a) + Q_N(a) \qquad (a > N^{1/4}).$$

Proof. Let n be a child of a. Then $n = at$, where every prime factor of t exceeds a. Since $a > N^{1/4}$,

$$t \le N/a < a^3.$$

Because every prime factor of t is $> a$, the integer t cannot have three or more prime factors. Thus t is either a prime $p > a$ or a product pq with primes $a < p \le q$.

Conversely, every ap with $p > a$ prime is a child by Lemma 7. Now let $n = apq$ with primes $a < p \le q$ and $apq \le N$. In the ordered factorization of n, all prime factors of a come first, then p, then q. By Lemma 2, the only possible proper ancestor of n strictly above a is ap, and this would require $q > ap$. But

$$ap > a^2 > N^{1/2}, \qquad q \le \frac{N}{ap} < N^{1/2} < ap,$$

so $q > ap$ is impossible. Hence $\pi(apq) = a$.

Therefore the children of a are exactly the displayed numbers, and summing $1/n$ over them gives

$$C_N(a) = \sum_{a<p\le N/a} \frac{1}{ap} + \sum_{\substack{a<p\le q \\ apq\le N}} \frac{1}{apq} = \frac{P_N(a) + Q_N(a)}{a}.$$

□

We now determine the sign of the divergence.

Lemma 9 (negative divergence below the quarter-power layer)**.** *Let $L := \lfloor N^{1/4} \rfloor$. Then for all sufficiently large N,*

$$d_N(a) < 0 \qquad (1 \le a \le L).$$

Proof. By Lemma 7, each prime p with $a < p \le N/a$ contributes the child ap, so

$$aC_N(a) \ge \sum_{a<p\le N/a} \frac{1}{p} = P_N(a) \ge P_N(L)$$

for every $a \le L$.

By Mertens' theorem,

$$\sum_{p\le x} \frac{1}{p} = \log\log x + B_1 + o(1),$$

so

$$P_N(L) = \sum_{L<p\le N/L} \frac{1}{p} = \log \frac{\log(N/L)}{\log L} + o(1) = \log 3 + o(1) > 1.$$

Hence $aC_N(a) > 1$ for all $a \le L$ once N is large enough, and therefore

$$d_N(a) = \frac{1}{a} - C_N(a) < 0.$$

□

Proposition 10 (eventual sign threshold and optimal frontier)**.** *For all sufficiently large N, there exists an integer $K_N \in [L, \lfloor\sqrt{N}\rfloor]$ such that*

$$d_N(a) \le 0 \iff a \le K_N.$$

Consequently,

$$M(N) = S_N(K_N),$$

so the frontier $A_N(K_N)$ is an optimal admissible set.

Proof. By Lemma 9, we have $d_N(a) < 0$ for all $a \le L$ once N is large enough.

Now let $a > L$. By Lemma 8,

$$d_N(a) = \frac{1 - P_N(a) - Q_N(a)}{a}.$$

As a increases, the summation domains defining $P_N(a)$ and $Q_N(a)$ only shrink, so $P_N(a) + Q_N(a)$ is nonincreasing in a. Also $P_N(a) + Q_N(a) = 0$ for $a > \sqrt{N}$, because then $N/a < a$ and there are no children at all. Since $P_N(L) > 1$ for large N, there is therefore a largest integer $K_N \ge L$ with $d_N(K_N) \le 0$, and it satisfies $K_N \le \lfloor \sqrt{N} \rfloor$. By monotonicity,

$$d_N(a) \le 0 \iff a \le K_N.$$

Now let A be any admissible antichain, and put $\Omega := \Omega(A)$. By Proposition 5,

$$\sum_{a \in A} \frac{1}{a} = \sum_{n \in \Omega} d_N(n).$$

Since $d_N(n) \le 0$ for $n \le K_N$ and $d_N(n) > 0$ for $n > K_N$, we have

$$\sum_{n \in \Omega} d_N(n) \le \sum_{n \in \Omega \cap (K_N, N]} d_N(n) \le \sum_{n > K_N} d_N(n).$$

The set $\Omega_{K_N} = \{n > K_N\}$ is an upset, so by Proposition 5 and Corollary 6 its boundary antichain is precisely $A_N(K_N)$ and has weight

$$\sum_{n > K_N} d_N(n) = S_N(K_N).$$

Therefore every admissible antichain has weight at most $S_N(K_N)$, while $A_N(K_N)$ attains that value. Hence $M(N) = S_N(K_N)$. $\square$

This is the direct analogue of Tao's proof of problem #1196: the optimizer is the boundary of the positive-divergence region.

4 Asymptotics of the threshold and the maximum

We now locate K_N and evaluate $M(N)$.

Lemma 11 (Mertens on polynomial intervals)**.** *Fix $\beta \in (0, 1)$. Uniformly for real numbers x, y with*

$$N^\beta \le x \le y \le N,$$

one has

$$\sum_{x < p \le y} \frac{1}{p} = \log \frac{\log y}{\log x} + o(1) \qquad (N \to \infty).$$

Proof. Let

$$B(x) := \sum_{p \le x} \frac{1}{p}.$$

Mertens' theorem gives $B(x) = \log \log x + B_1 + r(x)$ with $r(x) \to 0$ as $x \to \infty$. Hence

$$\sum_{x < p \le y} \frac{1}{p} = B(y) - B(x) = \log \frac{\log y}{\log x} + r(y) - r(x).$$

If $N^\beta \le x \le y \le N$, then both x and y tend to infinity uniformly with N, so $r(x)$ and $r(y)$ are uniformly $o(1)$. $\square$

Lemma 12 (prime-harmonic Riemann sums)**.** *Fix α with $1/4 \le \alpha < 1/3$, and let*

$$D_\alpha := \{(u, v) : \alpha \le u \le 1/3,\ u \le v \le (1-u)/2\}.$$

If $G : D_\alpha \to \mathbb{R}$ is continuous, then uniformly for $u \in [\alpha, 1/3]$,

$$\sum_{N^u < p \le N^{(1-u)/2}} \frac{G\Big(u, \frac{\log p}{\log N}\Big)}{p} = \int_u^{(1-u)/2} G(u, v)\, \frac{dv}{v} + o(1).$$

Proof. This is the usual Riemann-sum argument on short multiplicative slabs. Since D_α is compact, G is uniformly continuous and bounded there. Partition the v-interval into subintervals of mesh at most δ. On each slab $N^{v_j} < p \le N^{v_{j+1}}$, Lemma 11 gives

$$\sum_{N^{v_j} < p \le N^{v_{j+1}}} \frac{1}{p} = \log \frac{v_{j+1}}{v_j} + o(1) = \int_{v_j}^{v_{j+1}} \frac{dv}{v} + o(1),$$

uniformly in u and j. Replacing $G\Big(u, \frac{\log p}{\log N}\Big)$ on each slab by the constant $G(u, v_j)$ introduces an error bounded by the modulus of continuity of G at scale δ, times the uniformly bounded prime-harmonic mass of the whole interval. Summing over the slabs therefore yields a Riemann sum for

$$\int_u^{(1-u)/2} G(u, v)\, \frac{dv}{v},$$

plus an error that is $o(1)$ as $N \to \infty$ and then $\delta \to 0$, uniformly in u. □

Lemma 13 (harmonic Riemann sums)**.** *Let $0 < \alpha < \beta \le 1$, let $f : [\alpha, \beta] \to \mathbb{R}$ be continuous, and suppose*

$$\alpha \le \lambda_N \le \mu_N \le \beta.$$

Then

$$\frac{1}{\log N} \sum_{N^{\lambda_N} < a \le N^{\mu_N}} \frac{f\Big(\frac{\log a}{\log N}\Big)}{a} = \int_{\lambda_N}^{\mu_N} f(u)\, du + o(1),$$

uniformly in the choice of λ_N and μ_N.

Proof. Put $u_a := \frac{\log a}{\log N}$. Then

$$u_{a+1} - u_a = \frac{\log(1 + 1/a)}{\log N} = \frac{1}{a \log N} + O\bigg(\frac{1}{a^2 \log N}\bigg),$$

uniformly in a. Hence the displayed sum is a left-endpoint Riemann sum for

$$\int_{\lambda_N}^{\mu_N} f(u)\, du,$$

with total error

$$O\Bigg(\frac{1}{\log N} \sum_{a \ge N^\alpha} \frac{1}{a^2}\Bigg) = o(1).$$

Uniformity follows because f is uniformly continuous on the compact interval $[\alpha, \beta]$. □

For $1/4 \le u \le 1/2$, recall the definition

$$\Phi(u) := \log \frac{1-u}{u} + \int_u^{(1-u)/2} \frac{1}{v} \log \frac{1-u-v}{v}\, dv,$$

with the integral interpreted as 0 when $u \ge 1/3$.

Lemma 14 (uniform prime-semiprime asymptotics)**.** *Fix α with $1/4 \le \alpha < 1/2$. Then uniformly for integers a with*

$$N^\alpha \le a \le \sqrt{N}, \qquad u := \frac{\log a}{\log N},$$

one has

$$P_N(a) + Q_N(a) = \Phi(u) + o(1).$$

Proof. By Lemma 11,

$$P_N(a) = \sum_{a<p\le N/a} \frac{1}{p} = \log \frac{1-u}{u} + o(1)$$

uniformly for $N^\alpha \le a \le \sqrt{N}$.

If $u \ge 1/3$, then $Q_N(a) = 0$, because $apq \le N$ and $a < p \le q$ would imply

$$a^3 < apq \le N,$$

contradicting $a \ge N^{1/3}$. Thus only the range $u \in [\alpha, 1/3]$ needs analysis; this range is empty when $\alpha \ge 1/3$.

Assume therefore that $\alpha < 1/3$ and $u \in [\alpha, 1/3]$. Define

$$G(u, v) := \log \frac{1-u-v}{v}$$

on the compact triangle D_α. This function is continuous. For each outer prime p with $a < p \le N^{(1-u)/2}$, put

$$v := \frac{\log p}{\log N}.$$

Then another application of Lemma 11 gives

$$\sum_{p\le q\le N/(ap)} \frac{1}{q} = \log \frac{1-u-v}{v} + o(1) = G(u, v) + o(1),$$

uniformly for $(u, v) \in D_\alpha$. Since the prime-harmonic mass of the outer range is uniformly bounded, the total contribution of the uniform $o(1)$ term is still $o(1)$. Therefore

$$Q_N(a) = \sum_{a<p\le N^{(1-u)/2}} \frac{G\Big(u, \frac{\log p}{\log N}\Big)}{p} + o(1),$$

uniformly for $u \in [\alpha, 1/3]$. Lemma 12 now yields

$$Q_N(a) = \int_u^{(1-u)/2} \frac{1}{v} \log \frac{1-u-v}{v}\, dv + o(1),$$

uniformly in that range. Combining this with the asymptotic for $P_N(a)$ proves the claim. □

Lemma 15 (monotonicity of Φ)**.** *The function Φ is continuous and strictly decreasing on $[1/4, 1/2]$. Moreover,*

$$\Phi(1/4) > 1, \qquad \Phi(1/3) = \log 2 < 1.$$

Hence there is a unique $\alpha_2 \in (1/4, 1/3)$ with $\Phi(\alpha_2) = 1$.

Proof. Continuity is clear away from $u = 1/3$, and also at $u = 1/3$ because the interval of integration collapses there.

For $u \in [1/3, 1/2]$, the integral term vanishes, so

$$\Phi(u) = \log \frac{1-u}{u},$$

which is strictly decreasing.

For $u \in (1/4, 1/3)$, write

$$I(u) := \int_u^{(1-u)/2} \frac{1}{v} \log \frac{1-u-v}{v}\, dv.$$

Because the integrand vanishes at $v = (1-u)/2$, Leibniz' rule gives

$$I'(u) = -\frac{1}{u} \log \frac{1-2u}{u} - \int_u^{(1-u)/2} \frac{dv}{v(1-u-v)}.$$

Also,

$$\int_u^{(1-u)/2} \frac{dv}{v(1-u-v)} = \frac{1}{1-u} \log \frac{1-2u}{u}.$$

Hence

$$I'(u) = -\left(\frac{1}{u} + \frac{1}{1-u}\right) \log \frac{1-2u}{u} < 0,$$

since $u < 1/3$ implies $(1-2u)/u > 1$. The derivative of $\log \frac{1-u}{u}$ is also negative, so $\Phi'(u) < 0$ on $(1/4, 1/3)$ as well.

Finally,

$$\Phi(1/4) = \log 3 + \int_{1/4}^{3/8} \frac{1}{v} \log \frac{3/4 - v}{v}\, dv > \log 3 > 1,$$

while the integral vanishes at $u = 1/3$, giving $\Phi(1/3) = \log 2 < 1$. The existence and uniqueness of α_2 now follow from continuity and strict monotonicity. ∎

We can now locate the threshold.

Proposition 16 (location of the divergence threshold)**.** *The threshold in Proposition 10 satisfies*

$$K_N = N^{\alpha_2 + o(1)}.$$

Proof. Fix $\varepsilon > 0$.

For the upper bound, it is enough to consider $0 < \varepsilon < 1/2 - \alpha_2$, because larger values are weaker and in any case $K_N \le \sqrt{N}$. By Lemma 15, there exists $\eta > 0$ such that

$$\Phi(u) \le 1 - \eta \qquad (\alpha_2 + \varepsilon \le u \le 1/2).$$

Lemma 14 therefore gives, uniformly for integers a with

$$N^{\alpha_2 + \varepsilon} \le a \le \sqrt{N},$$

the estimate

$$P_N(a) + Q_N(a) \le 1 - \eta/2$$

for all sufficiently large N. Hence

$$d_N(a) = \frac{1 - P_N(a) - Q_N(a)}{a} \ge \frac{\eta}{2a} > 0$$

throughout that range. Since $d_N(a) \le 0$ for $a \le K_N$, this implies

$$K_N \le N^{\alpha_2 + \varepsilon}$$

for large N.

For the lower bound, first assume $0 < \varepsilon < \alpha_2 - 1/4$. By Lemma 15, there exists $\eta' > 0$ such that

$$\Phi(u) \ge 1 + \eta' \qquad (1/4 \le u \le \alpha_2 - \varepsilon).$$

Applying Lemma 14 uniformly on that interval, we get

$$P_N(a) + Q_N(a) \ge 1 + \eta'/2$$

for all integers a with

$$\lfloor N^{1/4} \rfloor < a \le N^{\alpha_2 - \varepsilon}$$

and all sufficiently large N. Hence

$$d_N(a) \le -\frac{\eta'}{2a} < 0$$

throughout that range. Since $d_N(a) \le 0$ exactly for $a \le K_N$, it follows that

$$K_N \ge N^{\alpha_2 - \varepsilon}$$

for all sufficiently large N. If $\varepsilon \ge \alpha_2 - 1/4$, the same conclusion follows a fortiori from any smaller positive value of ε.

Since $\varepsilon > 0$ was arbitrary, we obtain

$$K_N = N^{\alpha_2 + o(1)}.$$

□

Proof of Theorem 1. By Proposition 10 and Corollary 6,

$$M(N) = S_N(K_N) = \sum_{a > K_N} d_N(a).$$

For $a > \sqrt{N}$ there are no children, so $d_N(a) = 1/a$. Also $K_N \ge N^{1/4}$ for large N, and therefore Lemma 8 gives

$$d_N(a) = \frac{1 - P_N(a) - Q_N(a)}{a} \qquad (K_N < a \le \sqrt{N}).$$

Hence

$$M(N) = \sum_{\sqrt{N} < a \le N} \frac{1}{a} + \sum_{K_N < a \le \sqrt{N}} \frac{1 - P_N(a) - Q_N(a)}{a}.$$

The first sum is

$$\sum_{\sqrt{N} < a \le N} \frac{1}{a} = \frac{1}{2} \log N + O(1).$$

Set

$$\kappa_N := \frac{\log K_N}{\log N}.$$

By Proposition 16, we have $\kappa_N \to \alpha_2$. Choose $\delta > 0$ so small that $\alpha_2 - \delta > 1/4$. For large N, the lower limit K_N lies in $[N^{\alpha_2-\delta}, N^{\alpha_2+\delta}]$, so Lemma 14 applies uniformly on the interval $K_N < a \le \sqrt{N}$. Thus

$$1 - P_N(a) - Q_N(a) = 1 - \Phi\left(\frac{\log a}{\log N}\right) + o(1)$$

uniformly for $K_N < a \le \sqrt{N}$. Since

$$\sum_{K_N < a \le \sqrt{N}} \frac{1}{a} = O(\log N),$$

the total contribution of the uniform $o(1)$ term is $o(\log N)$. If

$$\mu_N := \frac{\log\lfloor\sqrt{N}\rfloor}{\log N},$$

then Lemma 13 yields

$$\frac{1}{\log N} \sum_{K_N < a \le \sqrt{N}} \frac{1 - P_N(a) - Q_N(a)}{a} = \int_{\kappa_N}^{\mu_N} (1 - \Phi(u))\, du + o(1).$$

Since $\kappa_N \to \alpha_2$, $\mu_N \to 1/2$, and $1 - \Phi(u)$ is continuous, the integral tends to

$$\int_{\alpha_2}^{1/2} (1 - \Phi(u))\, du.$$

Combining these estimates gives

$$\frac{M(N)}{\log N} \to \frac{1}{2} + \int_{\alpha_2}^{1/2} (1 - \Phi(u))\, du = c_2.$$

This proves the theorem. □

Remark 17. *Numerically one finds*

$$\alpha_2 \approx 0.28043830989, \qquad c_2 \approx 0.6187712111.$$

References

[1] T. F. Bloom, *Erdős Problem #858 - discussion thread*, https://www.erdosproblems.com/forum/thread/858.

[2] T. F. Bloom, *Erdős Problem #1196 - discussion thread*, https://www.erdosproblems.com/forum/thread/1196.

[3] P. Chojecki, *The asymptotic constant in Erdős problem #858*, https://www.ulam.ai/research/erdos858-asymptotic.pdf.

[4] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, 3rd ed., Graduate Studies in Mathematics 163, American Mathematical Society, 2015.