ARCHIVUM MATHEMATICUM (BRNO)
Tomus 56 (2020), 49–64

COMPOSITE POSITIVE INTEGERS WHOSE SUM OF PRIME FACTORS IS PRIME

Florian Luca and Damon Moodley

Abstract. In this note, we show that the counting function of the number of composite positive integers $n \le x$ such that $\beta(n) = \sum_{p\mid n} p$ is a prime is of order of magnitude at least $x/(\log x)^3$ and at most $x/\log x$.

1. Introduction

In this paper, for a positive integer n we set $\beta(n) = \sum_{p\mid n} p$. This function has been studied by a few authors before. It's first appearance was in papers of Pomerance and his co-authors, including Erdós (see [7], [8], [9]), where the positive integers n with $\beta(n) = \beta(n+1)$ were investigated. An example of such n is 714, a number which at that time appeared in the context of a baseball game and since then such numbers have been called Ruth-Aaron numbers. De Koninck and Luca studied positive integers n with $\beta(n) \mid n$ (see [3], [4]).

Let

$$\mathcal{B} := \{n \text{ composite} : \beta(n) \;\text{ is prime}\}\,.$$

For a subset $\mathcal{A}$ of positive integers and a real number $x \ge 1$, we write

$$\mathcal{A}(x) = \mathcal{A} \cap [1, x]\,.$$

We have the following theorem.

Theorem 1. *The estimates*

$$\frac{x}{(\log x)^3} \ll \#\mathcal{B}(x) \ll \frac{x}{\log x}$$

hold.

We believe the upper bound is closer to the truth and this would follow if we assume that $\beta(n)$ is "randomly distributed". In the last section, we provide a conditional proof of a lower bound of the same order of magnitude as the upper bound assuming a uniform version of the Bateman-Horn conjecture. In the same section, we conjecture that in fact $\#\mathcal{B}(x) = e^{\gamma}(1 + o(1))x/\log x$ holds as $x \to \infty$ and give some heuristics in support of our conjecture.

2020 *Mathematics Subject Classification*: primary 11N25; secondary 11N36.
Key words and phrases: primes, applications of sieve methods.
Received July 27, 2019, revised October 2019. Editor C. Greither.
DOI: 10.5817/AM2020-1-49

Throughout the paper, the Landau symbols O and o as well as the Vinogradov symbols $\ll$ and $\gg$ are used. The constants implied by them are absolute.

2. The upper bound

Here, we prove the upper bound from Theorem 1. We carve off a finite number of slices of $\mathcal{B}(x)$ denoted $\mathcal{B}_i(x)$ for $i = 1, 2, 3, 4$ each of cardinality $O(x/\log x)$. Then we will study the condition $\beta(n)$ is prime on the left-over subset of $n \le x$ denoted by $\mathcal{B}_5(x)$.

2.1. **Eliminating smooth numbers from $\mathcal{B}(x)$.** We let $P(n)$ be the largest prime factor of n. We put $y := \exp(\log x/\log\log x)$ and let

$$\mathcal{B}_1(x) := \{n \le x : P(n) \le y\}\,. \tag{1}$$

In classical notation, and in our range of y versus x, we have

$$\#\mathcal{B}_1(x) = \Psi(x, y) = \big(1 + o(1)\big)x\rho(u) = xu^{-(1+o(1))u}\,, \quad \text{where} \quad u := \frac{\log x}{\log y}\,,$$

(see Theorem 9.15 and Corollary 9.18 in [5]), where ρ is the Dickman function. Since for us $u = \log\log x$, we have $u^{-(1+o(1))u} = (\log x)^{-(1+o(1))\log\log\log x}$ as $x \to \infty$. Hence,

$$\#\mathcal{B}_1(x) = \frac{x}{(\log x)^{(1+o(1))\log\log\log x}} \ll \frac{x}{\log x}\,. \tag{2}$$

2.2. **Eliminating numbers divisible by the square of a large prime.** We put $z := (\log x)^{10}$ and let

$$\mathcal{B}_2(x) := \{n \le x : p^2 \mid n \text{ for some prime } p > z\}\,. \tag{3}$$

Given a prime p, the number of $n \le x$ divisible by p^2 is $\lfloor n/p^2\rfloor \le x/p^2$. Thus,

$$\#\mathcal{B}_2(x) \le \sum_{p>z} \frac{x}{p^2} < x\sum_{m>z}\frac{1}{m^2} \ll x\int_z^\infty \frac{dt}{t^2} \ll \frac{x}{z}.$$

In particular,

$$\#\mathcal{B}_2(x) \ll \frac{x}{\log x}\,. \tag{4}$$

2.3. **Applying the sieve.** Assume $n \in \mathcal{B}(x)\setminus(\mathcal{B}_1(x) \cup \mathcal{B}_2(x))$. Then $n = Pm$, where $P = P(n) > y$. Thus, $2 \le m < x/y$. For large x, we have $P > y > z$, so $P \nmid m$. Fix m. Then

$$\beta(n) = P + \beta(m) = P_1\,,$$

where P_1 is also a prime. By the sieve (apply Theorem 12.6 in [5] with $w(p) = 2$ if $p \nmid \beta(m)$ and $w(p) = 1$ otherwise), the number of primes $P \le x/m$ with the property that $P + \beta(m)$ is a prime is

$$\begin{aligned} &\ll \frac{x}{m} \prod_{3 \le p \le x/m} \left(1 - \frac{w(p)}{p}\right) \\ &\ll \frac{x}{m} \prod_{3 \le p \le x/m} \left(1 - \frac{2}{p}\right) \prod_{\substack{3 \le p \le x/m \\ p \mid \beta(m)}} \left(\left(1 - \frac{1}{p}\right)\left(1 - \frac{2}{p}\right)^{-1}\right) \\ &\ll \frac{x}{m(\log(x/m))^2} \left(\frac{\sigma(\beta(m))}{\beta(m)}\right), \end{aligned}$$

where we used the fact that

$$\prod_{\substack{3 \le p \le x/m \\ p \mid \beta(m)}} \left(\left(1 - \frac{1}{p}\right)\left(1 - \frac{2}{p}\right)^{-1}\right) \ll \prod_{p \mid \beta(m)} \left(1 + \frac{1}{p}\right) \ll \frac{\sigma(\beta(m))}{\beta(m)}.$$

Now we sum up over m. Note that since $n = Pm \le x$ and $P > P(m)$, it follows that the m's under scrutiny satisfy $mP(m) \le x$. Thus,

$$\#\left(\mathcal{B}(x) \backslash \left(\mathcal{B}_1(x) \cup \mathcal{B}_2(x)\right)\right) \ll \sum_{\substack{m \le x/y \\ mP(m) \le x}} \frac{x}{m(\log(x/m))^2} \left(\frac{\sigma(\beta(m))}{\beta(m)}\right).$$

If most of the sum over the above range of m is concentrated on m's that are not too small (say $\log(x/m) \asymp \log x$), and if in addition the factor $\sigma(\beta(m))/\beta(m)$ is $O(1)$ on the average over such m, then the sum of the reciprocals of these m will introduce at most another logarithm. Hence, we have to deal with $\log(x/m)$ and with $\sigma(\beta(m))/\beta(m)$. We first deal only with $\sigma(\beta(m))/\beta(m)$ when m is not too small as a warm-up example, and deal with the general case later.

2.4. **Eliminating smooth m's.** Let $w := \exp(\log x/(\log\log x)^3)$. Let

$$\mathcal{B}_3(x) := \left\{n \in \mathcal{B}(x) \backslash \left(\mathcal{B}_1(x) \cup \mathcal{B}_2(x)\right) \text{ and } P(m) \le w\right\}. \tag{5}$$

Then

$$\#\mathcal{B}_3(x) \le \sum_{\substack{m \le x/y \\ P(m) \le w}} \frac{x}{m(\log(x/m))^2} \left(\frac{\sigma(\beta(m))}{\beta(m)}\right).$$

Since $x/m \ge y$, we have $\log(x/m) \ge \log y = \log x/\log\log x$. Further, since $\beta(m) \le m \le x$, it follows, by the fact that the inequality $\sigma(n)/n \ll \log\log n$ holds for all positive integers $n \ge 3$ (see Proposition 8.5 in [5]), that $\sigma(\beta(m))/\beta(m) \ll \log\log x$. Hence,

$$\begin{aligned} \#\mathcal{B}_3(x) &\ll \frac{x(\log\log x)^3}{(\log x)^2} \sum_{\substack{m \le x/y \\ P(m) \le w}} \frac{1}{m} \ll \frac{x(\log\log x)^3}{(\log x)^2} \prod_{p \le w} \left(1 - \frac{1}{p}\right)^{-1} \\ &\ll \frac{x(\log\log x)^3 \log w}{(\log x)^2} = \frac{x}{\log x}. \end{aligned}$$

Hence,

$$\#\mathcal{B}_3(x) \ll \frac{x}{\log x}\,. \tag{6}$$

The next trick we use is to note that if d is a divisor of s, so is s/d, therefore

$$\frac{\sigma(s)}{s} = \sum_{d|s}\frac{1}{d} \le \sum_{\substack{d|s\\ d\le\sqrt{s}}}\left(\frac{1}{d}+\frac{1}{s/d}\right) \le 2\sum_{\substack{d|s\\ d\le\sqrt{s}}}\frac{1}{d}\,.$$

We apply this with $s=\beta(m)$ and change the order of summation getting that

$$\#\left(\mathcal{B}(x)\backslash\left(\mathcal{B}_1(x)\cup\mathcal{B}_2(x)\cup\mathcal{B}_3(x)\right)\right) \ll x\sum_{d\le\sqrt{x/y}}\frac{1}{d}\sum_{\substack{m\in\mathcal{M}\\ \beta(m)\ge d^2\\ \beta(m)\equiv 0 \pmod d}}\frac{1}{m(\log(x/m))^2}\,,$$

where

$$\mathcal{M} := \{m \le \min\{x/y, x/P(m)\} : P(m) > w,\ p^2 \nmid m \text{ for } p > z\}\,.$$

2.5. **The case $m \le x^{9/10}$.** We write $\mathcal{M}_1 := \mathcal{M}\cap[1,x^{9/10}]$ and

$$\mathcal{B}_4(x) := \{n\in\mathcal{B}(x)\backslash\left(\mathcal{B}_1(x)\cup\mathcal{B}_2(x)\cup\mathcal{B}_3(x)\right) : m\in\mathcal{M}_1\}\,. \tag{7}$$

When $m\in\mathcal{M}_1$, we have $x/m \ge x^{1/10}$, so $\log(x/m)\ge\log(x^{1/10})\gg\log x$. Thus,

$$\#\mathcal{B}_4(x) \ll \frac{x}{(\log x)^2}\sum_{d\le\sqrt{x/y}}\frac{1}{d}\sum_{\substack{m\in\mathcal{M}_1\\ \beta(m)\ge d^2\\ \beta(m)\equiv 0 \pmod d}}\frac{1}{m} := \frac{x}{(\log x)^2}\sum_{d\le\sqrt{x/y}}\frac{S_d}{d}\,. \tag{8}$$

We write $m =: Q\ell$, where $Q := P(m)$. Since $Q > w$ and $w > z$ for large x, it follows that for large x, $P(\ell) < Q$. Now $\beta(m) = Q + \beta(\ell) \ge Q$. Further, since $\omega(m) \ll \log m/\log\log m \le \log x/\log\log x$ holds for all our m (see Proposition 7.10 in [5]), and $Q > z$, it follows that

$$\beta(m) \ll Q(\log x/\log\log x), \quad \text{therefore} \quad \beta(m) < Q\log x = Qz^{1/10} < Q^{1.1}$$

holds for all large x. Since $d^2 \le \beta(m)$, it follows that $Q > d^{2/1.1} > d^{1.8}$. So, we fix d and ℓ. Then $\beta(m) = Q + \beta(\ell) \equiv 0 \pmod d$ puts Q into the arithmetic progression $-\beta(\ell)$ modulo d which depends on ℓ.

Also, $Q\in[\max\{P(\ell), d^{1.8}\}, x]$. By the sieve (see Theorem 12.7 in [5]), the counting function of primes $Q \le t$ in this arithmetic progression is

$$\pi(t,d,-\beta(\ell)) \ll \frac{t}{\phi(d)\log(t/d))} \ll \frac{t}{\phi(d)\log t} \tag{9}$$

for $t\in[\max\{P(\ell), d^{1.8}\}, x]$.

Fixing ℓ and summing up over all m with the fixed ℓ and the prime $P(m)=Q$ in the above progression modulo d, we get, by the Abel summation formula,

$$\begin{aligned}
S_{d,\ell} &\ll \frac{1}{\ell} \sum_{\substack{\max\{P(\ell), d^{1/8}\} \le Q \le x \\ Q \equiv -\beta(\ell) \pmod d}} \frac{1}{Q} \\
&\ll \frac{1}{\ell}\left(\frac{\pi(x,d,-\beta(\ell))}{x} + \int_{P(\ell)}^{x} \frac{\pi(t,d,-\beta(\ell))}{t^2}dt\right) \\
&\ll \frac{1}{\phi(d)\ell}\left(\frac{1}{\log x} + \int_{P(\ell)}^{x} \frac{dt}{t\log t}\right) \\
&\ll \frac{1}{\phi(d)\ell}\left(\frac{1}{\log x} + (\log\log x - \log\log P(\ell))\right). \qquad (10)
\end{aligned}$$

Inserting (10) into the right-hand side of (8) and summing up the first terms over $\ell \le m \le x^{0.9}$ and over $d \le x$, we get a contribution of at most

$$\frac{x}{(\log x)^3} \sum_{\ell \le x^{0.9}} \frac{1}{\ell} \sum_{d \ge 1} \frac{1}{d\phi(d)} \ll \frac{x}{(\log x)^2}$$

integers $n \in \mathcal{B}_4(x)$, where we used the fact that the last series is convergent because $\phi(d) \gg d/\log\log d$ (see Proposition 8.4 in [5]), therefore we have $d\phi(d) \gg d^2/\log\log d \gg d^{3/2}$. Summing up also the second terms in (10) over ℓ then over $d \le x$, we get

$$\begin{aligned}
\#\mathcal{B}_4(x) &\ll \frac{x}{(\log x)^2} \sum_{3 \le \ell \le x^{0.9}} \left(\frac{\log\log x - \log\log P(\ell)}{\ell}\right) \sum_{d \ge 1} \frac{1}{d\phi(d)} + \frac{x}{(\log x)^2} \\
&\ll \frac{x}{(\log x)^2} \sum_{3 \le \ell \le x^{0.9}} \frac{\log\log x - \log\log P(\ell)}{\ell}.
\end{aligned}$$

In the above, we discarded the cases $\ell = 1, 2$ since then $m = PQ,\ 2PQ$ with P and Q large primes, because for these ones $\beta(m) = P + Q,\ P + Q + 2$ is large and even so it cannot be a prime. We also absorbed the second term $x/(\log x)^2$ into the first term (say for $\ell = 3$). To continue, we may assume that $P(\ell) > y$. Indeed, the part of the above sum for with $P(\ell) \le y$ gives a contribution of

$$\frac{x\log\log x}{(\log x)^2} \sum_{P(\ell) \le y} \frac{1}{\ell} \ll \frac{x(\log\log x)\log y}{(\log x)^2} \ll \frac{x}{\log x}$$

to $\#\mathcal{B}_4(x)$, which is acceptable for us. For the rest of $\#\mathcal{B}_4(x)$, we put $P(\ell) \in [x^{1/(j+1)}, x^{1/j}]$ for some integer $j \in [1, \log\log x]$. For such ℓ, we have

$$\log\log P(\ell) = \log\log x + O\big(\log(j+1)\big),$$

so

$$\log\log x - \log\log P(\ell) = O\big(\log(j+1)\big).$$

For such an ℓ, we write $\ell =: R\ell_1$, where $R := P(\ell)$. We then get

$$\begin{aligned}
\#\mathcal{B}_4(x) &\ll \frac{x}{(\log x)^2} \sum_{3\le \ell\le x^{0.9}} \frac{\log\log x - \log\log P(\ell)}{\ell} + \frac{x}{\log x}\\
&\ll \frac{x}{\log x} + \frac{x}{(\log x)^2} \sum_{1\le j\le \log\log x} \sum_{\substack{3\le \ell\le x^{0.9}\\ x^{1/(j+1)}<P(\ell)\le x^{1/j}}} \frac{\log(j+1)}{\ell}\\
&\ll \frac{x}{\log x} + \frac{x}{(\log x)^2} \sum_{1\le j\le \log\log x} \log(j+1) \sum_{x^{1/j+1}<R\le x^{1/j}} \frac{1}{R} \sum_{P(\ell_1)\le x^{1/j}} \frac{1}{\ell_1}\\
&\ll \frac{x}{\log x} + \frac{x}{(\log x)^2}\\
&\quad \times \sum_{1\le j\le \log\log x} \log(j+1) \left(\log\log x^{1/j} - \log\log x^{1/j+1} + O\left(\frac{j}{\log x}\right)\right)\\
&\quad \times \left(\log x^{1/j} + O(1)\right)\\
&\ll \frac{x}{\log x} + \frac{x}{(\log x)^2} \sum_{1\le j\le \log\log x} \log(j+1) \left(\log\left(1+\frac{1}{j}\right) + O\left(\frac{j}{\log x}\right)\right)\\
&\quad \times \left(\frac{\log x}{j} + O(1)\right)\\
&\ll \frac{x}{\log x} + \frac{x(\log x)}{(\log x)^2} \sum_{j\ge 1} \frac{\log(j+1)}{j^2} \ll \frac{x}{\log x}\,, \qquad (11)
\end{aligned}$$

which is what we wanted. In the above, we used the fact that the series

$$\sum_{j\ge 1} \frac{\log(j+1)}{j^2}$$

is convergent. We also used the fact that

$$\sum_{a\le R\le b} \frac{1}{R} = \log\log b - \log\log a + O(1/\log a) \qquad (12)$$

with $a = x^{1/(j+1)}$ and $b = x^{1/j}$. With these choices, we have that

$$\log\log b - \log\log a = \log(1+1/j) = O(1/j)$$

and this last quantity dominates $O(1/\log a) = O(j/\log x)$ since $j \le \log\log x$.

2.6. **The case m large.** Here, we put

$$\mathcal{B}_5(x) := \mathcal{B}(x)\backslash \left(\cup_{i=1}^{4}\mathcal{B}_i(x)\right)\,. \qquad (13)$$

We put $\mathcal{M}_2 = \mathcal{M}\backslash\mathcal{M}_1$ and then, by the argument leading to (8), we have

$$\#\mathcal{B}_5(x) \ll x \sum_{d\le\sqrt{x/y}} \frac{1}{d} \sum_{\substack{m\in\mathcal{M}_2\\ \beta(m)\ge d^2\\ \beta(m)\equiv 0 \pmod d}} \frac{1}{m(\log(x/m))^2} := x \sum_{d\le\sqrt{x/y}} \frac{T_d}{d}\,. \qquad (14)$$

Every m participating in T_d satisfies $m > x^{9/10}$. We split the sum over m according to $x/m \in [x^{1/k+1}, x^{1/k})$. For such m, we have that the inequality $\log(x/m) \geq (\log x)/(k+1)$ holds. Further, $m \in (x^{1-1/k}, x^{1-1/(k+1)}]$, and since $m > x^{9/10}$, it follows that $k \geq 10$. Further,

$$w < P(m) < P < x/m \leq x^{1/k},$$

therefore $k \leq (\log\log x)^3$. Hence,

$$\begin{aligned}
T_d &\leq \frac{1}{(\log x)^2} \sum_{10 \leq k \leq (\log\log x)^3} (k+1)^2 \sum_{\substack{x^{1-1/k} \leq m \leq x^{1-1/(k+1)} \\ P(m) < x^{1/k} \\ \beta(m) \geq d^2 \\ \beta(m) \equiv 0 \pmod d}} \frac{1}{m} \\
&:= \frac{1}{(\log x)^2} \sum_{10 \leq k \leq (\log\log x)^3} (k+1)^2 T_{d,k}\,.
\end{aligned}$$

For each inner sum, we apply the argument used when bounding $\#\mathcal{B}_4(x)$. Namely, $\beta(m) = Q + \beta(\ell) \equiv 0 \pmod d$, which puts, for a fixed ℓ, the prime $Q = P(m)$ into the arithmetic progression $-\beta(\ell)$ modulo d. Further, $Q \geq \max\{P(\ell), d^{1.8}\}$ so the bound (9) applies. Fixing ℓ and summing up over all Q, we get, by the argument used at (10), that

$$\begin{aligned}
T_{d,k,\ell} &\ll \frac{1}{\ell} \sum_{\substack{\max\{P(\ell), d^{1/8}\} \leq Q \leq x \\ Q \equiv -\beta(\ell) \pmod d}} \frac{1}{Q} \\
&\ll \frac{1}{\ell}\left(\frac{\pi(x,d,-\beta(\ell))}{x} + \int_{P(\ell)}^{x} \frac{\pi(t,d,-\beta(\ell))}{t^2}\,dt\right) \\
&\ll \frac{1}{\phi(d)\ell}\left(\frac{1}{\log x} + \int_{P(\ell)}^{x} \frac{dt}{t\log t}\right) \\
&\ll \frac{1}{\phi(d)\ell}\left(\frac{1}{\log x} + (\log\log x - \log\log P(\ell))\right) \\
&\ll \frac{\log\log x - \log\log P(\ell)}{\phi(d)\ell}\,. \qquad (15)
\end{aligned}$$

For the last estimate above, we used the fact that $P(\ell) \leq x^{1/k} \leq x^{1/10}$, so $\log\log x - \log\log P(\ell) \geq \log\log x - \log\log x^{1/10} \gg 1$ dominates $1/\log x$. Hence, we can write

$$T_d \ll \frac{1}{\phi(d)(\log x)^2} \sum_{10 \leq k < (\log\log x)^3} (k+1)^2 \sum_{\ell \in \mathcal{L}_k} \frac{\log\log x - \log\log P(\ell)}{\ell},$$

where $\mathcal{L}_k$ is the set of positive integers m such that $m = Q\ell$ for some prime $Q > P(\ell)$, and m satisfies the conditions that $m \in (x^{1-1/k}, x^{1-1/(k+1)}]$,

$\beta(m) \equiv 0 \pmod{d^2}$. Therefore,

$$\#\mathcal{B}_5(x) \ll \frac{x}{(\log x)^2} \sum_{10 \le k \le (\log\log x)^3} (k+1)^2 \left(\sum_{\ell \in \mathcal{L}_k} \frac{\log\log x - \log\log P(\ell)}{\ell} \right) \left(\sum_{d \ge 1} \frac{1}{d\phi(d)} \right).$$

The last sum is $O(1)$. We may assume that $P(\ell) \ge w_1 := \exp(\log x/(\log\log x)^{10})$, since for $P(\ell) \le w_1$, we have

$$\begin{aligned}
&\frac{x}{(\log x)^2} \sum_{10 \le k \le (\log\log x)^3} (k+1)^2 \left(\sum_{P(\ell) \le w_1} \frac{\log\log x - \log\log P(\ell)}{\ell} \right) \\
&\ll \frac{x(\log\log x)^6}{(\log x)^2} \sum_{10 \le k \le (\log\log x)^3} \sum_{P(\ell) \le w_1} \frac{\log\log x}{\ell} \\
&\ll \frac{x(\log\log x)^{10}}{(\log x)^2} \sum_{P(\ell) \le w_1} \frac{1}{\ell} \\
&\ll \frac{x(\log\log x)^{10}\log w_1}{(\log x)^2} \ll \frac{x}{\log x}.
\end{aligned}$$

So, we have

$$\begin{aligned}
\#\mathcal{B}_5(x) \ll{}& \frac{x}{\log x} \\
&+ \frac{x}{(\log x)^2} \sum_{10 \le k \le (\log\log x)^3} (k+1)^2 \left(\sum_{\substack{\ell \in \mathcal{L}_k \\ P(\ell) \ge w_1}} \frac{\log\log x - \log\log P(\ell)}{\ell} \right),
\end{aligned}$$

and we may concentrate on the case $P(\ell) \ge w_1$. Note that since $Q \le x^{1/k}$, it follows that $\ell = m/Q \ge x^{1-2/k}$. We now let $j \ge 0$ and assume that $P(\ell) \in [x^{1/(k+j+1)}, x^{1/(k+j)}]$. Since $P(\ell) \ge w_1$, it follows that the inequality $j < j+k \le (\log\log x)^{10}$ holds. Write $\ell =: R\ell_1$, where $R := P(\ell)$. Then $\ell_1 = \ell/R \ge x^{1-2/k-1/(k+j)} \ge x^{1-3/k} \ge x^{1/3}$ is not too small. We easily verify this by noting that

$$\frac{2}{k} + \frac{1}{k+j} = \frac{3k+j}{k(k+j)} \le \frac{3(k+j)}{k(k+j)} = \frac{3}{k} \le \frac{1}{3},$$

since $k \ge 10$ and $j \ge 0$. Since $P(\ell_1) \le P(\ell)$, it follows that such numbers ℓ_1 are quite smooth, namely their $u = \log\ell_1/\log P(\ell_1)$ is at least

$$u = \frac{\log \ell_1}{\log P(\ell_1)} \ge \frac{\log(x^{1/3})}{\log x^{1/(k+j)}} = \frac{k+j}{3}.$$

For such numbers ℓ, we also have $\log\log x - \log\log P(\ell) = O(\log(k+j))$. We thus get that

$$\#\mathcal{B}_5(x) \ll \frac{x}{\log x} + \frac{x}{(\log x)^2} \sum_{10 \le k < (\log\log x)^3} (k+1)^2 \sum_{j \ge 0} \log(k+j) \sum_{x^{1/(k+j+1)} \le R \le x^{1/(k+j)}} \frac{1}{R} \sum_{\substack{\ell_1 \in [x^{1/3}, x] \\ u \ge (k+j)/3}} \frac{1}{\ell_1}.$$

The last inner sum is

$$\ll \rho((k+j)/3) \log x\,,$$

by applying the Abel summation formula for the sum of the reciprocals of the $x^{1/(k+j)}$-smooth numbers in $[x^{1/3}, x]$ as described in Section 2.1. The sum prior to it is, by estimate (12) with $a := x^{1/(k+j+1)}$ and $b := x^{1/(k+j)}$,

$$\begin{aligned} &\log\log(x^{1/(k+j)} - \log\log x^{1/(k+j+1)} + O\left(\frac{1}{\log x^{1/(k+j)}}\right) \\ &= \log\left(1 + \frac{1}{k+j}\right) + O\left(\frac{k+j}{\log x}\right) \ll \frac{1}{k+j} + O\left(\frac{k+j}{\log x}\right) \ll \frac{1}{k+j}\,, \end{aligned}$$

where we used the fact that $k + j = O((\log\log x)^{10})$. Thus,

$$\#\mathcal{B}_5(x) \ll \frac{x}{\log x} + \frac{x}{\log x} \sum_{10 \le k \le (\log\log x)^3} (k+1)^2 \sum_{0 \le j \le (\log\log x)^{10}} \frac{\rho((k+j)/3)\log(k+j)}{k+j}.$$

Since $\rho(u) = u^{-(1+o(1))u}$ as $u \to \infty$, it follows that $\rho((k+j)/3) \ll (k+j)^{-4}$. Hence, the inner sum above is bounded by

$$\begin{aligned} \sum_{0 \le j \le (\log\log x)^{10}} \frac{\rho((k+j)/3)\log(k+j)}{k+j} &\ll \sum_{j \ge 0} \frac{\log(k+j)}{(k+j)^5} \le \sum_{K \ge k} \frac{\log K}{K^5} \\ &\ll \int_k^\infty \frac{\log t dt}{t^5} \ll \frac{\log k}{k^4}\,, \end{aligned}$$

and so

(16)

$$\#\mathcal{B}_5(x) \ll \frac{x}{\log x} + \frac{x}{\log x} \sum_{10 \le k \le (\log\log x)^3} \frac{(k+1)^2 \log k}{k^4} \ll \frac{x}{\log x} \sum_{k \ge 9} \frac{\log k}{k^2} \ll \frac{x}{\log x}.$$

Thus, by inequalities (2), (4), (6), (11) and (16), we get that

$$\#\mathcal{B}(x) \le \sum_{i=1}^{5} \#\mathcal{B}_i(x) \ll \frac{x}{\log x}\,,$$

which is what we wanted to prove.

3. The lower bound

3.1. **A conditional lower bound.** Before we give the proof of the actual unconditional lower bound, let us give a conditional proof of a lower bound of the same order of magnitude as the upper bound. Before we can do that, let us recall the Bateman-Horn conjecture. We say that a system of non-constant polynomials $f_1(X),\dots,f_k(X)\in\mathbb{Z}[X]$ each having positive leading coefficient is *acceptable* if the following conditions hold:

(i) $f_i(X)$ is irreducible for all $i=1,\dots,k$.

(ii) there exists no prime number p such that $p\mid f_1(n)f_2(n)\cdots f_k(n)$ for all $n\ge 0$.

Given a system of acceptable polynomials, the Bateman-Horn conjecture (see Chapter 2.10 in [5]) is an heuristic statement addressing the frequency of the positive integers n such that

$$f_1(n),\ f_2(n),\ \dots,\ f_k(n)$$

are all prime numbers. The actual statement is as follows.

Conjecture 1 (Bateman, Horn [1]). *Let $f_1(X),\dots,f_k(X)$ be an acceptable system of polynomials. For each prime number p let*

$$\omega(p)=\#\{0\le n\le p-1\ :\ f_1(n)f_2(n)\dots f_k(n)\equiv 0 \pmod p\}\,,$$

and set

$$\pi_{f_1,\dots,f_k}(x):=\#\{n\le x\ :\ f_1(n),\ \dots,\ f_k(n)\ \text{are all primes}\}\,.$$

Then

$$\pi_{f_1,\dots,f_k}(x)=(1+o(1))C(f_1,\dots,f_k)\,\frac{1}{d_1\dots d_k}\,\frac{x}{(\log x)^k}\quad \text{as}\quad x\to\infty\,, \tag{17}$$

where $d_i:=\deg(f_i)$, and where the constant $C(f_1,\dots,f_k)$ is given by

$$C(f_1,\dots,f_k)=\prod_{p\ge 2}\frac{1-\omega(p)/p}{(1-1/p)^k}\,. \tag{18}$$

Proposition 1. *Assume that Bateman-Horn conjecture estimate* (17) *holds for the acceptable system of polynomials $X,\ X+h$ for all even positive integers h in the range $h<x^{\delta}$ with some small $\delta>0$. Then*

$$\#\mathcal{B}(x)\gg\frac{x}{\log x}\,.$$

Proof. For all even positive integers h, let $\pi_2(x,h)$ be the number of primes p such that $p+h\le x$ is also prime. The Bateman-Horn conjecture for the acceptable system of polynomials $X,\ X+h$ implies that

$$\pi_2(x,h)\asymp c_h\frac{x}{(\log x)^2},\quad \text{where}\quad c_h=2\left(\prod_{p\ne 2}\frac{p(p-2)}{(p-1)^2}\right)\left(\prod_{\substack{p\mid h\\ p>2}}\frac{p-1}{p-2}\right) \tag{19}$$

as $x \to \infty$. We assume further that $\pi_2(h) \gg c_h x/(\log x)^2$ holds uniformly for all even $h \le x^\delta$ for some small $\delta > 0$. We may assume that $\delta < 1/3$. Consider numbers $m \le x^{\delta/2}$ which are odd and for which $\omega(m)$ is even. Then $\beta(m)$ is even and $\beta(m) \le m \le x^{\delta/2} < (x/m)^\delta$. Thus, by the above assumption, with $h = \beta(m)$, we have

$$\pi_2(x/m, h) \gg c_h \frac{x}{m(\log(x/m))^2} \gg \frac{x}{m(\log x)^2}\,.$$

We may assume that $p > x^{1-\delta}$, since if not, then for a given m there are at most $\pi(x^{1-\delta}) < x^{1-\delta} < x/(m(\log x)^3)$ primes p failing the above inequality, so we can eliminate those situations from the above count and still keep $\gg x/(m(\log x)^2)$ primes p. We now sum up over $m \le x^{\delta/2}$ getting a count of

$$\gg \frac{x}{(\log x)^2} \sum_{\substack{m \le x^{\delta/2} \\ m \equiv 1 \pmod 2 \\ \omega(m) \equiv 0 \pmod 2}} \frac{1}{m}$$

pairs (m, p) with $m \le x^{\delta/2}$, $p \le x/m$ and $p + \beta(m)$ is prime. Let us now put $n = pm \le x$. Further, $p > x^{1-\delta} > m$, so $p = P(n)$ is uniquely determined out of n. Thus, this construction produces distinct integers $n \in \mathcal{B}(x)$. It is enough to show that the sum over m's is a positive proportion of $\log x$. Well, assume further $m \le x^{\delta/2}/3$ and that m is coprime to 6 and then

$$\sum_{\substack{m \le x^{\delta/2}/3 \\ \gcd(m,6)=1}} \frac{1}{m} \ge \sum_{\ell \le \lfloor x^\delta/18 \rfloor - 1} \frac{1}{6\ell+1} \gg \sum_{\ell \le \lfloor x^\delta/18 \rfloor - 1} \frac{1}{\ell} \gg \log\left(\frac{x^{\delta/2}}{18} - 1\right) \gg \log x\,.$$

Write the sum in the left as $S_0 + S_1$, where S_i denotes the contribution to the sum of m with $\omega(m) \equiv i \pmod 2$ for $i = 0, 1$. If $S_0 \ge S_1$, then $S_0 \ge (S_0 + S_1)/2 \gg \log x$, and we are through. Otherwise, $S_1 \gg \log x$ and

$$\sum_{\substack{m \le x^{\delta/2} \\ m \equiv 1 \pmod 2 \\ \omega(m) \equiv 0 \pmod 2}} \frac{1}{m} \ge \sum_{\substack{m \le x^{\delta/2}/3 \\ \gcd(m,6)=1 \\ \omega(m) \equiv 1 \pmod 2}} \frac{1}{3m} \ge \frac{S_1}{3} \gg \log x\,,$$

so we are done again. This finishes the proof of Proposition 1. $\square$

3.2. **The unconditional lower bound.** Here, we give a proof of the lower bound from Theorem 1. We use an average version of the Bateman-Horn conjecture due to Chudakov [2]. To state it, let again h be an even positive integer and let

$$T_h(x) := \sum_{n+h \le x} \frac{1}{\log n \log(n+h)}\,.$$

Note that $T_h(x) \asymp x/(\log x)^2$ uniformly for $h = o(x)$ as $x \to \infty$. Then Chudakov proved that

$$\sum_{2|h} |\pi_2(x,h) - c_h T_h(x)|^2 \ll_A \frac{x^3}{(\log x)^A}$$

for any constant $A > 0$. Here, c_h is given by (19). Note that the sum on the left is finite since when $h > x$ from the definition of $T_h(x)$ and if

$$T_h(x) \asymp \frac{x}{(\log x)^2} \gg \pi_2(x,h)$$

both $\pi_2(x,h)$ and $T_h(x)$ are zero. To apply Chudakov's theorem, we take $A := 10$ and let

$$\mathcal{H} := \{h < x/20; h \equiv 0 \pmod 2,\ \pi_2(x,h) < (c_h/2)T_h(x)\}\,.$$

Then

$$\sum_{h\in\mathcal{H}} (c_h T_h(x))^2 \ll \frac{x^3}{(\log x)^{10}}\,.$$

Since $T_h(x) \gg x/(\log x)^2$ for $h \le x/20$ and $c_h \gg 1$, it follows that the left–hand side of the above sum is

$$\gg \#\mathcal{H}\left(\frac{x}{(\log x)^2}\right)^2 \gg \#\mathcal{H}\frac{x^2}{(\log x)^4}.$$

Thus,

$$\#\mathcal{H}\frac{x^2}{(\log x)^4} \ll \frac{x^3}{(\log x)^{10}}\,,$$

implying that $\#\mathcal{H} \ll x/(\log x)^6$. Thus, the set of even integers $h \le x/20$ such that $\pi_2(x,h) \gg c_h T_h(x)$ is of cardinality $x/20 - O(x/(\log x)^6)$. In particular, this set contains most numbers of the form $h = Q + 3$, where $Q \in (x/40, x/20]$ is a prime. Consider pairs of numbers of the form (Q,p), where Q is prime, $h = Q+3 \in (x/40,x/20]\backslash\mathcal{H}$ and p is a prime counted by $\pi_2(x,h)$. The number of such pairs is

$$\begin{aligned}
&\ge \sum_{\substack{x/40\le Q+3\le x/20\\ Q+3\notin\mathcal{H}}} \pi_2(x,Q+3)\\
&\gg \sum_{\substack{x/40\le Q+3\le x/20\\ Q+3\notin\mathcal{H}}} c_{(Q+3)}T_{Q+3}(x)\\
&\gg \sum_{\substack{x/40\le Q+3\le x/20\\ Q+3\notin\mathcal{H}}} \frac{x}{(\log x)^2}\\
&\gg \frac{x}{(\log x)^2}\left(\pi\left(\frac{x}{20}\right) - \pi\left(\frac{x}{40}\right) - O\left(\frac{x}{(\log x)^6}\right)\right) \gg \frac{x^2}{(\log x)^3}\,.
\end{aligned}$$

For each one of these pairs, $n = 3pQ \le x^2$ satisfies that

$$\beta(n) = p + Q + 3 = p + h \quad \text{is prime}\,.$$

Thus, $n \in \mathcal{B}(x^2)$. Further, since P and Q are large, it follows that each n can appear at most twice in the above count (once from $\pi_2(x,Q+3)$ and once from $\pi_2(x,p+3)$). Thus, the number of such distinct $n \le x^2$ is $\gg x^2/(\log x)^3 \gg x^2/(\log(x^2))^3$. Replacing x by $\sqrt{x}$ in the above argument we get the desired lower bound. This finishes the proof of the lower bound in Theorem 1.

3.3. **A further conjecture on the count of $\mathcal{B}(x)$.** In this section, we offer the following conjecture.

Conjecture 2. *We have*

$$\#\mathcal{B}(x) = e^{\gamma}(1+o(1))\frac{x}{\log x} \qquad as \qquad x \to \infty\,.$$

In the previous sections, we already proved that $\#\mathcal{B}(x) \ll x/\log x$ unconditionally, and $\#\mathcal{B}(x) \gg x/\log x$ conditionally under the Bateman-Horn conjecture for all pairs of polynomials X, $X+h$ for even positive integers $h \le x^{\delta}$ for some $\delta > 0$. So, it is natural to conjecture that

$$\#\mathcal{B}(x) = (1+o(1))cx/\log x \quad \text{holds as} \quad x \to \infty \tag{20}$$

with some constant $c > 0$ and it remains to offer some guess for c. Well, assume that (20) holds. We shall estimate the sum

$$\sum_{n\in\mathcal{B}(x)} \frac{1}{n} \tag{21}$$

in two ways, as follows. In the first way, by estimate (20) and the Abel summation formula, expression (21) should be asymptotically $c(1+o(1))\log\log x$ as $x \to \infty$. Secondly, let us assume that $\beta(n)$ is "randomly distributed" and as such the probability of it to be prime is $1/\log\beta(n)$. Then the sum (21) should be

$$\sum_{2\le n\le x} \frac{1}{n\log\beta(n)}\,. \tag{22}$$

We have not seen the above sum evaluated in the literature, but we have seen the one for which $\beta(n)$ is replaced by $P(n)$. So, let compare them. We have $P(n) \le \beta(n) \le P(n)\omega(n)$. Let $s(n) := \log\beta(n) - \log P(n)$. If p_k denotes the kth prime, then $p_k > k$, so $P(n) \ge p_{\omega(n)} > \omega(n)$ holds for all $n \ge 2$. Thus, $s(n) \le \log\omega(n) \le \log P(n)$. Hence,

$$\begin{aligned}\frac{1}{\beta(n)} &= \frac{1}{\log P(n)+s(n)} = \frac{1}{\log P(n)}\left(\frac{1}{1+s(n)/\log P(n)}\right)\\ &= \frac{1}{\log P(n)} + O\left(\frac{s(n)}{(\log P(n))^2}\right) = \frac{1}{\log P(n)} + O\left(\frac{\log\omega(n)}{(\log P(n))^2}\right)\,.\end{aligned}$$

In the above, we used the fact that

$$\frac{1}{1+z} = 1 + O(z) \quad \text{for} \quad |z| < 1\,,$$

with $z := s(n)/\log P(n)$. Thus,

$$\sum_{2\le n\le x} \frac{1}{n\log\beta(n)} = \sum_{2\le n\le x} \frac{1}{n\log P(n)} + O\left(\sum_{n\le x} \frac{\log\omega(n)}{n(\log P(n))^2}\right)\,.$$

The first sum above was evaluated in [6]:

$$\sum_{2\le n\le x} \frac{1}{n\log P(n)} = e^{\gamma}\log\log x + O(1)\,.$$

It remains to bound the sum inside the O term. For this, we break it into two parts, namely the part for which $\omega(n)$ is large and the part for which $\omega(n)$ is small. Let

$$\mathcal{A} := \{n : \omega(n) \geq 10 \log\log x\}\,.$$

Recall that $\mathcal{A}(t) = \mathcal{A} \cap [1, t]$. Put $\tau(n)$ for the number of divisors of n. Since $\tau(n) \geq 2^{\omega(n)} \geq 2^{10\log\log x} > (\log x)^6$, we have that

$$\#\mathcal{A}(t)(\log x)^6 \leq \sum_{n\in\mathcal{A}(t)} \tau(n) \ll t\log t \ll t\log x\,.$$

The last inequality above is classical (see Theorem 4.9 n [5]). Thus,

$$\#\mathcal{A}(t) \ll \frac{t}{(\log x)^5} \quad \text{holds for all} \quad t \leq x\,.$$

By the Abel summation formula, we have that

$$\sum_{\substack{n\leq x\\ n\in\mathcal{A}}} \frac{1}{n} \leq \left.\frac{\#\mathcal{A}(t)}{t}\right|_{t=2}^{t=x} + \int_2^x \frac{\#\mathcal{A}(t)dt}{t^2} \ll \frac{1}{(\log x)^5}\left(1 + \int_2^x \frac{dt}{t}\right) \ll \frac{1}{(\log x)^4}\,.$$

Since $\omega(n) \leq \log n/\log\log n$, an estimate mentioned already during the proof of the upper bound for $\#\mathcal{B}_4(x)$, we have that $\log\omega(n) \leq \log\log x$ holds for all $n \leq x$ once x is sufficiently large. Thus,

$$\sum_{\substack{n\leq x\\ n\in\mathcal{A}}} \frac{\log\omega(n)}{n(\log P(n))^2} \ll (\log\log x)\left(\sum_{\substack{n\leq x\\ n\in\mathcal{A}}} \frac{1}{n}\right) \ll \frac{\log\log x}{(\log x)^4} = o(1)$$

as $x \to \infty$. Assume now that $n \leq x$ is not in $\mathcal{A}$. We then have that $\omega(n) \leq 10\log\log x$, so $\log\omega(n) \ll \log\log\log x$. Thus,

$$\sum_{\substack{2\leq n\leq x\\ n\notin\mathcal{A}}} \frac{\log\omega(n)}{n(\log P(n))^2} \ll (\log\log\log x) \sum_{2\leq n\leq x} \frac{1}{n(\log P(n))^2}\,.$$

The last sum on the right is $O(1)$. This follows easily by the Abel summation formula and the fact that

$$\sum_{2\leq n\leq x} \frac{1}{(\log P(n))^2} = (c_1 + o(1))\frac{x}{(\log x)^2} \qquad \text{as} \qquad x \to \infty\,,$$

which is a result of Wheeler from [10]. Further, $c_1 = \int_0^\infty \rho(t)(t+1)dt$, where $\rho(t)$ is the Dickman function mentioned in the upper bound for $\#\mathcal{B}_1(x)$ (for this, make $u := 1,\ \alpha := -2$ on the last display on page 516 in [10]). It thus follows that

$$\sum_{2\leq n\leq x} \frac{\log\omega(n)}{n(\log P(n))^2} = \sum_{\substack{n\leq x\\ n\in\mathcal{A}}} \frac{\log\omega(n)}{n(\log P(n)^2} + \sum_{\substack{2\leq n\leq x\\ n\notin\mathcal{A}}} \frac{\log\omega(n)}{n(\log P(n))^2} \ll \log\log\log x\,,$$

so

$$\sum_{2\le n\le x}\frac{1}{n\log\beta(n)}=\sum_{n\le x}\frac{1}{\log P(n)}+O\left(\sum_{n\le x}\frac{\log\omega(n)}{n(\log P(n))^2}\right)$$
$$=e^{\gamma}\log\log x+O(\log\log\log x)\,.$$

Since the sums (21) and (22) should asymptotically be the same under the assumption that $\beta(n)$ is "randomly distributed", we conclude that if asymptotic (20) holds, then we must indeed have $c=e^{\gamma}$.

We tested this computationally. We computed $r(x):=\#\mathcal{B}'(x)/\pi(x)$, where $\mathcal{B}'(x)=\{n\le x:\omega(n)\ge 2 \text{ and } \beta(n) \text{ is prime}\}$ for $x:=10^k$, and $k=3,4,5,6,7,8,9$ getting the values from the table below.

k	$x=10^k$	$\#\mathcal{B}'(x)$	$\pi(x)$	$r(x)$
3	10^3	130	168	0.77381
4	10^4	1196	1229	0.973149
5	10^5	11698	9592	1.21956
6	10^6	107315	78498	1.3671
7	10^7	961924	664579	1.44742
8	10^8	8641491	5761455	1.49988
9	10^9	78304633	50847534	1.53999
10	10^{10}	714962670	455052511	1.57117

This sequence $\{r(10^k)\}_{k\ge 3}$ in the right–most column above seems to increase with k and it is anyone's guess whether by continuing the calculations one would get a limit of $e^{\gamma}=1.78107\ldots$

Acknowledgement. We thank the referee for useful suggestions. The first author was supported in part by grant CPRR160325161141 from the NRF of South Africa, the Focus Area Number Theory grant RTNUM19 from CoEMaSS Wits, and by grant no. 17-02804S of the Czech Science Foundation.

References

[1] Bateman, P.T., Horn, R.A., *A heuristic asymptotic formula concerning the distribution of prime numbers*, Math. Comp. **16** (1962), 363–367.

[2] Chudakov, N., *On Goldbach-Vinogradov's theorem*, Ann. of Math. (2) **48** (1947), 515–545.

[3] De Koninck, J.-M., Luca, F., *Integers divisible by the sum of their prime factors*, Mathematika **52** (2005), 69–77.

[4] De Koninck, J.-M., Luca, F., *Integers divisible by sums of powers of their prime factors*, J. Number Theory **128** (2008), 557–563.

[5] De Koninck, J.-M., Luca, F., *Analytic number theory. Exploring the anatomy of integers*, Graduate Studies in Mathematics, vol. 134, American Mathematical Society, Providence, RI, 2012.

[6] De Koninck, J.-M., Sitaramachandra, R.R., *Sums involving the largest prime divisor of an integer*, Acta Arith. **48** (1987), 3–8.

[7] Erdös, P., Pomerance, C., *On the largest prime factors of n and $n+1$*, Aequationes Math. **17** (1978), 311–321.

[8] Nelson, C., Penny, D.E., Pomerance, C., *714 and 715*, J. Recreational Math. **7** (1974), 87–89.

[9] Pomerance, C., *Ruth–Aaron numbers revisited*, Paul Erdös and his Mathematics, vol. 11, János Bolyai Math. Soc., Budapest 1999, Bolyai Soc. Math. Stud., 2002, pp. 567–579.

[10] Wheeler, F.S., *Two differential-difference equations arising in number theory*, Trans. Amer. Math. Soc. **318** (1990), 491–523.

School of Mathematics, University of the Witwatersrand,
1 Jan Smuts Avenue, Braamfontein 2000,
Johannesburg, South Africa,
and
Research Group Algebraic Structures and Applications,
King Abdulaziz University, Jeddah, Saudi Arabia,
and
Max Planck Institute for Mathematics,
Vivatsgasse 7, Bonn 53111, Germany,
and
Department of Mathematics, University of Ostrava,
30. dubna 22, 701 03 Ostrava 1, Czech Republic
E-mail: `florian.luca@wits.ac.za`

School of Mathematics, University of the Witwatersrand,
1 Jan Smuts Avenue, Braamfontein 2000,
Johannesburg, South Africa
E-mail: `747011@students.wits.ac.za`