ON THE NUMBER OF PRIME FACTORS OF VALUES OF THE SUM-OF-PROPER-DIVISORS FUNCTION

LEE TROUPE

Abstract. Let $\omega(n)$ (resp. $\Omega(n)$) denote the number of prime divisors (resp. with multiplicity) of a natural number n. In 1917, Hardy and Ramanujan proved that the normal order of $\omega(n)$ is $\log\log n$, and the same is true of $\Omega(n)$; roughly speaking, a typical natural number n has about $\log\log n$ prime factors. We prove a similar result for $\omega(s(n))$, where $s(n)$ denotes the sum of the proper divisors of n: For any $\epsilon > 0$ and all $n \le x$ not belonging to a set of size $o(x)$,

$$|\omega(s(n)) - \log\log s(n)| < \epsilon \log\log s(n)$$

and the same is true for $\Omega(s(n))$.

1. Introduction

Let $s(n)$ denote the sum of the proper divisors of a positive integer n. The function $s(n)$ has been of interest to number theorists since antiquity; for example, the ancient Greeks wanted to know when $s(n) = n$, calling such integers *perfect*. In modern times, open problems concerning $s(n)$ abound, such as the famous Catalan-Dickson conjecture [Dic13]: For any positive integer n, the aliquot sequence at n (that is, the sequence of iterates of the function s on n) either terminates at 0 or is eventually periodic.

Another conjecture pertaining to $s(n)$ is the following, due to Erdös, Granville, Pomerance and Spiro [EGPS90]:

Conjecture 1.1. *If $\mathcal{A}$ is a set of natural numbers of asymptotic density zero, then $s^{-1}(\mathcal{A})$ also has density zero.*

Recall now Hardy and Ramanujan's normal order result for $\omega(n)$, where $\omega(n)$ denotes the number of distinct prime divisors of n (see [HR17]):

Theorem 1.2. *For any $\epsilon > 0$ and all numbers n not belonging to a set of asymptotic density zero,*

$$|\omega(n) - \log\log n| < \epsilon \log\log n.$$

Then Conjecture 1.1 would imply the following:

Theorem 1.3. *For any $\epsilon > 0$ and all numbers $n \le x$ not belonging to a set of size $o(x)$,*

$$|\omega(s(n)) - \log\log s(n)| < \epsilon \log\log s(n).$$

Though Conjecture 1.1 remains intractable, we are able to prove Theorem 1.3 unconditionally in the present article.

If a number n is composite, then n has a proper divisor greater than $n^{1/2}$; so for all composite $n \in (x^{1/2}, x]$, we have $x^{1/4} \le n^{1/2} \le s(n) \le x^2$. Hence, for all but $o(x)$ numbers n up to x, $\log\log s(n) = \log\log x + O(1)$. Therefore, it suffices to show that, given $\epsilon > 0$,

$$|\omega(s(n)) - \log\log x| < \epsilon \log\log x \qquad (1)$$

for all n except those belonging to a set of size $o(x)$.

2000 *Mathematics Subject Classification.* 11N37; 11N56; 11N64.

Key words and phrases. sum of proper divisors; number of prime divisors; normal order.

Our normal order result follows from the following estimate.

Theorem 1.4. *As $x \to \infty$,*

$$\sum_{n \le x \,:\, n \notin \mathcal{E}(x)} (\omega(s(n)) - \log\log x)^2 = o(x(\log\log x)^2).$$

where $\mathcal{E}(x) \subset \{1, 2, \ldots, \lfloor x \rfloor\}$ is of size $o(x)$.

For large enough x, if there are more than δx numbers up to x for which (1) fails, where $\delta > 0$ is any positive number, then at least $\frac{\delta}{2}x$ of them (say) do not belong to $\mathcal{E}(x)$, whence

$$\sum_{n \le x \,:\, n \notin \mathcal{E}(x)} (\omega(s(n)) - \log\log x)^2 \ge \tfrac{\delta}{2}\epsilon^2 x(\log\log x)^2;$$

this contradicts Theorem 1.4 for sufficiently large x, and Theorem 1.3 follows.

Remark. As one might expect, we obtain a true statement by replacing $\omega(s(n))$ with $\Omega(s(n))$ in the theorem above; this is shown in §5.

Notation. The letters p and q will denote prime variables unless otherwise stated. The symbols $P(n)$ and $P_2(n)$ will denote the largest and second-largest prime factors of n, respectively; if $n = 1$, we set $P(n) = P_2(n) = 1$, and if n is prime, we set $P_2(n) = 1$. For $x > 0$, define $\log_1 x = \max\{1, \log x\}$, and let $\log_k x$ denote the kth iterate of $\log_1$. Sums over $n \notin \mathcal{E}(x)$ carry the implicit condition that $n \le x$. The symbols $\ll$ and $\sim$ have their usual meanings, and we will make frequent use of O and o-notation. Other notation may be introduced as necessary.

2. Preliminaries

First we show that it suffices to consider a truncated version of the sum in question.

Lemma 2.1.

$$\sum_{n \le x} \omega(s(n)) = \left(\sum_{\log_2 x < p \le x^{1/\sqrt{\log_2 x}}} \sum_{\substack{n \le x \\ p \mid s(n)}} 1 \right) + o(x \log_2 x).$$

Proof. We have

$$\sum_{n \le x} \omega(s(n)) = \sum_{n \le x} \sum_{p : p \mid s(n)} 1.$$

If $n \le x$ then $s(n) \le x^2$. The number of primes $p \notin (\log_2 x, x^{1/\sqrt{\log_2 x}}]$ dividing a number $m \le x^2$ is $\le 2\sqrt{\log_2 x} + \pi(\log_2 x) = o(\log_2 x)$, and the lemma follows. □

2.1. The exceptional set.

Define $\mathcal{E}(x) := \{n \le x : \text{at least one of } A, B, C, D, E, F \text{ fails}\}$, where

A. $P(n) > x^{1/\log_3 x}$,
B. $P(n)^2 \nmid n$,
C. $P_2(n) > x^{1/\log_3 x}$,
D. $P_2(n) < xP(n)/2n$,
E. $P_2(n)^2 \nmid n$
F. if a prime q divides $\gcd(n/P(n), \sigma(n/P(n)))$, then $q < \log_2 x$.

Lemma 2.2. *$\#\mathcal{E}(x) = o(x)$.*

Let $\Psi(x, y)$ denote the count of y-smooth numbers up to x; that is, the count of numbers $n \le x$ such that $p \mid n \implies p \le y$. The following upper bound estimate of de Bruijn [dB66, Theorem 2] will be useful in proving the above lemma.

Proposition 2.3. *Let $x \ge y \ge 2$ satisfy $(\log x)^2 \le y \le x$. Whenever $u := \frac{\log x}{\log y} \to \infty$, we have*

$$\Psi(x, y) \le x/u^{u+o(u)}.$$

Proof of Lemma. Let $\mathcal{E}_j(x) = \{n \le x : \text{condition } j \text{ fails and all previous conditions hold}\}$, $j \in \{A, B, C, D, E, F\}$. If $n \in \mathcal{E}_A(x) \cup \mathcal{E}_B(x)$, then either $P(n) \le x^{1/\log_3 x}$ or $P(n) > x^{1/\log_3 x}$ and $P(n)^2 \mid n$. By Proposition 2.3, the number of $n \le x$ for which the former holds is $O(x/(\log_2 x)^4)$. The number of $n \le x$ for which the latter holds is $\ll x \sum_{p > x^{1/\log_3 x}} p^{-2} \ll x \exp(-\log x/\log_3 x)$, and this is also $O(x/(\log_2 x)^4)$.

If $n \in \mathcal{E}_C(x)$, then either (i) $P(n) \le x^{1/\log_4 x}$ or (ii) $P(n) > x^{1/\log_4 x}$ and $P_2(n) \le x^{1/\log_3 x}$. The number of $n \le x$ for which (i) holds is $O(x/(\log_3 x)^{10})$ by Proposition 2.3. For $n \le x$ such that (ii) holds, write $n = mP$, where $P = P(n) > x^{1/\log_4 x}$ and $P(m) \le x^{1/\log_3 x}$. Then $x/m \ge P > x^{1/\log_4 x}$, and the number of such P given m is, by the prime number theorem,

$$\ll \frac{x/m}{\log(x/m)} \le \frac{x \log_4 x}{m \log x}.$$

Now we sum this over m such that $P(m) \le x^{1/\log_3 x}$, obtaining that the number of $n \le x$ for which (ii) holds is

$$\ll \frac{x \log_4 x}{\log x} \sum_{m \,:\, P(m) \le x^{1/\log_3 x}} \frac{1}{m} = \frac{x \log_4 x}{\log x} \prod_{p \le x^{1/\log_3 x}} \left(1 - \frac{1}{p}\right)^{-1}$$
$$\ll \frac{x \log_4 x}{\log x} \left(\frac{\log_3 x}{\log x}\right)^{-1} = \frac{x \log_4 x}{\log_3 x}.$$

Thus $\#\mathcal{E}_C(x) = O(x \log_4 x/\log_3 x)$.

For $\#\mathcal{E}_D(x)$, note first that $xP(n)/2n > P(n)/2$. Thus any n in $\mathcal{E}_D(x)$ has a prime factor in $(P(n)/2, P(n))$, and the number of such integers n with $P(n) > x^{1/\log_3(x)}$ is, by the prime number theorem,

$$\le x \sum_{q > x^{1/\log_3 x}} \sum_{q/2 < q' < q} \frac{1}{qq'} \ll x \sum_{q > x^{1/\log_3 x}} \frac{1}{q \log q} = O\left(\frac{x \log_3 x}{\log x}\right).$$

To bound $\#\mathcal{E}_E(x)$, we consider $n \le x$ with $P_2(n)^2 \mid n$ and $P_2(n) > x^{1/\log_3 x}$. The number of such n is $\ll x \sum_{p > x^{1/\log_3 x}} p^{-2} \ll x \exp(-\log x/\log_3 x)$.

Finally, suppose $q \mid \big(n/P(n), \sigma(n/P(n))\big)$ and $q \ge \log_2 x$. Then $q \mid n$ and $q \mid \sigma(n)$. Write $n = q^e s$, with $q \nmid s$. If $e \ge 2$, then n has a squarefull divisor $\ge (\log_2 x)^2$, and the number of such n is $O(x/\log_2 x)$.

So assume $e = 1$. $q \mid \sigma(n)$ implies $q \mid \sigma(s)$. Write $s = p_1^{e_1} \cdots p_k^{e_k}$; then $q \mid \sigma(p_i^{e_i})$ for some i. If $e_i \ge 2$, then we have $2p_i^{e_i} > \sigma(p_i^{e_i}) \ge q$, so s has a squarefull divisor $\ge q/2$. The number of such $s \le x/q$ is $O(x/q^{3/2})$, and summing this over $q \ge \log_2 x$ gives that the number of possible n is $O(x/\sqrt{\log_2 x})$.

If $e_i = 1$, then $p_i \equiv -1 \pmod q$. By Brun-Titchmarsh and partial summation, the number of $s \le x/q$ divisible by such a prime p_i is

$$\frac{x}{q} \sum_{\substack{p \le x/q \\ p \equiv -1 \pmod q}} \frac{1}{p} \ll \frac{x \log_2 x}{q^2},$$

and summing this over $q \geq \log_2 x$ we obtain that the number of possible n in this case is $O(x/\log_3 x)$.

This finishes the proof, noting that all size bounds established are $o(x)$. $\square$

2.2. **Primes in arithmetic progressions.** In the proof of Theorem 3.1, we will need asymptotic estimates for the count of primes in fairly short progressions. The following theorem is sufficient, though we must exclude moduli which are multiples of a certain integer.

Theorem 2.4. *Let X and T satisfy $X \geq T \geq 2$, and suppose $q \leq T^{2/3}$. Then*

$$\pi(X; q, a) \sim \pi(X)/\varphi(q) \quad \text{as} \quad \frac{\log X}{\log T} \to \infty$$

uniformly for all coprime pairs (a, q), except possibly for those q which are multiples of some integer $q_1(T)$.

Proof. Referring to the proof of Linnik's theorem in [Bom74, p. 55], we have

$$\sum_{\substack{p \leq X \\ p \equiv a \pmod q}} \log p = \frac{X}{\varphi(q)} + O\left(\frac{X}{\varphi(q)} \exp(-c_1 A)\right) + O\left(\frac{X \log X}{T}\right) + O\left(\frac{1}{\varphi(q)} X^{1/2} T^5\right),$$

unless q is divisible by a certain exceptional modulus $q_1 = q_1(T)$, and where A is such that $T = X^{1/A}$. Notice that the quantity A tends to infinity. Therefore, the first and last O-terms are $o(X/\varphi(q))$, as is the middle O-term, provided $T > (\log X)^6$; but if $T \leq (\log X)^6$, Theorem 2.4 follows from the Siegel-Walfisz theorem. One now obtains the desired asymptotic for $\pi(X; q, a)$ by standard arguments. $\square$

We will also make use of the following fact [Polar, Lemma 2.7]:

Proposition 2.5. *Let q be a natural number with $q \leq x^{\frac{1}{2 \log_3 x}}$. The number of $n \leq x$ not belonging to $\mathcal{E}_A(x) \cup \mathcal{E}_B(x)$ for which q divides $s(n)$ is*

$$\ll \frac{\tau(q)}{\varphi(q)} \cdot x \log_3 x.$$

3. An average result for $\omega(s(n))$

First we prove an average order result for $\omega(s(n))$:

Theorem 3.1. *As $x \to \infty$,*

$$\sum_{n \leq x \,:\, n \notin \mathcal{E}(x)} \omega(s(n)) \sim x \log \log x.$$

This result will serve as a stepping stone towards Theorem 1.4.

By Lemma 2.1,

$$\sum_{n \leq x \,:\, n \notin \mathcal{E}(x)} \omega(s(n)) = \left(\sum_{\substack{\log_2 x < p \leq x^{1/\sqrt{\log_2 x}} \\ p \neq p_1(T)}} \sum_{\substack{n \leq x \,:\, n \notin \mathcal{E}(x) \\ p \mid s(n)}} 1 \right) + o(x \log_2 x),$$

where $p_1(T)$ is the exceptional modulus coming from Theorem 2.4. Eventually we will be counting primes in certain progressions modulo $p \leq x^{1/\sqrt{\log_2 x}}$, and so it will suffice to take $T = x^{1.5/\sqrt{\log_2 x}}$. Since p is prime, this excludes at most one summand above. Fix a prime $p \in (\log_2 x, x^{1/\sqrt{\log_2 x}}]$, $p \neq p_1(T)$ and consider the inner sum above.

Since $n \notin \mathcal{E}(x)$, we can write $n = mP$, $P := P(n)$, where $P \nmid m$, $P > x^{1/\log_3 x}$, $x^{1/\log_3 x} < P(m) < x/2m$ (this follows from condition D, noting that $m = n/P$), and $p \mid (m, \sigma(m)) \implies p < \log_2 x$. Now

$$\sum_{\substack{n\le x\,:\,n\notin\mathcal{E}(x)\\ p\mid s(n)}} 1 = \sum_{\substack{n\le x\,:\,n\notin\mathcal{E}(x)\\ p\mid s(n)\\ p\mid s(m)}} 1 + \sum_{\substack{n\le x\,:\,n\notin\mathcal{E}(x)\\ p\mid s(n)\\ p\nmid s(m)}} 1.$$

But this first sum is actually empty! If $p \mid s(n) = s(mP) = Ps(m) + \sigma(m)$, and $p \mid s(m)$, then $p \mid m$ and $p \mid \sigma(m)$. Since $n \notin \mathcal{E}(x)$, this forces $p < \log_2 x$, contradicting our choice of the fixed prime p.

If $p \nmid s(m)$, then since $p \mid s(n) = Ps(m) + \sigma(m)$, we have

$$P \equiv -\sigma(m)s(m)^{-1} \pmod p.$$

For convenience, we now define the following notation: Write

$$\sideset{}{'}\sum_{m} := \sum_{\substack{x^{1/\log_3 x} < m \le x^{1-1/\log_3 x}\\ p\nmid s(m)\\ x^{1/\log_3 x} < P(m) < x/2m\\ P(m)^2 \nmid m\\ q\mid(m,\sigma(m)) \implies q<\log_2 x}}.$$

Now

$$\sum_{\substack{n\le x\,:\,n\notin\mathcal{E}(x)\\ p\mid s(n)\\ p\nmid s(m)}} 1 = \sideset{}{'}\sum_{m} \sum_{\substack{P(m)<P\le x/m\\ P\equiv -\sigma(m)s(m)^{-1} \pmod p}} 1.$$

The inner sum is equal to

$$\pi(x/m; p, -\sigma(m)s(m)^{-1}) - \pi(P(m); p, -\sigma(m)s(m)^{-1}).$$

We now use Theorem 2.4 to rewrite the terms above, with $T = x^{1.5/\sqrt{\log_2 x}}$. Note that $P(m), x/m > x^{1/\log_3 x}$, so $P(m)$ and x/m are both greater than any fixed power of T for large enough x. Since $p \ne p_1(T)$, we have

$$\begin{aligned}
&\sideset{}{'}\sum_{m} \Big(\pi(x/m; p, -\sigma(m)s(m)^{-1}) - \pi(P(m); p, -\sigma(m)s(m)^{-1})\Big)\\
&\qquad = \frac{1}{p-1}\sideset{}{'}\sum_{m}\Big(\pi(x/m) - \pi(P(m))\Big) + o\left(\frac{1}{p-1}\sideset{}{'}\sum_{m}\Big(\pi(x/m) + \pi(P(m))\Big)\right).
\end{aligned}$$

To deal with the o-term, observe that from our conditions on m,

$$\begin{aligned}
\pi(x/m) + \pi(P(m)) &\le \pi(x/m) + \pi(x/2m)\\
&\le 2\pi(x/m) \le 10\big(\pi(x/m) - \pi(P(m))\big)
\end{aligned}$$

for large enough x, and so the above sum on m is

$$\sim \frac{1}{p-1}\sideset{}{'}\sum_{m}\Big(\pi(x/m) - \pi(P(m))\Big).$$

A prime P counted by the term $\pi(x/m) - \pi(P(m))$ corresponds to an integer $n = Pm$, with $P^2 \nmid n$ and $p \nmid s(m)$; the conditions imposed on m guarantee that $n \le x$ and $n \notin \mathcal{E}(x)$.

It is clear that every $n \leq x$ with $n \notin \mathcal{E}(x)$ and $p \nmid s(m)$ will be counted by such a summand. Thus

$$\begin{aligned}\sum_m{}' \pi(x/m) - \pi(P(m)) &= \#\{n \leq x : n = mP(n) \notin \mathcal{E}(x), p \nmid s(m)\} \\ &= \#\{n \leq x : n \notin \mathcal{E}(x)\} - \#\{n \leq x : n = mP(n) \notin \mathcal{E}(x), p \mid s(m)\}.\end{aligned}$$

The following lemma allows us to estimate the subtrahend.

Lemma 3.2. *For a fixed prime $\log_2 x < q \leq x^{1/\sqrt{\log_2 x}}$, the number of $n \leq x$ not belonging to $\mathcal{E}(x)$ (so $n = mP(n)$) such that $q \mid s(m)$ is*

$$\ll \frac{x \log_3 x \log_4 x}{q-1}.$$

Proof. Write $n = mP$, with $P = P(n)$. Since $n \notin \mathcal{E}(x)$, $P(m)^2 \nmid m$ and $P(m) > x^{1/\log_3 x}$. Note that $q \leq x^{1/\sqrt{\log_2 x}}$, and $(x/P)^{\frac{1}{2\log_3 x}} > x^{\frac{1}{2(\log_3 x)^2}} > x^{1/\sqrt{\log_2 x}}$ for large enough x. By Proposition 2.5, the number of $m \leq x/P$ not in $\mathcal{E}(x)$ for which $q \mid s(m)$ is

$$\ll \frac{1}{q-1} \cdot \frac{x \log_3 x}{P}.$$

Summing over $P \in (x^{1/\log_3 x}, x]$ gives the result. $\square$

By the above lemma, we have

$$\sum_m{}' \Big(\pi(x/m) - \pi(P(m))\Big) = x + O\left(\frac{x \log_3 x \log_4 x}{p-1}\right) + O(\#\mathcal{E}(x)).$$

Putting everything together, we have shown that

$$\sum_{n \leq x \,:\, n \notin \mathcal{E}(x)} \omega(s(n)) = \sum_{\substack{\log_2 x < p \leq x^{1/\sqrt{\log_2 x}} \\ p \neq p_1(T)}} \left(\frac{x}{p-1} + O\left(\frac{x \log_3 x \log_4 x}{(p-1)^2}\right) + o\left(\frac{x}{p-1}\right)\right).$$

The O-term contributes

$$\ll x \log_3 x \log_4 x \sum_{t > \log_2 x} \frac{1}{t^2} \ll \frac{x \log_3 x \log_4 x}{\log_2 x} = o(x \log_2 x).$$

By Mertens' first theorem,

$$\begin{aligned}\sum_{\substack{\log_2 x < p \leq x^{1/\sqrt{\log_2 x}} \\ p \neq p_1(T)}} \frac{1}{p-1} &= \log_2(x^{1/\sqrt{\log_2 x}}) - \log_2(\log_2(x)) + O(1) \\ &= (\log_2 x)(1 + o(1)).\end{aligned}$$

Thus,

$$\sum_{n \leq x \,:\, n \notin \mathcal{E}(x)} \omega(s(n)) = x \log_2 x + o(x \log_2 x),$$

as desired.

4. Proof of Theorem 1.4

Notice that

$$\begin{aligned}\sum_{n\notin\mathcal{E}(x)}(\omega(s(n))-\log_2 x)^2 &= \sum_{n\notin\mathcal{E}(x)}\omega(s(n))^2 - 2\log_2 x\sum_{n\notin\mathcal{E}(x)}\omega(s(n)) + x(\log_2 x)^2(1+o(1))\\ &= \sum_{n\notin\mathcal{E}(x)}\omega(s(n))^2 - x(\log_2 x)^2(1+o(1)).\end{aligned}$$

by Theorem 3.1. Hence, the following lemma implies Theorem 1.4.

Lemma 4.1. *As $x\to\infty$,*

$$\sum_{n\le x\,:\,n\notin\mathcal{E}(x)}\omega(s(n))^2 \sim x(\log_2 x)^2.$$

We have

$$\sum_{n\le x\,:\,n\notin\mathcal{E}(x)}\omega(s(n))^2 = \sum_{n\notin\mathcal{E}(x)}\left(\sum_{p\mid s(n)}1\right)^2 = \sum_{n\notin\mathcal{E}(x)}\sum_{p,q\,:\,p,q\mid s(n)}1$$

where the sum is over pairs of primes p, q. This inner sum is equal to

$$\sum_{\substack{p,q\,:\,p,q\mid s(n)\\ p\neq q}}1+\sum_{p\mid s(n)}1,$$

and hence

$$\sum_{n\le x\,:\,n\notin\mathcal{E}(x)}\omega(s(n))^2 = \sum_{n\notin\mathcal{E}(x)}\sum_{\substack{p,q\,:\,p,q\mid s(n)\\ p\neq q}}1+o(x(\log_2 x)^2)$$

by Theorem 3.1.

We can again truncate the sum in question.

Lemma 4.2.

$$\sum_{n\notin\mathcal{E}(x)}\omega(s(n))^2 = \left(\sum_{n\notin\mathcal{E}(x)}\sum_{\substack{\log_2 x<p\le x^{1/\sqrt{\log_2 x}}\\ p\mid s(n)}}\sum_{\substack{\log_2 x<q\le x^{1/\sqrt{\log_2 x}}\\ q\mid s(n)\\ q\neq p}}1\right)+o(x(\log_2 x)^2).$$

Proof. By Lemma 2.1, we have $\omega(s(n))=\omega'(s(n))+o(\log_2 x)$, where $\omega'(m)$ denotes the number of prime divisors p of m with $p\in(\log_2 x, x^{1/\sqrt{\log_2 x}}]$. Hence,

$$\sum_{n\notin\mathcal{E}(x)}\omega(s(n))^2 = \sum_{n\notin\mathcal{E}(x)}\omega'(s(n))^2+o\left(\log_2 x\sum_{n\notin\mathcal{E}(x)}\omega'(s(n))\right)+o(x(\log_2 x)^2).$$

By Theorem 3.1, $\sum_{n\notin\mathcal{E}(x)}\omega'(s(n))\sim x\log_2 x$, and the lemma follows. □

4.1. **Proof of Lemma 4.1.**

As before, we fix primes $p,q\in(\log_2 x, x^{1/\sqrt{\log_2 x}}]$ and count the number of $n\notin\mathcal{E}(x)$ where $p,q\mid s(n)$. Eventually we will need to use Theorem 2.4 to count primes modulo pq, and so we take $T=x^{3/\sqrt{\log_2 x}}$ in the theorem. However, we must ensure that pq is not a multiple of some integer $q_1(T)$. Let $p_1(T)$ denote the largest prime factor of $q_1(T)$. If pq is a multiple of $q_1(T)$, then $p_1(T)$ divides pq, which forces $p_1(T)=p$ or $p_1(T)=q$. Therefore, we will insist that $p,q\neq p_1(T)$, which excludes at most one each of p and q.

Write $n = mP$, $P := P(n)$, where $P \nmid m$, $P > x^{1/\log_3 x}$, $x^{1/\log_3 x} < P(m) < x/2m$, and a prime $\ell \mid (n, \sigma(n)) \implies \ell < \log_2 x$. As before, we can split the above sum in two, with one sum over n such that $p, q \nmid s(m)$ and the other over n such that either p or q divides $s(m)$; and as before, the latter sum will be empty, by the same argument. Thus we are reduced to considering $n \notin \mathcal{E}(x)$ with $p, q \mid s(n)$ and $p, q \nmid s(m)$.

If both p and q divide $s(n)$ but not $s(m)$, then since $s(n) = Ps(m) + \sigma(m)$, we have

$$P \equiv -\sigma(m)s(m)^{-1} \pmod{pq}.$$

Therefore

$$\sum_{\substack{n \notin \mathcal{E}(x) \\ p,q \mid s(n) \\ p,q \nmid s(m)}} 1 = {\sum_m}' \sum_{\substack{P(m) < P \le x/m \\ P \equiv -\sigma(m)s(m)^{-1} \pmod{pq}}} 1,$$

where now ${\sum_m}'$ includes the condition $q \nmid s(m)$.

The inner sum is equal to

$$\pi\big(x/m; pq, -\sigma(m)s(m)^{-1}\big) - \pi\big(P(m); pq, -\sigma(m)s(m)^{-1}\big).$$

We once again use Theorem 2.4 on the terms above, with $T = x^{3/\sqrt{\log_2 x}}$. The analysis proceeds exactly as before, with the factor of $1/\varphi(p)$ replaced by $1/\varphi(pq)$. We have in the end that the inner sum (over n) is asymptotic to

$$\frac{1}{\varphi(pq)}\big(\#\{n \le x : n \notin \mathcal{E}(x)\} - \#\{n \le x : n = mP(n) \notin \mathcal{E}(x), \text{ either } p \text{ or } q \mid s(m)\}\big).$$

Applying Lemma 3.2 twice, we have that the number of $n \le x$ not belonging to $\mathcal{E}(x)$ such that either p or q divides $s(m)$ is

$$\ll x \log_3 x \log_4 x \left(\frac{1}{p-1} + \frac{1}{q-1}\right).$$

Putting everything together, we have shown that

$$\sum_{n \le x : n \notin \mathcal{E}(x)} \omega(s(n))^2 = x \sum_p \sum_q \frac{1}{(p-1)(q-1)}$$
$$+ \sum_p \sum_q O\left(\frac{x \log_3 x \log_4 x}{(p-1)^2(q-1)} + \frac{x \log_3 x \log_4 x}{(p-1)(q-1)^2}\right) + o(x(\log_2 x)^2),$$

where the sums over p and q have the restrictions $p, q \in (\log_2 x, x^{1/\sqrt{\log_2 x}}]$, $q \neq p$, and $p, q \neq p_1(T)$ (from Theorem 2.4). Now, since $q \le x^{1/\sqrt{\log_2 x}}$ and $\sum_p 1/p^2$ converges,

$$\sum_p \sum_q \frac{1}{(p-1)^2(q-1)} = O(\log_2 x)$$

by Mertens' first theorem. Hence the O-term contributes

$$\ll x \log_2 x \log_3 x \log_4 x = o(x(\log_2 x)^2),$$

and so

$$\sum_{n \le x : n \notin \mathcal{E}(x)} \omega(s(n))^2 = x \sum_p \sum_q \frac{1}{(p-1)(q-1)} + o(x(\log_2 x)^2).$$

Another application of Mertens' theorem tells us

$$\sum_{n \le x : n \notin \mathcal{E}(x)} \omega(s(n))^2 = x(\log_2 x)^2 + o(x(\log_2 x)^2),$$

which completes the proof.

5. From $\omega(s(n))$ to $\Omega(s(n))$

We conclude by showing that the result proved in the previous section holds with $\omega(s(n))$ replaced by $\Omega(s(n))$.

Lemma 5.1.
$$\sum_{n\notin\mathcal{E}(x)} \big(\Omega(s(n))-\omega(s(n))\big)^2 = o(x(\log_2 x)^2).$$

It follows quickly from this lemma that $\sum_{n\notin\mathcal{E}(x)} \big(\Omega(s(n))-\log_2 x\big)^2 = o(x(\log_2 x)^2)$. Adding and subtracting $\omega(s(n))$ inside the square and expanding, we have
$$\sum_{n\notin\mathcal{E}(x)} \big(\Omega(s(n))-\log_2 x\big)^2 = \sum_{n\notin\mathcal{E}(x)} \big(\Omega(s(n))-\omega(s(n))\big)^2 + \sum_{n\notin\mathcal{E}(x)} \big(\omega(s(n))-\log_2 x\big)^2$$
$$+\,2\sum_{n\notin\mathcal{E}(x)} \Big[\big(\Omega(s(n))-\omega(s(n))\big)\big(\omega(s(n))-\log_2 x\big)\Big].$$

The first and second sums are $o(x(\log_2 x)^2)$, by the above lemma and Theorem 1.4, respectively. An application of the Cauchy-Schwarz inequality shows us that the last term squared is
$$\ll \left(\sum_{n\notin\mathcal{E}(x)} \big(\Omega(s(n))-\log_2 x\big)^2\right)\left(\sum_{n\notin\mathcal{E}(x)} \big(\omega(s(n))-\log_2 x\big)^2\right);$$
using Lemma 5.1 and Theorem 1.4 once more and taking square roots, we see that this is also $o(x(\log_2 x)^2)$.

Proof of Lemma 5.1. We wish to estimate from above the quantity
$$\sum_{n\notin\mathcal{E}(x)} \big(\Omega(s(n))-\omega(s(n))\big)^2 = \sum_{n\le x\,:\,n\notin\mathcal{E}(x)} \Bigg(\sum_{\substack{p^k\mid s(n)\\ k\ge 2}} 1\Bigg)^2$$
$$= \sum_{n\notin\mathcal{E}(x)} \sum_{\substack{p^k,q^j\mid s(n)\\ k,j\ge 2\\ p\ne q}} 1 + \sum_{n\notin\mathcal{E}(x)} \sum_{\substack{p^k,p^j\mid s(n)\\ k,j\ge 2}} 1.$$

We handle the "$p=q$" sum first. We obtain an upper bound by insisting $j\le k$ and inserting a factor of 2. If $p^k\mid s(n)$, the condition $p^k,p^j\mid s(n)$ is satisfied for each j with $2\le j\le k$, and so the sum is
$$\ll \sum_{n\notin\mathcal{E}(x)} \sum_{\substack{p^k\mid s(n)\\ k\ge 2}} k = \sum_{k\ge 2} k \sum_{n\notin\mathcal{E}(x)} \sum_{p\,:\,p^k\mid s(n)} 1.$$

A number $m\le x^2$ has at most $\frac{20}{k}\log_3 x$ primes $p>x^{1/10\log_3 x}$ with $p^k\mid m$. Define $L_1:=\lfloor 30\log_3 x\rfloor$. If $p>x^{1/10\log_3 x}$, then $p^{L_1}>x^2$, so certainly for $k>L_1$ the condition $p^k\mid s(n)$ cannot be met. Therefore,
$$\sum_{k\ge 2} k \sum_{n\notin\mathcal{E}(x)} \sum_{\substack{p>x^{1/10\log_3 x}\\ p^k\mid s(n)}} 1 \le 20x\log_3 x \sum_{k=2}^{L_1} 1 = O(x\log_3 x\cdot L_1),$$

and $x \log_3 x \cdot L_1 \ll x(\log_3 x)^2$.

It remains to consider

$$\sum_{n \notin \mathcal{E}(x)} \sum_{\substack{p \le x^{1/10 \log_3 x} \\ p^k | s(n) \\ k \ge 2}} k = \sum_{p \le x^{1/10 \log_3 x}} \sum_{\substack{n \notin \mathcal{E}(x) \\ p^k | s(n) \\ k \ge 2}} k = \sum_{p \le x^{1/10 \log_3 x}} \sum_{\substack{n \notin \mathcal{E}(x) \\ p^k | s(n) \\ p^k \le x^{1/2 \log_3 x} \\ k \ge 2}} k + \sum_{p \le x^{1/10 \log_3 x}} \sum_{\substack{n \notin \mathcal{E}(x) \\ p^k | s(n) \\ p^k > x^{1/2 \log_3 x} \\ k \ge 2}} k.$$

For the first sum, the condition $p^k \le x^{1/2 \log_3 x}$ allows us to apply Proposition 2.5, which gives

$$\sum_{p \le x^{1/10 \log_3 x}} \sum_{\substack{n \notin \mathcal{E}(x) \\ p^k | s(n) \\ p^k \le x^{1/2 \log_3 x} \\ k \ge 2}} k \ll x \log_3 x \sum_{k=2}^{L_2} \sum_{p \le x^{1/10 \log_3 x}} \frac{k^2}{p^k},$$

with $L_2 = \lfloor 3 \log x \rfloor$. But the sum of k^2/p^k over all $p \ge 2$ and all $k \ge 2$ is $O(1)$, so this is $O(x \log_3 x)$.

For the second sum, define $\ell(p) := \max\{m \in \mathbb{N} : p^m \le x^{1/2 \log_3 x}\}$. Notice that $\ell(p) < \log x$ trivially and that $p^{\ell(p)} > x^{1/2 \log_3 x}/p$. The second sum is bounded from above by

$$\sum_{\substack{p \le x^{1/10 \log_3 x} \\ p^k > x^{1/2 \log_3 x} \\ k \ge 2}} \sum_{\substack{n \notin \mathcal{E}(x) \\ p^{\ell(p)} | s(n)}} k \ll x \log_3 x \sum_{k=2}^{L_2} k \sum_{p \le x^{1/10 \log_3 x}} \frac{\ell(p)}{p^{\ell(p)}},$$

using Proposition 2.5 once more. The above inequalities then show that this sum is

$$\ll \frac{x (\log x)^3 \log_3 x}{x^{1/2 \log_3 x}} \cdot x^{1/10 \log_3 x} = o(x).$$

Thus $\sum_{n \notin \mathcal{E}(x)} \big(\Omega(s(n)) - \omega(s(n))\big) \ll x \log_3 x \log_4 x$, which is $o(x \log_2 x)$.

We now turn our attention to $\sum_{n \notin \mathcal{E}(x)} \sum_{p^k, q^j | s(n)} 1$, with $k, j \ge 2$ and $p \ne q$. Arguments similar to those used before show

$$\sum_{n \notin \mathcal{E}(x)} \sum_{\substack{p^k | s(n) \\ k \ge 2}} \sum_{\substack{q^j | s(n) \\ j \ge 2 \\ q \ne p}} 1 = \sum_{n \notin \mathcal{E}(x)} \sum_{\substack{p^k | s(n) \\ k \ge 2}} \sum_{\substack{q \le x^{1/10 \log_3 x} \\ q^j | s(n) \\ j \ge 2 \\ q \ne p}} 1 + o(x \log_2 x \log_3 x),$$

and p can certainly be restricted in the same way. Rearranging the sum, we have

$$\sum_{n \notin \mathcal{E}(x)} \sum_{\substack{p \le x^{1/10 \log_3 x} \\ p^k | s(n) \\ k \ge 2}} \sum_{\substack{q \le x^{1/10 \log_3 x} \\ q^j | s(n) \\ j \ge 2 \\ q \ne p}} 1 = \sum_{p \le x^{1/10 \log_3 x}} \sum_{\substack{q \le x^{1/10 \log_3 x} \\ q \ne p}} \sum_{\substack{n \notin \mathcal{E}(x) \\ p^k, q^j | s(n) \\ k, j \ge 2}} 1.$$

We now proceed in essentially the same way as before. Write the inner sum over q as two sums, one over $q^j \le x^{1/4 \log_3 x}$ and the other over $q^j > x^{1/4 \log_3 x}$. Do the same for the outer sum over p. When the dust settles, we are left with four sums to handle, one for each possible combination of ranges for p^k and q^j.

First we handle the case $p^k, q^j \le x^{1/4\log_3 x}$. Using Proposition 2.5, we obtain

$$\sum_{p\le x^{1/10\log_3 x}} \sum_{\substack{q\le x^{1/10\log_3 x}\\ q\ne p}} \sum_{\substack{n\notin\mathcal{E}(x)\\ p^k,q^j\mid s(n)\\ p^k,q^j\le x^{1/4\log_3 x}\\ k,j\ge 2}} 1 \ll x\log_3 x \sum_{k\ge 2}\sum_{j\ge 2}\sum_{p\le x^{1/10\log_3 x}} \sum_{\substack{q\le x^{1/10\log_3 x}\\ q\ne p}} \frac{kj}{p^k q^j}$$
$$\ll x\log_3 x.$$

If $p^k, q^j > x^{1/4\log_3 x}$, we define $\ell(p) = \max\{m : p^m \le x^{1/4\log_3 x}\}$ and, as before, obtain that this sum is

$$\ll x\log_3 x \sum_{k\ge 2}\sum_{j\ge 2}\sum_{p\le x^{1/10\log_3 x}} \sum_{q\le x^{1/10\log_3 x}} \frac{\ell(p)\ell(q)}{p^{\ell(p)}q^{\ell(q)}} \ll \frac{x(\log x)^4\log_3 x}{x^{1/2\log_3 x}}\cdot x^{1/5\log_3 x},$$

which is $o(x)$.

Assume now that $q^j \le x^{1/4\log_3 x} \le p^k$; the final case is completely similar. Then this sum is bounded from above by

$$\sum_{\substack{p\le x^{1/10\log_3 x}\\ p^k> x^{1/4\log_3 x}\\ k\ge 2}} \sum_{\substack{q\le x^{1/10\log_3 x}\\ q^j\le x^{1/4\log_3 x}\\ j\ge 2}} \sum_{\substack{n\notin\mathcal{E}(x)\\ p^{\ell(p)},q^j\mid s(n)}} 1 \ll x\log_3 x \sum_{k=2}^{L_2}\sum_{j\ge 2}\sum_{p\le x^{1/10\log_3 x}} \sum_{\substack{q\le x^{1/10\log_3 x}\\ q\ne p}} \frac{\ell(p)j}{p^{\ell(p)}q^j}$$
$$\ll \frac{x(\log x)^2\log_3 x}{x^{1/3\log_3 x}}. \qquad \square$$

Remark. We know, by the celebrated Erdős - Kac theorem, that (roughly speaking) $\omega(n)$ is normally distributed with mean and variance $\log\log n$. The corresponding theorem for $\omega(\sigma(n))$ follows from methods of Erdős and Pomerance [EP85]: $\omega(\sigma(n))$ is also normally distributed, but with mean $\frac{1}{2}(\log\log n)^2$ and standard deviation $\frac{1}{\sqrt{3}}(\log\log n)^{3/2}$. One hopes that something similar can be said about $\omega(s(n))$. The results of the present article indicate that $s(n)$ typically has just as many prime factors as n; for this reason, we expect the Erdős - Kac theorem to hold with $\omega(s(n))$ in place of $\omega(n)$.

Acknowledgements

The problem tackled here first arose in conversations between Paul Pollack and Carl Pomerance; the author is indebted to both. The author would like to extend further gratitude to Paul Pollack for advice and suggestions in the course of writing this article.

References

[Bom74] E. Bombieri, *Le grand crible dans la théorie analytique des nombres*, Astérisque **18** (1974), 1 – 87.

[dB66] N. G. de Bruijn, *On the number of positive integers $\le x$ and free of prime factors $> y$. II*, Indag. Math. **28** (1966), 239 – 247.

[Dic13] L. E. Dickson, *Theorems and tables on the sum of the divisors of a number*, Quart. J. Math. **44** (1913), 264 – 296.

[EGPS90] P. Erdős, A. Granville, C. Pomerance, and C. Spiro, *On the normal behavior of the iterates of some arithmetic functions*, Analytic Number Theory, Proc. Conf. in honor of Paul T. Bateman (B. C. Berndt et al., eds.), Birkhäuser, Boston, 1990, pp. 165 – 204.

[EP85] P. Erdős and C. Pomerance, *The normal number of prime factors of $\varphi(n)$*, Rocky Mtn. J. Math. **15** (1985), 343 – 352.

[HR17] G. H. Hardy and S. Ramanujan, *The normal number of prime factors of a number n*, Quart. J. Math. **48** (1917), 76 – 92.

[Polar] P. Pollack, *Some arithmetic properties of the sum of proper divisors and the sum of prime divisors*, Illinois J. Math. (to appear).

Department of Mathematics, Boyd Graduate Studies Research Center, University of Georgia, Athens, GA 30602, USA
E-mail address: ltroupe@math.uga.edu