ON THE EQUATION $\varphi(n) = \varphi(n+1)$

PAUL KINLAW[1], MITSUO KOBAYASHI[2], AND CARL POMERANCE[3]

ABSTRACT. We consider solutions of the equation $\varphi(n) = \varphi(n+1)$, where φ denotes Euler's function. Improving on previous work, we show that the reciprocal sum over all such n is less than 8.

1. INTRODUCTION.

We study solutions of the equation $\varphi(n) = \varphi(n+1)$, where φ denotes Euler's function. Let $\mathcal{S} = \{n \in \mathbb{N} : \varphi(n) = \varphi(n+1)\} = \{1, 3, 15, \dots\}$ and let $S(x)$ denote the number of $n \in \mathcal{S}$ not exceeding x. In 1936, Erdős [4] proved that $\mathcal{S}$ has asymptotic density zero. In 1987, Erdős et al. [5, Theorem 3] proved that $S(x) < x/e^{\sqrt[3]{\log x}}$ for all sufficiently large x. The cube root of $\log x$ was improved recently to the square root by Yamada [11].

It is still not known if there are infinitely many solutions. However, it is conjectured in [5] that $S(x) > x^{1-\varepsilon}$ for all $\varepsilon > 0$ and $x > C_\varepsilon$.

From the upper bound results for $S(x)$ it follows that the reciprocal sum is finite. As with Brun's constant, where one attempts to get good estimates for the reciprocal sum of primes p with $p + 2$ also prime, it is a challenge to get good estimates for the reciprocal sum of members of $\mathcal{S}$. It is shown in [1] that the reciprocal sum is less than 441702 and conjectured that the value is less than 2. We improve the upper bound.

Theorem 1.1. *We have*

$$\sum_{n \in \mathcal{S}} \frac{1}{n} < 7.6472.$$

The proof makes use of the exact computation of $\mathcal{S}$ up to 10^{13}. Beyond that point, an averaging argument is employed to greatly limit the possibilities for the odd member of $\{n, n+1\}$ for $n \in \mathcal{S}$. Indeed, for $n \in \mathcal{S}$ we have $\varphi(n)/n \approx \varphi(n+1)/(n+1)$, and the even member has this ratio at most $1/2$. The averaging argument shows that only a small density of odd numbers n have $\varphi(n)/n$ so small.

To be sure, even if a set has a very small density, if that density is positive, then the reciprocal sum will be infinite. So averaging arguments can take us only so far. Several new techniques are used to deal with the large range, $n > e^{150}$. These include methods suggested by Patrick Letendre, and similar to the methods employed by Yamada [11]. We use several techniques from [7] on the distribution of numbers with no large prime factors. Most helpful

HUSSON UNIVERITY, 1 COLLEGE CIRCLE, BANGOR, ME 04401, USA[1]
MATHEMATICS DEPARTMENT, DARTMOUTH COLLEGE, HANOVER, NH 03755, USA[2–3]
E-mail addresses: KinlawP@Husson.edu, mitsuo.kobayashi@dartmouth.edu, carl.pomerance@dartmouth.edu.

2010 *Mathematics Subject Classification.* 11N25, 11Y99.
Key words and phrases. Euler's function, reciprocal sum, smooth numbers.

is a new paper of Bennett et al. [2] on numerically explicit estimates for the distribution of primes in residue classes.

2. Notation and Preliminary Lemmas

We split the sum into three intervals, with cutoffs at 10^{13} and $X_0 = e^{150}$. We let $\exp(x)$ and $\log x$ denote the natural exponential and logarithmic functions. We let x denote a real number, m and n denote positive integers, p, q, r denote prime numbers, $P(n)$ denote the largest prime factor of n, and $\pi(x)$ denote the prime counting function.

We state several preliminary lemmas that will be used in the proof of Theorem 1.1. We will use the bounds [9, (3.5, 3.6)] of Rosser and Schoenfeld and [3, Cor 5.2, Thm. 5.6] of Dusart for the prime counting function.

Lemma 2.1. *For all $x > 1$, we have*

$$\pi(x) < 1.25506x/\log x,$$
$$\pi(x) \le x/\log x\,(1 + 1.2762/\log x),$$
$$\pi(x) \le x/\log x\left(1 + 1/\log x + 2.53816/\log^2 x\right).$$

For all $x \ge 17$, we have $\pi(x) > x/\log x$.

Lemma 2.2. *For all $x \ge 2278383$ we have*

$$\left|\sum_{p\le x}\frac{1}{p} - (\log\log x + B)\right| \le \frac{0.2}{\log^3 x},$$

where $B = 0.2614972128\ldots$ denotes the Mertens constant.

Let $\pi(x; m, a) = |\{p \le x : p \equiv a \pmod m\}|$.

Lemma 2.3. *For $m < C < D$, we have*

$$\sum_{\substack{C<p\le D\\ p\equiv a\,(m)}}\frac{1}{p} < \frac{2}{\varphi(m)}\left(\log\log(D/m) - \log\log(C/m) + \frac{1}{\log(D/m)}\right).$$

Lemma 2.3 follows directly from the Brun-Titchmarsh theorem by partial summation, see for instance [7, Lem. 2.8]. A more elementary result that can complement Lemma 2.3 is the following.

Lemma 2.4. *Suppose that m is a positive integer coprime to 6. We have*

$$\sum_{\substack{p\le 398m\\ p\equiv 1\,(m)}}\frac{1}{p} < \frac{2.0156}{m}.$$

Proof. Since m is odd, the primes in the sum are the primes in the set $\{2m+1, 4m+1, \ldots, 396m+1\}$. If $m \equiv 1 \pmod 3$ then the numbers $2jm+1$ with $j \equiv 1 \pmod 3$ are divisible by 3, and if $m \equiv 2 \pmod 3$, the numbers $2jm+1$ with $j \equiv 2 \pmod 3$ are divisible by 3. Thus, the sum in the lemma is either at most

$$\frac{1}{m}\sum_{\substack{j\le 198\\ j\not\equiv 1\,(3)}}\frac{1}{2j} \quad\text{or}\quad \frac{1}{m}\sum_{\substack{j\le 198\\ j\not\equiv 2\,(3)}}\frac{1}{2j}.$$

The second sum here is larger than the first sum, and the second sum is < 2.0156. $\square$

Corollary 2.5. *For $r > 3$ prime and $x > 398r$, we have*

$$\sum_{\substack{p \le x \\ p \equiv 1\,(r)}} \frac{1}{p} \le \frac{2}{r-1}\left(\log\log(x/r) - 0.78169 + \frac{1}{\log(x/r)}\right).$$

The corollary follows from Lemmas 2.3 and 2.4 since $-\log\log 398 + 2.0156/2 < -0.78169$. We will also use the following inequality.

Lemma 2.6. *For a positive integer $m \le 1200$ and $x > 50m^2$, we have*

$$\sum_{\substack{50m^2 < p \le x \\ p \equiv 1\,(m)}} \frac{1}{p} < \frac{1}{\varphi(m)}\left(\log\log x - \log\log(50m^2) - \frac{1.5}{\log x} + \frac{2.5}{\log^2 x} + \frac{1.5}{\log(50m^2)}\right).$$

Proof. This follows from a partial summation argument and the following new result, see [2, Cor. 1.6]: under the hypotheses of the lemma,

$$\frac{x}{\varphi(m)\log x} < \pi(x; m, 1) < \frac{x}{\varphi(m)\log x}\left(1 + \frac{2.5}{\log x}\right).$$

$\square$

We also use the following bound [7, Lemma 2.7].

Lemma 2.7. *For all $y > 1$ we have*

$$\sum_{p > y} \frac{1}{p^2} < \frac{1}{y \log y}.$$

Corollary 2.8. *For all $y \ge 6241$, we have*

$$\sum_{\substack{p^a > y \\ a \ge 2}} \frac{1}{p^a} < \frac{2.4}{\sqrt{y}\log y}.$$

Proof. Using the bound

$$\sum_{\substack{p^a > y \\ a \ge 2}} \frac{1}{p^a} = \sum_{p \ge 2} \frac{1}{p(p-1)} - \sum_{\substack{p^a \le y \\ a \ge 2}} \frac{1}{p^a} < 0.773157 - \sum_{\substack{p^a \le y \\ a \ge 2}} \frac{1}{p^a},$$

a computer check shows that the claim holds for $6241 \le y < 10^8$. Assume that $y \ge 10^8$. We split the sum into two cases, $p > \sqrt{y}$ and $p \le \sqrt{y}$. We bound the first case as

$$\sum_{\substack{p > \sqrt{y} \\ p^a > y,\ a \ge 2}} \frac{1}{p^a} = \sum_{p > \sqrt{y}} \sum_{a \ge 2} \frac{1}{p^a} = \sum_{p > \sqrt{y}} \frac{1}{p(p-1)} < \frac{\sqrt{y}}{\sqrt{y}-1} \sum_{p > \sqrt{y}} \frac{1}{p^2}$$
$$< \frac{2}{\sqrt{y}\log y}\left(1 + \frac{1}{\sqrt{y}-1}\right),$$

using Lemma 2.7. We next address the second case. For $p \le \sqrt{y}$ let a_p be the least integer such that $p^{a_p} > y$. We have

$$\sum_{\substack{p \le \sqrt{y} \\ p^a > y,\ a \ge 3}} \frac{1}{p^a} = \sum_{p \le \sqrt{y}} \frac{1}{p^{a_p}} \frac{1}{1 - 1/p}.$$

We consider two cases, $a_p = 3$ and $a_p > 3$. For the first case, we have

$$\sum_{\substack{p \le \sqrt{y} \\ p^3 > y}} \frac{1}{p^3} \frac{1}{1 - 1/p} < \frac{y^{1/3}}{y^{1/3} - 1} \sum_{p > y^{1/3}} \frac{1}{p^3}.$$

By partial summation and Lemma 2.1,

$$\sum_{p > y^{1/3}} \frac{1}{p^3} = -\frac{\pi(y^{1/3})}{y} + \int_{y^{1/3}}^{\infty} \frac{3\pi(t)}{t^4} dt < \frac{2.4356}{y^{2/3} \log y} < \frac{0.1131}{\sqrt{y} \log y}.$$

For the second case, we have

$$\sum_{p \le y^{1/3}} \frac{1}{y} \frac{1}{1 - 1/p} < \frac{27.5742}{y} + \frac{101}{100y} \left(\pi(y^{1/3}) - 25\right)$$
$$< \frac{2.3242}{y} + \frac{0.1699}{\sqrt{y} \log y}.$$

Combining these bounds, we have

$$\sum_{\substack{p^a > y \\ a \ge 2}} \frac{1}{p^a} < \frac{2.2878}{\sqrt{y} \log y}$$

for all $y \ge 10^8$. This completes the proof of Corollary 2.8. □

3. An Averaging Method

Let $N(x)$ denote the number of odd $n \le x$ with $\varphi(n)/n < 1/2$.

Proposition 3.1. *We have $N(x) < 0.017876x + 670.515\sqrt{x} + 5.4$ for all $x > 0$.*

We will prove Proposition 3.1 after noting the following corollary.

Proof of Proposition 3.1. For a real number $T \ge 1$, let g_T denote the multiplicative function supported on the squarefree numbers such that $g_T(p) = (p/(p-1))^T - 1$. Thus,

$$\sum_{d \mid n} g_T(d) = (n/\varphi(n))^T.$$

Noting that 323323 is the product of all of the primes from 7 to 19, we partition the odd numbers n such that $\varphi(n)/n < 1/2$ into four classes:

(1) $\gcd(n, 6) = 1$,
(2) $\gcd(n, 30) = 3$,
(3) $\gcd(n, 30) = 15$ and $\gcd(n, 323323) = 1$,
(4) $\gcd(n, 30) = 15$ and $\gcd(n, 323323) > 1$.

Let $B_i(x)$ denote the number of $n \le x$ in each case i. For any $T \ge 1$,

$$B_1(x) \le \frac{1}{2^T} \sum_{\substack{n \le x \\ (n,6)=1}} \left(\frac{n}{\varphi(n)} \right)^T = \frac{1}{2^T} \sum_{\substack{n \le x \\ (n,6)=1}} \sum_{d|n} g_T(d).$$

Changing the order of summation, we obtain

$$B_1(x) \le \left(\frac{1}{2} \right)^T \sum_{\substack{d \le x \\ (d,6)=1}} g_T(d) \left(\frac{x}{3d} + \frac{2}{3} \right),$$

using the bound $|\{n \le t : \gcd(n,6) = 1\}| \le t/3 + 2/3$. Thus,

$$B_1(x) \le x \left(\frac{1}{3 \cdot 2^T} \sum_{(d,6)=1} \frac{g_T(d)}{d} \right) + \frac{2}{3 \cdot 2^T} \sum_{\substack{d \le x \\ (d,6)=1}} g_T(d).$$

Let S_1 and S_2 denote the first and second sums, respectively. We have

$$\sum_{(d,6)=1} \frac{g_T(d)}{d} = \prod_{p \ge 5} \left(1 + \frac{g_T(p)}{p} \right) = \exp\left(\sum_{p \ge 5} \log\left(1 + \frac{g_T(p)}{p} \right) \right).$$

We choose T as 69. Computing the sum for $p < 10^9$ and then majorizing the tail using Lemmas 2.1 or 2.2, we get

$$\sum_{p \ge 5} \log\left(1 + \frac{g_T(p)}{p} \right) < 34.3844.$$

Thus, $S_1 < (4.839 \times 10^{-7})x$.

We next turn to S_2. By Rankin's trick,

$$\sum_{\substack{d \le x \\ (d,6)=1}} g_T(d) \le \sqrt{x} \sum_{\substack{d \le x \\ (d,6)=1}} \frac{g_T(d)}{\sqrt{d}} \le \sqrt{x} \prod_{p \ge 5} \left(1 + \frac{g_T(p)}{\sqrt{p}} \right)$$

$$= \sqrt{x} \exp\left(\sum_{p \ge 5} \log\left(1 + \frac{g_T(p)}{\sqrt{p}} \right) \right).$$

Splitting the sum at 10^9 as before, we compute

$$\sum_{p \ge 5} \log\left(1 + \frac{g_T(p)}{\sqrt{p}} \right) < 49.1683,$$

so that $S_2 < 2.549\sqrt{x}$. Thus $B_1(x) < (4.839 \times 10^{-7})x + 2.549\sqrt{x}$.

We next bound $B_2(x)$. For a positive integer u, let

$$f_u(m) = \prod_{\substack{p \,|\, m \\ p \nmid u}} \left(\frac{p}{p-1} \right) \tag{1}$$

and let $g_{T,u}$ be the multiplicative function supported on the squarefree numbers coprime to u such that $g_{T,u}(p) = g_T(p)$ for $p \nmid u$. Thus,

$$\sum_{d|m} g_{T,u}(d) = f_u(m)^T.$$

We have

$$B_2(x) \le \frac{1}{2^T} \sum_{\substack{n \le x \\ (n,30)=3}} \left(\frac{n}{\varphi(n)}\right)^T = \left(\frac{3}{4}\right)^T \sum_{\substack{m \le \frac{x}{3} \\ (m,10)=1}} f_3(m)^T$$

and so, using the bound $|\{n \le t : \gcd(n,10) = 1\}| \le 2t/5 + 4/5$,

$$\begin{aligned} B_2(x) &\le \left(\frac{3}{4}\right)^T \sum_{\substack{m \le \frac{x}{3} \\ (m,10)=1}} \sum_{d \,|\, m} g_{T,3}(d) \le \left(\frac{3}{4}\right)^T \sum_{\substack{d \le \frac{x}{3} \\ (d,10)=1}} g_{T,3}(d) \left(\frac{2x}{15d} + \frac{4}{5}\right) \\ &< \left(\frac{2}{15}\left(\frac{3}{4}\right)^T \sum_{(d,10)=1} \frac{g_{T,3}(d)}{d}\right) x + \frac{4}{5}\left(\frac{3}{4}\right)^T \sum_{\substack{d \le \frac{x}{3} \\ (d,10)=1}} g_{T,3}(d). \end{aligned}$$

Let S_1' and S_2' denote the left and right sums, respectively. We have

$$\sum_{(d,10)=1} \frac{g_{T,3}(d)}{d} = \prod_{p \ge 7} \left(1 + \frac{g_{T,3}(p)}{p}\right) = \exp\left(\sum_{p \ge 7} \log\left(1 + \frac{g_{T,3}(p)}{p}\right)\right).$$

We choose $T = 29$ and as before, we split the sum at 10^9, getting

$$\sum_{p \ge 7} \log\left(1 + \frac{g_{T,3}(p)}{p}\right) < 4.85969.$$

This gives $S_1' < 0.004095x$. By Rankin's method, we have

$$\begin{aligned} \sum_{\substack{d \le \frac{x}{3} \\ (d,10)=1}} g_{T,3}(d) &\le \sqrt{\frac{x}{3}} \prod_{p \ge 7} \left(1 + \frac{g_{T,3}(p)}{\sqrt{p}}\right) \\ &= \sqrt{\frac{x}{3}} \exp\left(\sum_{p \ge 7} \log\left(1 + \frac{g_{T,3}(p)}{\sqrt{p}}\right)\right). \end{aligned}$$

Splitting the sum at 10^9 as above, we obtain $S_2' < 6.765\sqrt{x}$, so that $B_2(x) < 0.004095x + 6.765\sqrt{x}$.

We next turn to $B_3(x)$. Noting that the product of the primes to 19 is 9699690, we have

$$B_3(x) < \frac{1}{2^T} \sum_{\substack{n \le x \\ (n,9699690)=15}} \left(\frac{n}{\varphi(n)}\right)^T = \left(\frac{15}{16}\right)^T \sum_{\substack{n \le \frac{x}{15} \\ (n,646646)=1}} f_{15}(n)^T.$$

Note that $\sum_{d \mid n} g_{T,15}(d) = f_{15}(n)^T$ and $\varphi(646646) = 207360$. One finds via a computer search among numbers to 646646 that for any $t > 0$, the number of $d \le t$ coprime to 646646 is at most $207360t/646646 + 5.525$. We have as above that $B_3(x)$ is less than

$$\frac{207360}{646646}\left(\frac{15}{16}\right)^T \frac{x}{15}\prod_{p\ge 23}\left(1+\frac{g_{T,15}(p)}{p}\right)$$
$$+\, 5.525\left(\frac{15}{16}\right)^T \sqrt{\frac{x}{15}}\prod_{p\ge 23}\left(1+\frac{g_{T,15}(p)}{\sqrt{p}}\right).$$

Taking $T = 72$ and estimating the products as above, we find that $B_3(x) < 0.00182x + 661.201\sqrt{x}$.

Finally, we obtain an upper bound for $B_4(x)$. The conditions that $\gcd(n, 30) = 15$ and $\gcd(n, 323323) > 1$ put n in one of 115963 residue classes modulo 9699690. We find the optimal bound

$$B_4(x) \le \frac{115963}{9699690}x + \frac{204775}{38038} < 0.01196x + 5.3835$$

by a computer search to 9699690.

Combining our bounds for $B_i(x)$ proves the proposition. □

Remark. After work of Schoenberg [10] we know the density δ of numbers n with $\varphi(n)/n < 1/2$ exists, and the second author of this paper has calculated [6] that its value lies in the interval $(0.51120, 0.51176)$. Since every even number that is not a power of 2 satisfies this inequality, we have that $\delta - 1/2$ is the density of odd n with $\varphi(n)/n < 1/2$, that is, the density of the numbers counted by $N(x)$. We see that the bound of 0.017876 in Proposition 3.1 is not too far off from the asymptotically best possible estimate.

The following results can be proved in a similar way as we proved Proposition 3.1.

Proposition 3.2. *Let $M(x)$ denote the number of odd $m \le x$ such that $\varphi(m)/m < 0.5001$. We have $M(x) < 0.01794x + 680.18\sqrt{x} + 5.4$ for all $x > 0$. Moreover, for all $x > 0$ and $D > 0$, we have*

$$M(D + x) - M(D) < 0.01794x + 1360.36\sqrt{D + x} + 10.8.$$

The proof of Proposition 3.2 is nearly identical to that of Proposition 3.1 with the following changes. For the first assertion, the factor of $1/2^T$ is replaced with 0.5001^T. For the second assertion, the factor of $\sqrt{x}$ is replaced with $\sqrt{D + x}$. For example, in the case that m is coprime to 6, and $D = 0$, we get the bound

$$(4.91 \times 10^{-7})x + 2.5844\sqrt{x}, \tag{2}$$

which can be compared with our estimate for $B_1(x)$ in the proof of Proposition 3.1. Also, we replace the bound for case (1) by

$$|\{n \in (D, D + x] : \gcd(n, 6) = 1\}| \le x/3 + 4/3,$$

where the constant term is doubled due to the periodicity and symmetry of $\gcd(n, 6)$ as well as the right-continuity of $|\{n \le x : \gcd(n, 6) = 1\}| - x/3$, and similarly for cases (2)–(4). This change does not affect the constant in the main term but each of the constants of lower order will be double those of $M(x)$.

We will also use the following proposition.

Proposition 3.3. *Suppose that n is odd with $\varphi(n)/n < \frac{1}{2}$, $p \mid n$ with $p > 5000$ and $s \mid n+1$ with $s > 1$ and s coprime to 30030. The number of $n \le t$ with these properties is at most*

$$0.02194\frac{t}{ps} + 225\sqrt{\frac{t}{ps}} + 23.36\sqrt{\frac{t}{p}} + 38.$$

This estimate holds equally if the roles of n and $n+1$ are reversed.

Proof. The proof parallels that of Proposition 3.1, and in particular we have the same 4 cases. But here we replace "323323" with "1001".

Write $n = mp$ and $\varphi(n)/n < \frac{1}{2}$, so that $\varphi(m)/m < \frac{1}{2} + \epsilon$, where $\epsilon = 10^{-4}$. We first count the number of choices for $n \le t$ with $\gcd(n, 6) = 1$. This is at most the number of $m \le t/p$ coprime to 6, with $\varphi(m)/m < \frac{1}{2} + \epsilon$ and $mp \equiv -1 \pmod{s}$. Let b be an integer with $bp \equiv -1 \pmod{s}$, so that $m \equiv b \pmod{s}$. We have

$$N_1 := \sum_{\substack{m \le t/p \\ \gcd(m,6)=1 \\ m \equiv b \pmod{s} \\ \varphi(m)/m < \frac{1}{2}+\epsilon}} 1 \le \left(\frac{1}{2} + \epsilon\right)^T \sum_{\substack{m \le t/p \\ \gcd(m,6)=1 \\ m \equiv b \pmod{s}}} (m/\varphi(m))^T.$$

Since $\sum_{d \mid m} g_T(d) = (m/\varphi(m))^T$, we have

$$N_1 \le \left(\frac{1}{2} + \epsilon\right)^T \sum_{\substack{d \le t/p \\ \gcd(d,6s)=1}} g_T(d) \sum_{\substack{k \le t/pd \\ \gcd(k,6)=1 \\ k \equiv bd^{-1} \pmod{s}}} 1.$$

If $d > t/ps$, then $k < s$, so there is at most one k in the inner sum, and the contribution to the expression is at most

$$N_{1,1} := \left(\frac{1}{2} + \epsilon\right)^T \sum_{\substack{d \le t/p \\ \gcd(d,6)=1}} g_T(d). \tag{3}$$

The remaining part is at most

$$N_{1,2} := \left(\frac{1}{2} + \epsilon\right)^T \sum_{\substack{d \le t/ps \\ \gcd(d,6s)=1}} g_T(d) \sum_{\substack{k \le t/pd \\ \gcd(k,6)=1 \\ k \equiv bd^{-1} \pmod{s}}} 1.$$

The inner sum on k is at most $t/3psd + 4$, using an inclusion-exclusion on the 4 divisors of 6. (The "+4" can be improved here, but this is unimportant.) Thus,

$$\begin{aligned} N_{1,2} &\le \left(\frac{1}{2} + \epsilon\right)^T \sum_{\substack{d \le t/ps \\ \gcd(d,6)=1}} g_T(d) \left(\frac{t}{3psd} + 4\right) \\ &= \left(\frac{1}{2} + \epsilon\right)^T \frac{t}{3ps} \sum_{\substack{d \le t/ps \\ \gcd(d,6)=1}} \frac{g_T(d)}{d} + 4\left(\frac{1}{2} + \epsilon\right)^T \sum_{\substack{d \le t/ps \\ \gcd(d,6)=1}} g_T(d). \end{aligned}$$

With this expression and (3) we have 3 sums to estimate. We take $T = 69$. We have

$$\left(\frac{1}{2}+\epsilon\right)^T \frac{t}{3ps} \sum_{\substack{d\le t/ps\\ \gcd(d,6)=1}} \frac{g_T(d)}{d} < 4.91\times 10^{-7}\frac{t}{ps}.$$

Also,

$$4\left(\frac{1}{2}+\epsilon\right)^T \sum_{\substack{d\le t/ps\\ \gcd(d,6)=1}} g_T(d) \le 4\sqrt{\frac{t}{ps}}\left(\frac{1}{2}+\epsilon\right)^T \sum_{\gcd(d,6)=1} \frac{g_T(d)}{\sqrt{d}} < 15.51\sqrt{\frac{t}{ps}}.$$

Similarly,

$$N_{1,1} \le \sqrt{\frac{t}{p}}\left(\frac{1}{2}+\epsilon\right)^T \sum_{\gcd(d,6)=1} \frac{g_T(d)}{\sqrt{d}} < 3.88\sqrt{\frac{t}{p}}.$$

Summing up, we have

$$N_1 \le 4.91\times 10^{-7}\frac{t}{ps} + 15.51\sqrt{\frac{t}{ps}} + 3.88\sqrt{\frac{t}{p}}.$$

We next consider

$$N_2 := \sum_{\substack{m\le t/p\\ \gcd(m,30)=3\\ m\equiv b \pmod{s}\\ \varphi(m)/m<\frac{1}{2}+\epsilon}} 1 \le \left(\frac{3}{4}+\frac{3\epsilon}{2}\right)^T \sum_{\substack{m\le t/3p\\ \gcd(m,10)=1\\ m\equiv b' \pmod{s}}} f_3(m)^T.$$

Then, as with the work for N_1, we get

$$N_2 \le \left(\frac{3}{4}+\frac{3\epsilon}{2}\right)^T \left(\sum_{\substack{d\le t/3p\\ \gcd(d,10)=1}} g_{T,3}(d) + \sum_{\substack{d\le t/3ps\\ \gcd(d,10)=1}} g_{T,3}(d)\left(\frac{2}{5}\frac{t}{3psd}+4\right)\right).$$

Choosing $T = 29$, we get

$$N_2 \le 0.00412\frac{t}{ps} + 34.02\sqrt{\frac{t}{ps}} + 8.51\sqrt{\frac{t}{p}}.$$

We also have

$$N_3 := \sum_{\substack{m\le t/p\\ \gcd(m,30030)=15\\ m\equiv b \pmod{s}\\ \varphi(m)/m<\frac{1}{2}+\epsilon}} 1 \le \left(\frac{15}{16}+\frac{15\epsilon}{8}\right)^T \sum_{\substack{m\le t/15p\\ \gcd(m,2002)=1\\ m\equiv b' \pmod{s}}} f_{15}(m)^T.$$

We introduce $g_{T,15}$ and note that the number of integers to $t/15pd$ coprime to 2002 and in a residue class mod s is at most $24t/1001psd + 16$. So N_3 is at most

$$\left(\frac{15}{16}+\frac{15\epsilon}{8}\right)^T \left(\sum_{\substack{d\le t/15p\\ (d,2002)=1}} g_{T,15}(d) + \sum_{\substack{d\le t/15ps\\ (d,2002)=1}} g_{T,15}(d)\left(\frac{24t}{1001psd}+16\right)\right).$$

Choosing $T = 36$, we get

$$N_3 \le 0.00846\frac{t}{ps} + 175.47\sqrt{\frac{t}{ps}} + 10.97\sqrt{\frac{t}{p}}.$$

We next consider the case when $\gcd(n, 30) = 15$ and $\gcd(n, 1001) > 1$. In this case, the number of integers $n \le t$ is at most

$$\frac{1}{30} \cdot \frac{281}{1001}\frac{t}{ps} + 38.$$

Putting our estimates together, we complete the proof. □

4. Proof of Theorem 1.1.

Recall that $X_0 = e^{150}$. We partition solutions of $\varphi(n) = \varphi(n+1)$ into a small range $n \le 10^{13}$, middle range $10^{13} < n < X_0$, and large range $n > X_0$.

4.1. The Small Range, $n \le 10^{13}$.

By computation using an exhaustive list of all 10755 solutions up to 10^{13} (see [8]) we have

$$\sum_{\substack{n\in\mathcal{S} \\ n\le 10^{13}}} \frac{1}{n} = 1.432488\ldots.$$

4.2. The Middle Range, $10^{13} < n \le X_0$.

It is shown in [1, Prop. 2.2] that for solutions to $\varphi(n) = \varphi(n+1)$ larger than 2^{32}, the odd member of the pair, say n, satisfies $\varphi(n)/n < \frac{1}{2}$. It follows then via partial summation and doubling the estimate in Proposition 3.1 (to allow for the possibility that an odd number n may be in two pairs of numbers with equal φ-values) that

$$\sum_{\substack{n\in\mathcal{S} \\ 10^{13}<n\le X_0}} \frac{1}{n} = \frac{S(X_0) - S(10^{13})}{X_0} + \int_{10^{13}}^{X_0} \frac{S(t) - S(10^{13})}{t^2}dt < 4.3293.$$

However, we can do a little better as follows.

Consider odd numbers $n > 10^{13}$ divisible by 105. These are part of case (4) in the proof of Proposition 3.1 and according to the accounting there, the number of them in $[1, x]$ is at most $x/210 + 1$. However, the number of these with $\varphi(n) = \varphi(m)$ with $m = n \pm 1$ is considerably smaller. Note that m is even and $\varphi(m)/m \le (1 + 1/m)\varphi(105)/105$. Further, $m \equiv \pm 1 \pmod{105}$. Fix $a = \pm 1$ and let $B(x)$ denote the number of such numbers $m \le x$ with $n \equiv a \pmod{105}$. Since $\varphi(105)/105 = 16/35$ and letting $\epsilon = 10^{-13}$, we have for any $T > 0$,

$$B(x) = \sum_{\substack{m\le x \\ 2\,|\,m \\ m\equiv a \pmod{105} \\ \varphi(m)/m<\frac{16}{35}(1+\epsilon)}} 1 \le \left(\frac{16}{35}(1+\epsilon)\right)^T \sum_{\substack{m\le x \\ 2\,|\,m \\ m\equiv a \pmod{105}}} \left(\frac{m}{\varphi(m)}\right)^T = \left(\frac{32}{35}+\epsilon\right)^T \sum_{\substack{l\le x/2 \\ l\equiv b \pmod{105}}} f_2(n)^T,$$

where b is such that $2b \equiv a \pmod{105}$. Thus,

$$B(x) \le \left(\frac{32}{35}+\epsilon\right)^T \sum_{\substack{d\le x/2\\(d,105)=1}} g_{T,2}(d) \sum_{\substack{k\le x/2d\\ k\equiv bd^{-1} \pmod{105}}} 1.$$

The inner sum is at most $x/210d+1$, so that

$$B(x) \le \left(\frac{32}{35}+\epsilon\right)^T \sum_{\substack{d\le x/2\\(d,105)=1}} g_{T,2}(d)\frac{x}{210d} + \left(\frac{32}{35}+\epsilon\right)^T \sum_{\substack{d\le x/2\\(d,105)=1}} g_{T,2}(d)$$

Choosing $T=18$ and using the methods of Section 3, we have

(4) $$B(x) \le 0.002578x + 7.7\sqrt{x}.$$

Subtracting $x/210-1/2$ from the estimate in Proposition 3.1, adding in the estimate in (4), and doubling, we have

$$S(x)-S(10^{13}) \le 0.031385x + 1360\sqrt{x} + 11.3.$$

We thus have

(5) $$\sum_{\substack{n\in\mathcal{S}\\10^{13}<n\le X_0}} \frac{1}{n} \le 3.8006.$$

4.3. **The Large Range,** $n > X_0$**.** Here is the plan for the proof. Let $n \in \mathcal{S}$. We show that, but for a small number of exceptions, $P(n)$ and $P(n+1)$ are large and that neither n nor $n+1$ is divisible by a large proper power of a prime. We then deal with the situation when the largest prime q dividing $n(n+1)$ is very large (approximately, it is $> n^{0.3}$). Here we consider the two cases: $P(q-1)$ is large and $P(q-1)$ is small. Finally, we have the situation when q is not so large. Here we concentrate on the odd member of the pair, doubling our estimate since we do not know which of n, $n+1$ is odd. The advantage to us of working with the odd member is that we can bring in Proposition 3.3 to help with the estimate.

Let $I_k = (e^k, e^{k+1})$ and $\mathcal{S}_k = I_k \cap \mathcal{S}$. Let $\alpha_k = 3.5$ for $150 \le k < 400$ and $\alpha_k = 4$ for $k \ge 400$. Let $\beta_k = 4$ for $150 \le k < 200$, $\beta_k = 4.5$ for $200 \le k < 400$, and $\beta_k = 5$ for $k \ge 400$. Let

$$x_k = e^{k/\lfloor \alpha_k \log k\rfloor},\ x_k' = e^{0.3k}, \quad z_k = e^{\sqrt{k}/\beta_k},\ z_k' = e^{0.7\sqrt{k}}.$$

Also, let

$$x' = x'(t) = x'_{\lfloor \log t\rfloor}, \quad z' = z'(t) = z'_{\lfloor \log t\rfloor}.$$

Define the following sets of natural numbers:

$$\mathcal{C}_0^k = \{n \in \mathcal{S}_k : q^a | n(n+1) \text{ for some } a\ge 2, \text{ where } q^a > x_k \text{ or } q > z_k'\},$$
$$\mathcal{C}_1^k = \{n \in \mathcal{S}_k : \omega(n) \text{ or } \omega(n+1) \ge \alpha_k \log\lfloor \log n\rfloor\},$$
$$\mathcal{C}_2^k = \mathcal{S}_k \setminus (C_0^k \cup \mathcal{C}_1^k).$$

We will use the convention $\mathcal{C}_i = \bigcup_{k\ge 150} \mathcal{C}_i^k$. We first bound the contribution to the reciprocal sum from $\mathcal{C}_0$.

Proposition 4.1. *We have*

$$\sum_{n\in\mathcal{C}_0}\frac{1}{n}<0.2516.$$

Proof. We handle the case when $q^a \mid n$ and double the estimate to allow for the parallel case $q^a \mid n+1$. Let $T_k=\{q^a : a\ge 2, q^a>x_k\}$. By [7, Lem. 2.2], we have

$$\sum_{k\ge 150}\sum_{\substack{e^k<n\le e^{k+1}\\ \exists s\in T_k: s\mid n}}\frac{1}{n}\le\sum_{k\ge 150}\sum_{\substack{s\in T_k\\ s\le e^{k+1}}}\frac{1}{s}+\sum_{k\ge 150}\sum_{\substack{s\in T_k\\ s\le e^{k+1}}}\frac{1}{e^k}.$$

The right sum is

$$\sum_{k\ge 150}\frac{1}{e^k}\sum_{\substack{s\in T_k\\ s\le e^{k+1}}}1\le\sum_{k\ge 150}\frac{e^{(k+1)/2}}{e^k}=\frac{1}{(\sqrt{e}-1)e^{74}}<2\cdot 10^{-32}.$$

Here we used inequality (3.7) in the proof of [7, Prop. 3.3] to bound the number of proper prime powers up to t as less than $t^{1/2}$ for $t\ge 200$. For the left sum, we use Corollary 2.8 to bound

$$\sum_{k\ge 150}\sum_{s\in T_k}\frac{1}{s}\le\sum_{k\ge 150}\frac{2.4}{\sqrt{x_k}\log x_k}.$$

Computing the sum directly to $k=10^8$ and bounding the remaining sum with an integral, this expression is less than $0.12345+0.00155=0.12500$, the two numbers coming from the ranges $150\le k\le 399$ and $k\ge 400$, respectively.

We proceed in the same way, but now use Lemma 2.7 and $T'_k=\{q^2 : q>z'_k\}$. The reciprocal sum is bounded above by

$$\sum_{k\ge 150}\sum_{\substack{e^k<n\le e^{k+1}\\ \exists s\in T'_k: s\mid n}}\frac{1}{n}\le\sum_{k\ge 150}\sum_{\substack{s\in T'_k\\ s\le e^{k+1}}}\frac{1}{s}+\sum_{k\ge 150}\sum_{\substack{s\in T'_k\\ s\le e^{k+1}}}\frac{1}{e^k}.$$

By Lemma 2.7, we compute that this expression is smaller than 0.00079. Noting that $2(0.12500+0.00079)<0.2516$, completes the proof. □

Proposition 4.2. *We have*

$$\sum_{n\in\mathcal{C}_1}\frac{1}{n}<0.1430.$$

Proof. As before, we treat the case of n, doubling the estimate to account for the case of $n+1$. Following [7, Prop. 3.2], we have $\tau_5(n)\ge 5^{\omega(n)}$, where $\tau_5(n)$ denotes the number of ordered factorizations of n into five positive integers. By [7, Lem. 2.5] we have

$$\begin{aligned}\sum_{\substack{e^{150}<n<e^{400}\\ \omega(n)\geq 3.5\log\lfloor\log n\rfloor}}\frac{1}{n} &= \sum_{151\leq k\leq 400}\ \sum_{\substack{e^{k-1}<n<e^{k}\\ \omega(n)\geq 3.5\log(k-1)}}\frac{1}{n}\\ &\leq \sum_{151\leq k\leq 400} 5^{-3.5\log(k-1)}\sum_{n<e^k}\frac{\tau_5(n)}{n}\\ &\leq \sum_{151\leq k\leq 400}\frac{1}{120}\frac{(k+5)^5}{(k-1)^{3.5\log 5}} < 0.07006.\end{aligned}$$

Note that this sum, if extended to infinity, diverges. However, by changing 3.5 to 4, the sum converges, and we have

$$\sum_{\substack{n>e^{400}\\ \omega(n)\geq 4\log\lfloor\log n\rfloor}}\frac{1}{n}\leq \sum_{k\geq 401}\frac{1}{120}\frac{(k+5)^5}{(k-1)^{4\log 5}} < 0.00142.$$

Noting that $2(0.07006+0.00142) < 0.1430$, the proof is complete. □

For $n\in\mathcal{C}_2^k$, we may assume that $\omega(n) < \alpha_k\log\lfloor\log n\rfloor$, since $n\notin\mathcal{C}_1$. Therefore, the largest prime power dividing n exceeds $n^{1/\lfloor\alpha_k\log\lfloor\log n\rfloor\rfloor} > e^{k/\lfloor\alpha_k\log k\rfloor}$. It follows that this prime exactly divides n since $n\notin\mathcal{C}_0$, so that $P(n) > x_k$ and $P(n)\,\|\,n$. These conclusions hold as well for $n+1$.

We use the notation $q = P(n(n+1))$ and $p = P(n)$. We define the following sets:

$$\begin{aligned}\mathcal{C}_3^k &= \{n\in\mathcal{C}_2^k : q > x_k',\ P(q-1)\leq z_k'\},\\ \mathcal{C}_4^k &= \{n\in\mathcal{C}_2^k : q > x_k',\ P(q-1) > z_k'\},\\ \mathcal{C}_5^k &= \{n\in\mathcal{C}_2^k\setminus(\mathcal{C}_3\cup\mathcal{C}_4) : P(p-1)\leq z_k\},\\ \mathcal{C}_6^k &= \{n\in\mathcal{C}_2^k\setminus(\mathcal{C}_3\cup\mathcal{C}_4) : P(p-1) > z_k\}.\end{aligned}$$

We continue with the convention $\mathcal{C}_i = \cup_{k\geq 150}\mathcal{C}_i^k$.

Proposition 4.3. *We have*

$$\sum_{n\in\mathcal{C}_3}\frac{1}{n} < 0.2543.$$

Proof. Write the one of $n, n+1$ which is a multiple of q as qm. We will sum $1/qm$ and double the estimate to allow for the ambiguity of whether $q\mid n$ or $q\mid n+1$. We first consider the case that $q > e^{0.45k}$. Let $S(x,y)$ denote the reciprocal sum of those integers $j > x$ with $P(j)\leq y$. By [7, Lem. 2.2, 2.10],

$$\begin{aligned}\sum_{k\geq 150}\sum_{\substack{q>e^{0.45k}\\ P(q-1)\leq z_k'}}\frac{1}{q}\sum_{\frac{e^k}{q}<m<\frac{e^{k+1}}{q}}\frac{1}{m} &\leq \sum_{k\geq 150}\frac{1}{2}S\left(\frac{e^{0.45k}-1}{2}, z_k'\right)(1+e/2)\\ &< 0.00063(1+e/2),\end{aligned}$$

noting that $q-1$ is even. Also, we bound

$$\sum_{k\ge 150}\sum_{\substack{x'_k<q<e^{0.45k}\\ P(q-1)\le z'_k}}\frac{1}{q}\sum_{\frac{e^k}{q}<m<\frac{e^{k+1}}{q}}\frac{1}{m}\le\sum_{k\ge 150}\frac{1}{2}S\left(\frac{x'_k-1}{2},z'_k\right)\left(1+e^{-0.55k}\right)$$
$$<0.12564.$$

Here we used [7, Lem. 2.10] to sum over $k\ge 300$, obtaining a bound of 0.00801, and [7, Lem. 2.9] with $s_k=\log(e^{0.2}u_k\log u_k)/\log z'_k$ to sum over $150\le k\le 299$, obtaining a bound of 0.11763. Combining and doubling, we complete the proof of Proposition 4.3. □

Proposition 4.4. *We have*

$$\sum_{n\in\mathcal{C}_4}\frac{1}{n}<0.8542.$$

Proof. Let $n\in\mathcal{C}_4$. Since $n\notin\mathcal{C}_2$, $r=P(q-1)\mid\varphi(n)$, and $r>z'_{\lfloor\log n\rfloor}$, there are primes p,p' with $q=\max\{p,p'\}$, $p\|n$, $p'\|n+1$ and $p\equiv p'\equiv 1\pmod r$. Writing $n=pm$ and $n+1=p'm'$, we have $pm+1=p'm'$ and $(p-1)\varphi(m)=(p'-1)\varphi(m')$. Thus,

$$p'(m'\varphi(m)-m\varphi(m'))=(m+1)\varphi(m)-m\varphi(m').$$

If the left side is zero, consider that since $\gcd(m,m')=1$, we would then have $m\mid\varphi(m)$ and $m'\mid\varphi(m')$, so that $m=m'=1$. But this does not occur for $n>1$, so the left side is not zero. Therefore p' (and also p) are fixed by the ordered pair (m,m'), so that n is completely determined by the pair (m,m').

Let $\mathcal{A}(t)=\{n\le t:n\in\mathcal{C}_4\}$ and let $y_k=ke^k\sqrt{z'_k}/20$ and $y=y_{\lfloor\log t\rfloor}$. Then

$$\mathcal{A}(t)=\mathcal{A}_1(t)\cup\mathcal{A}_2(t),$$

where

$$\mathcal{A}_1(t)=\{n\in\mathcal{A}(t):pp'\le y\}\text{ and }\mathcal{A}_2(t)=\{n\in\mathcal{A}(t):pp'>y\}.$$

Let $A_i(t)$ denote the cardinality of $\mathcal{A}_i(t)$, $i=1,2$. The system of congruences $n\equiv 0\pmod p$, $n+1\equiv 0\pmod{p'}$ has a unique solution n modulo pp' by the Chinese remainder theorem. Thus,

$$A_1(t)\le\sum_{r>z'}\sum_{\substack{pp'\le y\\ \max\{p,p'\}>x'\\ p\equiv p'\equiv 1\,(r)}}\left(\frac{t}{pp'}+1\right).$$

For a prime $r>z'$ let

$$v_r=\sum_{\substack{pp'\le y\\ \max\{p,p'\}>x'\\ p\equiv p'\equiv 1\,(r)}}\frac{t}{pp'}.$$

Let $x'' = x'^{0.4} = e^{0.12\lfloor \log t \rfloor}$. Consider the case when $r > x''$. We have

$$\sum_{r>x''} v_r \le \sum_{r>x''} t \left(\sum_{j<t/2r} \frac{1}{2jr} \right)^2 < \sum_{r>x''} \frac{t}{(2r)^2} (\log(t/2r)+1)^2$$
$$\le \sum_{r>x''} \frac{t}{(2r)^2} (\log(t/2x'')+1)^2 \le \frac{t(\log(t/2x'')+1)^2}{4x'' \log x''},$$

using Lemma 2.7. Applying partial summation, the contribution to the reciprocal sum is $< 5 \times 10^{-5}$. Now assume that $r \in (z', x'']$. We have

(6) $$v_r \le 2 \sum_{\substack{x'<p<y/(2r) \\ p \equiv 1\,(r)}} \sum_{\substack{p' \le y/x' \\ p' \equiv 1\,(r)}} \frac{t}{pp'},$$

doubled because we assume $p > x'$. We have by Lemma 2.3 and Corollary 2.5 that

$$\sum_{\substack{x'<p \le y/2r \\ p \equiv 1\,(r)}} \frac{1}{p} \le \frac{s_1(r)}{r}, \qquad \sum_{\substack{p' \le y/x' \\ p' \equiv 1\,(r)}} \frac{1}{p'} \le \frac{s_2(r)}{r},$$

where

$$s_1(r) = \frac{2r}{r-1} \Big(\log\log \frac{y}{(2r)^2} - \log\log \frac{x'}{2r} + \frac{1}{\log(y/(2r)^2)} \Big),$$
$$s_2(r) = \frac{2r}{r-1} \Big(\log\log \frac{y}{2rx'} - 0.78169 + \frac{1}{\log(y/(2rx'))} \Big).$$

We assemble these estimates into (6). Note that $s_1(r)s_2(r)$ is increasing in the variable r for $z' < r \le x''$. Let $x''_k = x''(e^k) = e^{0.12k}$. Via partial summation, we have the reciprocal sum in this case at most

$$2 \sum_{k \ge 150} \sum_{z'_k < r \le x''_k} \frac{s_1(r)s_2(r)}{r^2} \le 2 \sum_{k \ge 150} \sum_{r > z'_k} \frac{s_1(x''_k)s_2(x''_k)}{r^2} \le \sum_{k \ge 150} \frac{2s_1(x''_k)s_2(x''_k)}{z'_k \log z'_k},$$

using Lemma 2.7. We have the contribution to the reciprocal sum for $r \in (z', x'']$ is less than 0.04665.

We next estimate the sum of the error term 1. This is

(7) $$2 \sum_{r>z'} \sum_{\substack{p'<y/x' \\ p' \equiv 1\,(r)}} \sum_{\substack{x'<p \le y/p' \\ p \equiv 1\,(r)}} 1.$$

Writing $p = 2ar+1$, $p' = 2br+1$, the contribution when $r > x''$ is at most

$$2 \sum_{r>x''} \sum_{ab \le y/4r^2} 1 \le \sum_{r>x''} \frac{y}{2r^2} \Big(\log \frac{y}{4r^2} + 1 \Big) < \frac{y(\log(y/4x''^2)+1)}{2x'' \log x''},$$

using Lemma 2.7 and the elementary estimate that the number of pairs a, b with $ab \le x$ is at most $x \log x + x$. Dividing our expression by t and integrating from X_0 to ∞, we get less than 0.00030.

So now we assume that $z' < r \le x''$. Using the Brun–Titchmarsh inequality, the inner sum in (7) is at most $2(y/p')/((r-1)\log(y/(p'r)))$. Note that not both $2r+1$, $4r+1$ can be prime, since one of them is divisible by 3. Thus, the contribution to (7) when $p' \le 6r$ is at most

$$\sum_{z'<r\le x''} \frac{4y}{(2r+1)(r-1)\log(y/((2r+1)r))} < \frac{2y}{\log(y/2x''^2)}\left(1+\frac{1}{z'}\right)\sum_{r>z'}\frac{1}{r^2}.$$

Using Lemma 2.7 and partial summation, the contribution to the reciprocal sum in this case is less than 0.01432. We now assume that $p' > 6r$ in (7). We find that for a given r, the expression is at most

$$\frac{8y}{r^2}\frac{r^2}{(r-1)^2}(A+B),$$

where $A = 1/(\log(y/x'r)\log(x'/r))$ and

$$B = \frac{1}{\log(y/r^2)}\big(\log\log(y/x'r) - \log\log(x'/r) - \log\log 6 + \log\log(y/6r^2)\big).$$

Using that $(1+1/(r-1))^2(A+B)$ is increasing in r on $(z', x'']$, using partial summation and Lemma 2.7 we get that the contribution to the reciprocal sum is less than 0.33245.

We next consider an upper bound for $A_2(t)$. If $n \in \mathcal{A}_2(t)$ then $pp' > y$, and since $pp'mm' = n(n+1) \le t(t+1)$, we have

$$mm' < t(t+1)/y = 20t(t+1)/(ke^k\sqrt{z_k'}) = w = w(t), \text{ say.}$$

Further, one of m, m' is odd and the other is even, so assume m is odd, m' is even. We double our estimate to take into account the other possibility. There are two cases: $3 \mid m$ and $3 \nmid m$. Let $\mathcal{A}_{2,1}(t)$ denote the set of such ordered pairs (m, m') when $3 \mid m$, and $\mathcal{A}_{2,2}(t)$ the set of such pairs with $3 \nmid m$. Let $A_{2,i}(t)$ denote their cardinalities for $i = 1, 2$, respectively. Since the pair (m, m') fixes p and p' (and therefore n), we have

$$A_2(t) = A_{2,1}(t) + A_{2,2}(t).$$

Note that $p \ge 2r+1 > 2z'+1 > 5000$ since $p \equiv 1 \pmod r$ and $r > z'$. Thus, $\varphi(m)/m < 0.5001$, so we may apply the averaging argument in Proposition 3.2. Since m, m' are coprime,

$$A_{2,1}(t) \le 2 \sum_{\substack{m\le w\\ \gcd(m,6)=3\\ \varphi(m)/m<0.5001}} \sum_{\substack{m'\le w/m\\ 2\mid m'\\ 3\nmid m'}} 1 \le \frac{2}{3} \sum_{\substack{m\le w\\ \gcd(m,6)=3\\ \varphi(m)/m<0.5001}} \left(\frac{w}{m}+2\right),$$

$$A_{2,2}(t) \le 2 \sum_{\substack{m\le w\\ \gcd(m,6)=1\\ \varphi(m)/m<0.5001}} \sum_{\substack{m'\le w/m\\ 2\mid m'}} 1 \le \sum_{\substack{m\le w\\ \gcd(m,6)=1\\ \varphi(m)/m<.5001}} \frac{w}{m}$$

Letting $M_1(x)$ be the number of $m \le x$ with $\gcd(m, 6) = 1$ and $\varphi(m)/m < 0.5001$ and noting that the first such m is $m_1 := 37182145$, we have from (2) and partial summation that

$$\sum_{\substack{m\le w\\ \gcd(m,6)=1\\ \varphi(m)/m<0.5001}} \frac{1}{m} = \frac{M_1(w)}{w} + \int_{m_1}^{w} \frac{M_1(x)}{x^2}dx$$

$$\le 5\times 10^{-7} + 5\times 10^{-7}(\log w - \log m_1) + \frac{2\cdot 2.6}{\sqrt{m_1}}$$

$$< 5\times 10^{-7}\log w + 8.6\times 10^{-4}.$$

For the sum of $w/m+2$ for $m \le w$, $\gcd(m,6)=3$, and $\varphi(m)/m < 0.5001$, we use Proposition 3.2, and relax the condition $\gcd(m,6)=3$ to $\gcd(m,2)=1$. Computing directly the sum of $1/m$ to 10^{10}, an upper bound for the sum is 0.21322. Thus,

$$\sum_{\substack{m\le w\\ \gcd(m,2)=1\\ \varphi(m)/m<.5001}} \frac{1}{m} < 0.01794\log w - 0.172656.$$

Further, using $w \ge 2\times 10^{62}$ and Proposition 3.2, we have the number of integers m in the sum at most $0.01795w$. Thus,

$$A_2(t) < 0.0119605w\log w - 0.09031w.$$

The contribution to the reciprocal sum from this term is at most

$$\int_{X_0}^{\infty} \frac{1}{t^2}A_2(t)dt < \sum_{k\ge 150}\int_{e^k}^{e^{k+1}} \frac{1}{t^2}(0.0119605w\log w - 0.09031w)dt < 0.46042.$$

Combining these bounds, we complete the proof of Proposition 4.4. □

Proposition 4.5. *We have*

$$\sum_{n\in\mathcal{C}_5}\frac{1}{n} < 0.2790.$$

Proof. Assume that $n \in \mathcal{C}_5$ and write $n = pm$. We also assume that n is odd. The case when n is even is completely parallel, so we double our estimates to reflect this case. We bound the reciprocal sum for $x_k < p \le x_k'$ and $r = P(p-1) \le z_k$ by

$$\sum_{k\ge 150}\sum_{\substack{x_k<p\le x_k'\\ P(p-1)\le z_k}} \frac{1}{p} \sum_{\substack{\frac{e^k}{p}<m<\frac{e^{k+1}}{p}\\ m \text{ odd},\ \varphi(m)/m<0.5001}} \frac{1}{m},$$

noting that $\varphi(m)/m < p/(2(p-1)) < 0.5001$ for $p > x_{150}$. We first bound the inner sum. Recall that $M(x) = |\{m \le x : 2 \nmid m,\ \varphi(m)/m < 0.5001\}|$. Let $D = e^k/p$. By partial summation,

$$\sum_{\substack{\frac{e^k}{p}<m<\frac{e^{k+1}}{p}\\ m \text{ odd},\ \varphi(m)/m<0.5001}} \frac{1}{m} = \frac{M(De)}{De} - \frac{M(D)}{D} + \int_{D}^{De}\frac{M(t)}{t^2}dt.$$

Let $a = 0.01794$, $b = 1360.36$, $c = 10.8$. By Proposition 3.2,

$$\frac{M(De)}{De} < \frac{M(D) + a(e-1)D + b\sqrt{De} + c}{De}$$

and

$$\int_D^{De} \frac{M(t)}{t^2} dt < \frac{M(D)}{D} - \frac{M(D)}{De} + \frac{a}{e} - \frac{2b}{\sqrt{De}} + \frac{2b}{\sqrt{D}} - \frac{c}{De} + \frac{c}{D}.$$

Combining terms, the sum over m is less than 0.01795.

Turning to the sum over p, we first bound this sum over $k \geq 2000$. Using the notation of [7, Lem. 2.10] and observing that $p-1$ is even,

$$\begin{aligned}\sum_{k\geq 2000} \sum_{\substack{p>x_k\\ P(p-1)\leq z_k}} \frac{1}{p} &\leq \frac{1}{2} \sum_{k\geq 2000} S\left(\frac{x_k - 1}{2}, z_k\right)\\ &< \frac{1}{2} \sum_{k\geq 2000} \frac{20e^{(1+\epsilon)u_k}(2^{\log(u_k \log u_k)/\log z_k} - 1)^{-1}}{(u_k \log u_k)^{u_k}}\\ &< 0.04598,\end{aligned}$$

where $\epsilon = 2.3 \cdot 10^{-8}$ and $u_k = \log((x_k - 1)/2)/\log z_k$. Here we computed the sum over $2000 \leq k \leq 10^8$ directly and then compared the remaining series to an integral. Using the first inequality of [7, Lem. 2.9] with $s_k = \log(e^\gamma u_k \log u_k)/\log z_k$ and noting that $p-1$ is even, we have

$$\sum_{1500\leq k\leq 1999} \sum_{\substack{p>x_k\\ P(p-1)\leq z_k}} \frac{1}{p} \leq \frac{1}{2} \sum_{1500\leq k\leq 1999} S\left(\frac{x_k - 1}{2}, z_k\right) < 0.010329.$$

We next sum over $1000 \leq k \leq 1499$, $700 \leq k \leq 999$, $556 \leq k \leq 699$, with parameters $s_k = \log(e^c u_k \log u_k)/\log z_k$, $c = 0.5, 0.45, 0.4$, to obtain bounds 0.120102, 0.643079, 1.211382, respectively.

Finally, for the interval $150 \leq k \leq 555$, we directly evaluate the sum of reciprocals of even z_k-smooth numbers $p - 1 > x_k - 1$ as follows. The sum of reciprocals of all even z_k-smooth numbers is equal to $\prod_{3\leq p\leq z_k} \frac{p}{p-1}$. For each $150 \leq k \leq 555$ we subtract from this quantity the sum of reciprocals of even z_k-smooth numbers not exceeding $x_k - 1$. Summing over $150 \leq k \leq 199$, $200 \leq k \leq 399$, $400 \leq k \leq 555$, we obtain the bounds 3.439039, 1.941653, 0.35777, respectively.

Summing these bounds, multiplying by 0.01795, and doubling, we complete the proof of Proposition 4.5. □

Proposition 4.6. *We have*

$$\sum_{n\in\mathcal{C}_6} \frac{1}{n} < 0.6320.$$

Proof. We assume that n is odd and double the bound, noting that a symmetric argument applies to the case that $n+1$ is odd. Recall that $p = P(n)$. There is a prime $r > z_{\lfloor \log n \rfloor}$ such that $r \mid p-1$, and thus $r|\varphi(n) = \varphi(n+1)$. Either $r^2 \mid n+1$ or there is a prime $p' \mid n+1$ with $p' \equiv 1 \pmod r$. In this proof we let the letter s denote either r^2 or p'. Since $n \notin \mathcal{C}_0, \mathcal{C}_4$, we have $s \le x'_{\lfloor \log n \rfloor}$. Consider the counting function of such $n \le t$. Noting that $p, p', r^2 < t^{0.3}$ and applying Proposition 3.3, we find that the counting function is bounded above by

$$\sum_{r>z} \sum_{\substack{x<p\le x' \\ p\equiv 1\,(r)}} \sum_{s\le x'} \left(\frac{0.02194t}{ps} + 225\sqrt{\frac{t}{ps}} + 23.36\sqrt{\frac{t}{p}} + 38 \right),$$

where $x = x_{\lfloor \log t \rfloor}$, $z = z_{\lfloor \log t \rfloor}$, and s runs over primes $p' \equiv 1 \pmod r$ or $s = r^2$. For $t \ge X_0$ and $p, s \le x' \le t^{0.3}$, we have

$$23.36\sqrt{\frac{t}{p}} + 38 < 23.37\sqrt{\frac{t}{p}}, \quad 0.02194\frac{t}{ps} + 225\sqrt{\frac{t}{ps}} < 0.02195\frac{t}{ps}.$$

Decoupling the possibilities for s, our counting function is majorized by $S_1 + S_2$, where

$$S_1 = \sum_{r>z} \sum_{\substack{x<p\le x' \\ p\equiv 1\,(r)}} \left(\frac{0.02195t}{pr^2} + 23.37\sqrt{\frac{t}{p}} \right),$$

$$S_2 = \sum_{r>z} \sum_{\substack{x<p\le x' \\ p\equiv 1\,(r)}} \sum_{\substack{p'\le x' \\ p'\equiv 1\,(r)}} \left(\frac{0.02195t}{pp'} + 23.37\sqrt{\frac{t}{p}} \right).$$

We can make a further consolidation in S_1, since $n \notin \mathcal{C}_0$ implies that $r < z'$. Thus, for $t > X_0$, we have

$$S_1 < \sum_{r>z} \sum_{\substack{x<p\le x' \\ p\equiv 1\,(r)}} \frac{0.02196t}{pr^2}.$$

We use Lemma 2.3 to sum $1/p$, Lemma 2.7 to sum $1/r^2$, and we majorize $1/(r-1)$ (from Lemma 2.3) with $1/(z-1)$. After partial summation to extract the reciprocal sum from the counting function, we have a contribution of at most

$$2(0.00206 + 0.00328 + 0.00085) = 0.01238$$

to the reciprocal sum. (The three terms correspond to the three expressions for z_k.)

We now turn to S_2. Via partial summation, the reciprocal sum of integers counted by S_2 is bounded by

$$\sum_{k\ge 150} \sum_{r>z_k} \sum_{\substack{x_k<p\le x'_k \\ p\equiv 1\,(r)}} \sum_{\substack{p'\le x'_k \\ p'\equiv 1\,(r)}} \left(\frac{0.02195}{pp'} + 23.37e^{-k/2}\frac{1}{\sqrt{p}} \right).$$

For the the term involving $1/pp'$, let

$$P(k,r) = \sum_{\substack{x_k<p<x'_k \\ p\equiv 1\,(r)}} \frac{1}{p}, \quad Q(k,r) = \sum_{\substack{p'\le x'_k \\ p'\equiv 1\,(r)}} \frac{1}{p'}.$$

We split up the range for the variables k, r into 3 regions:

- $k \ge 1258$,
- $150 \le k \le 1257$, $r \ge 1201$,
- $150 \le k \le 1257$, $r < 1201$.

In the first region, for each k we segment the interval of primes $r > z_k$ into intervals $(100^{j-1}z_k, 100^j z_k]$ for j such that $100^j z_k < 100x_k^{1/2}$. In each of these intervals we use Lemma 2.3 to bound $P(k,r)$ and Corollary 2.5 to bound $Q(k,r)$. In doing this, note that our bound for $r^2 P(k,r)Q(k,r)$ is increasing in r on each interval, so we replace r in the expression with the upper bound of the interval, and then use Lemma 2.7 to bound the sum of $1/r^2$ in each interval. After applying partial summation, multiplying by 0.02195, and doubling, we get a contribution of less than 0.09481 to the reciprocal sum. For larger values of r, we use Corollary 2.5 to majorize both $P(k,r)$ and $Q(k,r)$, so that now the upper bound for $r^2 P(k,r)Q(k,r)$ is decreasing in r. After using Lemma 2.7 for the sum on r, summing on k, and performing the requisite doubling, we find that the contribution to the reciprocal sum is less than 2.2×10^{-9}. For the second region we proceed in a similar manner, except that we use $x_k^{0.6}$ instead of $x_k^{1/2}$ and for the interval when $j = 1$ we use the lower bound 1201 for r (in the upper intervals, it is larger). We get an upper bound, after doubling,

$$0.003893 + 0.011933 + 0.010115 < 0.02595,$$

the 3 numbers corresponding to the changing choices for x_k, z_k.

For the third region, we use Lemma 2.6. Write $P(k,r) = P_1(k,r) + P_2(k,r)$ and $Q(k,r) = Q_1(k.r) + Q_2(k,r)$, where

$$P_1(k,r) = \sum_{\substack{x_k < p \le 50r^2 \\ p \equiv 1\,(r)}} \frac{1}{p}, \quad P_2(k,r) = \sum_{\substack{\max\{x_k, 50r^2\} < p \le x'_k \\ p \equiv 1\,(r)}} \frac{1}{p}$$

$$Q_1(k,r) = \sum_{\substack{p' \le 50r^2 \\ p' \equiv 1\,(r)}} \frac{1}{p'}, \quad Q_2(k,r) = \sum_{\substack{50r^2 < p' \le x'_k \\ p' \equiv 1\,(r)}} \frac{1}{p'}$$

Since $r < 1201$ and $k \le 1257$, we can compute P_1 and Q_1 directly, and as mentioned, we use Lemma 2.6 on the remaining sums. We get that the contribution to the reciprocal sum is at most

$$2(0.04007 + 0.11131 + 0.09799) = 0.49874.$$

To complete the proof, we deal with

$$\sum_{k \ge 150} \sum_{r > z_k} 23.37 e^{-k/2} \sum_{\substack{x_k < p \le x'_k \\ p \equiv 1\,(r)}} \sum_{\substack{p' \le x'_k \\ p' \equiv 1\,(r)}} \frac{1}{\sqrt{p}}.$$

Using the Brun–Titchmarsh inequality and partial summation, we have

$$\sum_{\substack{x_k < p \le x'_k \\ p \equiv 1\,(r)}} \frac{1}{\sqrt{p}} < \frac{2\sqrt{x'_k}}{(r-1)\log(x'_k/r)} + \frac{\sqrt{r}}{r-1} \mathrm{li}\big(\sqrt{x'_k/r}\big),$$

where li is the logarithmic integral function. Splitting the sum on r at $e^{\sqrt{k}}$, we have the contribution here smaller than

$$2(2.2 \times 10^{-5} + 9 \times 10^{-7} + 4 \times 10^{-12} + 10^{-6}) < 5 \times 10^{-5},$$

where the first three terms correspond to the changing choices for z_k and the last term corresponds to the case that $r > e^{\sqrt{k}}$.

Adding the various contributions, we have that the reciprocal sum is smaller than

$$0.01238 + 0.09481 + 0.02595 + 0.49874 + 5 \times 10^{-5} < 0.6320.$$

$\square$

In sum, the large range bound for the reciprocal sum is

$$0.2516 + 0.1430 + 0.2543 + 0.8542 + 0.2790 + 0.6320 = 2.4141.$$

Combining the bounds from the small, middle, and large ranges,

$$\sum_{n\in\mathcal{S}} \frac{1}{n} < 1.4325 + 3.8006 + 2.4141 = 7.6472.$$

Acknowledgments

We are very grateful to Patrick Letendre for helpful ideas at the beginning of the project, including ideas used in Sections 3 and 4. We are also grateful to Giovanni Resta for his help extending the computer search for all solutions to 10^{13}.

References

[1] J. Bayless and P. Kinlaw, Consecutive coincidences of Euler's function. *International Journal of Number Theory.* **12** (2015), 1011–1026.
[2] M. A. Bennett, G. Martin, K. O'Bryant, and A. Rechnitzer, Explicit bounds for primes in arithmetic progressions. *Illinois J. Math.* **62** (2018), 427–532.
[3] P. Dusart, Explicit estimates of some functions over primes. *Ramanujan J.* **45** (2018), 227–251.
[4] P. Erdős, On a problem of Chowla and some related problems. *Math. Proc. Cambridge Phil. Soc.* **32** (1936), 530–540.
[5] P. Erdős, C. Pomerance, and A. Sárközy, On locally repeated values of certain arithmetic functions. II. *Acta Math. Hungarica* **49** (1987), 251–259.
[6] M. Kobayashi, A generalization of a series for the density of abundant numbers. *Int. J. Number Theory* **12** (2016), 671–677.
[7] H. M. Nguyen and C. Pomerance, The reciprocal sum of the amicable numbers. *Math. Comp.* **88** (2019), 1503–1526.
[8] G. Resta, Online encyclopedia of integer sequences, `http://oeis.org/A001274`, (2019).
[9] J. B. Rosser and L. Schoenfeld, Approximate formulas for some functions of prime numbers. *Illinois J. Math*, **6** (1962), 64–94.
[10] I. J. Schoenberg, Über die asymptotische Verteilung reeller Zahlen mod 1. *Math. Z.*, **28**, (1928), 171–199.
[11] T. Yamada, On equations $\sigma(n) = \sigma(n+k)$ and $\varphi(n) = \varphi(n+k)$. *J. Combinatorics and Number Theory*, **9** (2017), 15–21.