POPULAR SUBSETS FOR EULER'S φ -FUNCTION

PAUL POLLACK

ABSTRACT. Let $\varphi(n) = \#(\mathbb{Z}/n\mathbb{Z})^{\times}$ (Euler's totient function). Let $\epsilon > 0$, and let $\alpha \in (0,1)$. We prove that for all $x > x_0(\epsilon, \alpha)$ and every subset \mathscr{S} of [1, x] with $\#\mathscr{S} \leq x^{1-\alpha}$, the number of $n \leq x$ with $\varphi(n) \in \mathscr{S}$ is at most $x/L(x)^{\alpha-\epsilon}$, where

 $L(x) = \exp(\log x \cdot \log_3 x / \log_2 x).$

Under plausible conjectures on the distribution of smooth shifted primes, this upper bound is best possible, in the sense that the number α appearing in the exponent of L(x) cannot be replaced by anything larger.

1. INTRODUCTION

Let $\varphi(n) = \#(\mathbb{Z}/n\mathbb{Z})^{\times}$ (Euler's totient function), and let $N(m) = \#\varphi^{-1}(m)$. In 1935, Erdős [8] showed that $\varphi(\cdot)$ fails to be injective rather spectacularly: there is a constant c > 0 and an infinite, increasing sequence of positive integers m along which $N(m) > m^c$. Probably this result holds with any c < 1; already in the same paper, Erdős showed that would follow if, roughly speaking, shifted primes p - 1 are 'smooth' (free of large prime factors) with the same frequency as ordinary integers of the same size. Applying a theorem of Baker and Harman in this direction [1], one can show that c = 0.7039 is admissible above.

In the opposite direction, it was shown by Pomerance [24] (see also [25, 26]) that as $m \to \infty$,

(1)
$$N(m) \le m/L(m)^{1+o(1)},$$

where

$$L(x) = \exp\left(\frac{\log x \cdot \log \log \log x}{\log \log x}\right).$$

(Erdős [9] had earlier obtained that $N(m) \leq m/L(m)^{c+o(1)}$, for some unspecified positive constant c.) Under the same kind of assumption alluded to above concerning smooth shifted primes, Pomerance showed that there is a sequence of m tending to infinity along which equality holds in (1).

Since $n/\log \log n \ll \varphi(n) \leq n$ (see, for instance, [27, Theorem 5.6, p. 115] for the nontrivial half), one can restate (1) as a theorem about preimages of singleton subsets of [1, x]. Specifically, for any single-element subset $\mathscr{S} \subset [1, x]$, the number of $n \leq x$ with $\varphi(n) \in \mathscr{S}$ is at most $x/L(x)^{1+o(1)}$, as $x \to \infty$. Our main purpose here is to show that an upper bound of the same shape holds for considerably larger sets \mathscr{S} . As a special case of our result, Pomerance's upper bound of $x/L(x)^{1+o(1)}$ holds for all sets of size at most $x^{o(1)}$.

²⁰¹⁰ Mathematics Subject Classification. 11N64 (primary), 11N25, 11N37 (secondary).

Theorem 1. Let $\epsilon > 0$, and let $\alpha \in (0,1)$. For all $x > x_0(\epsilon, \alpha)$, and each set \mathscr{S} of integers in [1,x] with $\#\mathscr{S} \leq x^{1-\alpha}$, the number of $n \leq x$ with $\varphi(n) \in \mathscr{S}$ does not exceed

 $x/L(x)^{\alpha-\epsilon}$.

Theorem 1 is probably sharp for every $\alpha \in (0, 1)$, in the sense that the conclusion becomes false if the number α appearing in the exponent of L(x) is replaced by anything larger. One can see why by specializing \mathscr{S} to the set of $(\log x)^{1/\alpha}$ -smooth numbers contained in [1, x]. In this case, $\#\mathscr{S} = x^{1-\alpha+o(1)}$, as $x \to \infty$ (see, e.g., [27, Theorem 5.2, p. 513]). On the other hand, it is a consequence of work in [4] that under the aforementioned conjectures on shifted primes, the number of $n \leq x$ with $\varphi(n) \in \mathscr{S}$ is $x/L(x)^{\alpha+o(1)}$. The upper bound follows, unconditionally, from Theorem 3.1 of [4] (see [3] for corrections and a somewhat sharper theorem), while the (conditional) lower bound comes from taking $y = (\log x)^{1/\alpha}$ in eq. (8.4) of [4].

According to Erdős [10, 11] (see also [24]), Davenport and Heilbronn corresponded about the second moment of N(m), with Heilbronn showing that $\frac{1}{x} \sum_{m \leq x} N(m)^2 \to \infty$ as $x \to \infty$. Pomerance's conditional lower bound proof in [24] shows that for all large x, there is an $m \leq x$ with $N(m) \geq x/L(x)^{1+o(1)}$ (as $x \to \infty$). Thus (conditionally),

$$\sum_{m \le x} N(m)^2 \ge \frac{x^2}{L(x)^{2+\epsilon}}$$

for any $\epsilon > 0$ and all large x. Theorem 1 has the following (unconditional) consequence.

Corollary 2. For each $\epsilon > 0$ and all large x,

(2)
$$\sum_{m \le x} N(m)^2 \le \frac{x^2}{L(x)^{2-\epsilon}}.$$

The short proof of Corollary 2 is given in §4 below.

In the early 20th century, Carmichael [7] conjectured that for every n, the equation $\varphi(n') = \varphi(n)$ has a solution $n' \neq n$. In other words, if we let $C(n) = N(\varphi(n))$, we always have C(n) > 1. While Carmichael's conjecture remains unresolved, Ford has shown that C(n) assumes every integer value > 1 infinitely often [14].

One can view Carmichael's conjecture as a statement on the minimal order of C(n). Understanding the maximal order of C(n) is essentially equivalent to understanding the maximal order of N(m), while results on the average order of C(n) are essentially equivalent to results on $\sum_{m \leq x} N(m)^2$. This leaves the question of the typical behavior of C(n). In [18], it is proved that

(3)
$$L^*(n)^{\frac{1}{2}-\epsilon} < C(n) < L^*(n)^{\frac{3}{2}+\epsilon}$$

for almost all n (i.e., all n outside of a set of density zero), where

$$L^*(x) = \exp((\log \log x)^2 \log \log \log x).$$

In $\S5$, we show that the lower bound in (3) is essentially optimal.

Corollary 3. For almost all n,

(4)
$$C(n) < L^*(n)^{\frac{1}{2}+\epsilon}$$

Corollary 3 should be contrasted with Theorem 3 of [13]. There Ford shows that N(m), restricted to the set of m with $N(m) \ge 1$, is almost always nearly bounded: the relative upper density of m with with $N(m) \ge k$ tends to 0 as k tends to infinity.

We conclude the introduction by discussing an application to a problem considered in [4] and [2]. Fix a polynomial $f(T) \in \mathbb{Z}[T]$ of degree $d \geq 2$. It is shown in [2, Theorem 5.1] that the number of $n \leq x$ for which $\varphi(n)$ belongs to the range of f is at most $x/\exp((\kappa_d + o(1))\sqrt{\log x})$, where $\kappa_d = \sqrt{(2-2/d)\log 2}$. (When $f(T) = T^2$, a slightly stronger bound, but still of the shape $x/\exp((\log x)^{1/2+o(1)})$, was proved in [4].) Theorem 1 immediately improves this to $x/L(x)^{1-\frac{1}{d}+o(1)}$. The only property of the range of f relevant to this argument is the rate of growth of its counting function. Adapting the methods of this paper to take advantage of additional structure might be expected to lead to improve estimates in this problem. A first step in this direction is taken in [22], where it is shown that an upper bound of $x/L(x)^{1+o(1)}$ holds for each polynomial $f(T) = T^k$ (with k any fixed positive integer); that this upper bound is sharp, under the same plausible conjectures on smooth shifted primes alluded to earlier, is implicit in [4] (and fleshed out in [22]).

Notation. The letter p is reserved throughout for primes. We say that a positive integer n is y-smooth if every prime dividing n is at most y; the y-smooth component of n is its largest y-smooth divisor. If p^e is a prime power, we say that $p^e \parallel n$ if $p^e \mid n$ but $p^{e+1} \nmid n$. For Y > 0, we define arithmetic functions $\omega_{>Y}(\cdot)$ and $\Omega_{>Y}(\cdot)$ by

$$\omega_{>Y}(n) = \sum_{\substack{p|n\\p>Y}} 1, \quad \text{and} \quad \Omega_{>Y}(n) = \sum_{\substack{p^e \mid |n\\p>Y}} e;$$

when Y = 0, we omit the subscripts and write simply $\omega(\cdot)$ and $\Omega(\cdot)$. We write $p^{-}(n)$ for the smallest prime dividing n, with the convention that $p^{-}(1) = \infty$. The kth iterate of the natural logarithm is abbreviated as $\log_{k}(\cdot)$.

2. Lemmata

In this section we collect preliminary results needed for the proof of Theorem 1.

Given a real number $z \ge 2$, we let $W_z(\cdot)$ denote the additive arithmetic function whose value at prime powers p^e is given by $W_z(p^e) = \omega_{>z}(\varphi(p^e))$. The following estimate for a high exponential moment of $W_z(\cdot)$ implies that $W_z(n)$ is 'fairly small' for 'most' $n \le x$.

Lemma 4. Let

$$z = \exp((\log_2 x)^{1/2}).$$

Fix any $\eta \in (0, 1)$, and let

$$A = (\log_2 x)^{1-\eta}.$$

As $x \to \infty$,

Proof. We borrow ideas from the argument used in [19] to bound $\sum_{n \leq x} \tau(\varphi(n))$ from above. Let $c \in (1, 2)$ be a parameter to be specified later. Then (Rankin's trick)

(5)
$$\sum_{n \le x} A^{W_z(n)} \le x^c \sum_{n \le x} \frac{A^{W_z(n)}}{n^c} \le x^c \prod_{p \le x} \left(1 + \sum_{k=1}^{\infty} \frac{A^{\omega_{>z}(p^{k-1}(p-1))}}{p^{kc}} \right)$$
$$\le x^c \exp\left(\sum_{p \le x} \sum_{k=1}^{\infty} \frac{A^{\omega_{>z}(p^{k-1}(p-1))}}{p^{kc}}\right)$$

Since $\omega_{>z}(p(p-1)) = \omega_{>z}(p^2(p-1)) = \omega_{>z}(p^3(p-1)) = \dots,$ $\sum_{k=1}^{\infty} \frac{A^{\omega_{>z}(p^{k-1}(p-1))}}{k_2} = \frac{A^{\omega_{>z}(p-1)}}{k_2} + A^{\omega_{>z}(p(p-1))} \left(\frac{1}{2k} + \frac{1}{2k}\right)$

$$\sum_{k=1}^{\infty} \frac{p^{kc}}{p^{kc}} = \frac{p^{c}}{p^{c}} + A^{\omega_{>z}(p(p-1))} \left(\frac{1}{p^{2c}} + \frac{1}{p^{3c}} + \dots\right)$$
$$\leq \frac{A^{\omega_{>z}(p-1)}}{p^{c}} + 2\frac{A^{\omega_{>z}(p(p-1))}}{p^{2c}}.$$

Now if $p \leq z$, then

$$\frac{A^{\omega_{>z}(p(p-1))}}{p^{2c}} = \frac{A^{\omega_{>z}(p-1)}}{p^{2c}} < \frac{A^{\omega_{>z}(p-1)}}{p^{c}},$$

while if p > z, then

$$\frac{A^{\omega_{>z}(p(p-1))}}{p^{2c}} = \frac{A}{p^c} \frac{A^{\omega_{>z}(p-1)}}{p^c} \le \frac{A}{z} \frac{A^{\omega_{>z}(p-1)}}{p^c} < \frac{A^{\omega_{>z}(p-1)}}{p^c}$$

Thus, in either case,

(6)
$$\sum_{k=1}^{\infty} \frac{A^{\omega_{>z}(p^{k-1}(p-1))}}{p^{kc}} \le 3 \frac{A^{\omega_{>z}(p-1)}}{p^c}.$$

Hence, to bound the right-hand side of (5), it is enough to bound $\sum_{p \leq x} A^{\omega_{>z}(p-1)}/p^c$. We can derive such an estimate by partial summation, given upper bounds for the sums

$$S(T) := \sum_{p \le T} A^{\omega_{>z}(p-1)}, \quad \text{valid for all } T > 1.$$

Let $K = \lceil A \rceil$. Observe that for every positive integer n,

$$A^{\omega(n)} \le \tau_K(n),$$

where as usual

$$\tau_K(n) = \sum_{\substack{d_1, \dots, d_K \\ d_1 \cdots d_K = n}} 1;$$

indeed, both sides of the claimed inequality are multiplicative and the inequality is obvious on prime powers. Furthermore, if $d_1 \cdots d_K = n$, then some $d_i \ge n^{1/K}$. Since the d_i play symmetric roles,

$$\tau_{K}(n) \leq K \sum_{\substack{d_{1},\dots,d_{K} \\ d_{1}\cdots d_{K}=n \\ d_{K} \geq n^{1/K}}} 1 \leq K \sum_{\substack{d_{1},\dots,d_{K-1} \\ d_{1}\cdots d_{K-1} \mid n \\ d_{1}\cdots d_{K-1} \leq n^{1-\frac{1}{K}}}} 1.$$

Writing $(p-1)_{>z}$ for the largest divisor of p-1 composed of primes exceeding z,

$$A^{\omega_{>z}(p-1)} = A^{\omega((p-1)_{>z})} \le \tau_K((p-1)_{>z}).$$

Consequently,

$$S(T) = \sum_{p \le T} A^{\omega_{>z}(p-1)} \le K \sum_{p \le T} \sum_{\substack{d_1, \dots, d_{K-1} \\ d_1 \cdots d_{K-1} | p-1 \\ p^-(d_i) > z \ \forall i \\ d_1 \cdots d_{K-1} \le T^{1-\frac{1}{K}}} 1.$$

Reversing the order of summation and applying the Brun–Titchmarsh inequality, this is seen to be at most

$$K \sum_{\substack{d_1, \dots, d_{K-1} \\ p^-(d_i) > z \ \forall i \\ d_1 \cdots d_{K-1} \le T^{1-\frac{1}{K}}}} \pi(T; d_1 \cdots d_{K-1}, 1) \ll K^2 \frac{T}{\log T} \sum_{\substack{d_1, \dots, d_{K-1} \\ p^-(d_i) > z \ \forall i \\ d_1 \cdots d_{K-1} \le T^{1-\frac{1}{K}}}} \frac{1}{\varphi(d_1 \cdots d_{K-1})} \ll A^2 \frac{T}{\log T} \sum_{\substack{d \le T^{1-\frac{1}{K}} \\ p^-(d) > z}} \frac{\tau_{K-1}(d)}{\varphi(d)}.$$

Continuing,

(7)
$$\sum_{\substack{d \leq T^{1-\frac{1}{K}} \\ p^{-}(d) > z}} \frac{\tau_{K-1}(d)}{\varphi(d)} \leq \prod_{z
$$\leq \exp\left(\sum_{z$$$$

Observe that

$$\sum_{p>z} \sum_{k=2}^{\infty} \frac{\tau_{K-1}(p^k)}{\varphi(p^k)} = \sum_{k=2}^{\infty} \binom{k+K-2}{K-2} \sum_{p>z} \frac{1}{\varphi(p^k)} \ll \sum_{k=2}^{\infty} \binom{k+K-2}{K-2} z^{1-k}.$$

Moreover, in the final sum, each ratio of consecutive terms satisfies

$$\frac{\binom{k+K-1}{K-2}z^{-k}}{\binom{k+K-2}{K-2}z^{1-k}} = \frac{1}{z}\left(1+\frac{K-2}{k+1}\right) < \frac{K}{z} < \frac{2A}{z} < \frac{1}{2}$$

(if x is large). Hence, that sum is dominated by its first term:

$$\sum_{k=2}^{\infty} \binom{k+K-2}{K-2} z^{1-k} \ll \binom{K}{K-2} z^{-1} \ll \frac{A^2}{z} \ll 1.$$

So the second factor in (7) is O(1). The first factor in (7) is trivially bounded when $T \leq z$. On the other hand, if T > z, then

$$\sum_{z$$

(Here we estimated the sum on p using the prime number theorem with error term.) Since $A < \log_2 x = (\log z)^2$, we conclude that

$$\exp\left(\sum_{z z. \end{cases}$$

Collecting the above estimates, we have shown that

$$S(T) \ll A^2 \frac{T}{\log T} \cdot \begin{cases} 1 & \text{if } T \leq z, \\ (\log T / \log z)^{K-1} & \text{if } T > z. \end{cases}$$

We now return to estimating $\sum_{p \leq x} A^{\omega_{>z}(p-1)}/p^c$. Keeping in mind that 1 < c < 2,

$$\sum_{p \le x} \frac{A^{\omega_{>z}(p-1)}}{p^c} \le \int_{2^-}^{\infty} t^{-c} \, dS(t) = c \int_{2}^{\infty} \frac{S(t)}{t^{c+1}} \, dt \ll \int_{2}^{z} \frac{S(t)}{t^{c+1}} \, dt + \int_{z}^{\infty} \frac{S(t)}{t^{c+1}} \, dt$$
$$\ll A^2 \int_{2}^{z} \frac{dt}{t \log t} + \frac{A^2}{(\log z)^{K-1}} \int_{z}^{\infty} \frac{(\log t)^{K-2}}{t^c} \, dt.$$

Now $\int_2^z \frac{dt}{t \log t} \ll \log \log z \ll \log_3 x$, while

$$\int_{z}^{\infty} \frac{(\log t)^{K-2}}{t^{c}} dt \le \int_{1}^{\infty} \frac{(\log t)^{K-2}}{t^{c}} dt = \frac{(K-2)!}{(c-1)^{K-1}}$$

(The final equality can be seen by making the change of variables $t = e^{u/(c-1)}$ and invoking Euler's formula for the Gamma function.) It follows that

$$\sum_{p \le x} \frac{A^{\omega_{>z}(p-1)}}{p^c} \ll A^2 \log_3 x + \frac{A^2}{(\log z)^{K-1}} \frac{(K-2)!}{(c-1)^{K-1}}$$

Using this estimate together with (5), (6) shows that (8)

$$\sum_{n \le x} A^{W_z(n)} \le x \exp(O(A^2 \log_3 x)) \cdot \exp\left((c-1)\log x + O\left(\frac{A^2}{(\log z)^{K-1}}\frac{(K-2)!}{(c-1)^{K-1}}\right)\right).$$

Now $\exp(O(A^2 \log_3 x)) \le \exp((\log_2 x)^2) = L(x)^{o(1)}$. We complete the proof by showing that when

$$c = 1 + \frac{K}{(\log x)^{\frac{1}{K}} (\log z)^{1 - \frac{1}{K}}}$$

the final right-hand factor in (8) is also $L(x)^{o(1)}$. With this choice of c,

$$(c-1)\log x + O\left(\frac{A^2}{(\log z)^{K-1}}\frac{(K-2)!}{(c-1)^{K-1}}\right) \ll A^2 K\left(\frac{\log x}{\log z}\right)^{1-\frac{1}{K}}$$

Since $A, K \ll (\log_2 x)^{1-\eta}$, this last expression is

$$\ll (\log_2 x)^{O(1)} (\log x)^{1-\frac{1}{K}} \ll \frac{\log x}{\exp((\log_2 x)^{\eta/2})}$$

In particular, it is $o(\log L(x))$, as desired.

The next two lemmas — which are somewhat crude, but effective for our purpose — allow us to control, for most n, the squarefull parts of p-1 for primes p dividing n.

Lemma 5. Let $Y, Z \ge 1$. The number of positive integers $n \le x$ with

(9)
$$\sum_{p|n} (\Omega_{>Y}(p-1) - \omega_{>Y}(p-1)) \ge Z$$

is at most $xL(x)^{2+o(1)}Y^{-Z/2}$, as $x \to \infty$ (uniformly in Y, Z).

		L	
		L	
		L	
		L	

Proof. Assume that (9) holds. For each $p \mid n$, let $E_p = \Omega_{>Y}(p-1) - \omega_{>Y}(p-1)$. We may view the nonzero elements of the multiset $\{E_p\}_{p\mid n}$ as partitioning a certain positive integer S (say). Then $S \geq Z$ by assumption, while trivially

$$S \le \sum_{p|n} \Omega_{>Y}(p-1) \le \Omega(\varphi(n)) < 2\log x.$$

Our strategy will be to bound the number of n corresponding to a particular partition of S and then to sum over all possible choices of this partition.

Changing notation slightly, label the elements of the partition as E_1, \ldots, E_k . Then there are distinct primes p_1, \ldots, p_k dividing n, where each $p_i - 1$ has squarefull part d_i (say) with

$$E_i = \Omega_{>Y}(p_i - 1) - \omega_{>Y}(p_i - 1)$$
$$= \Omega_{>Y}(d_i) - \omega_{>Y}(d_i),$$

and

$$d_i \ge Y^{\Omega_{>Y}(d_i)} \ge Y^{E_i}.$$

The number of such $n \leq x$ is at most

(10)
$$x \prod_{i=1}^{k} \left(\sum_{p_i}' \frac{1}{p_i} \right),$$

where the ' indicates a restriction to $p_i \leq x$ for which $p_i - 1$ has squarefull part at least Y^{E_i} . For each positive integer $d \leq x$,

$$\sum_{\substack{p \equiv 1 \pmod{d}}} \frac{1}{p} \ll \frac{\log_2 x}{\varphi(d)}$$
$$\ll \frac{(\log_2 x)^2}{d}$$

(The first estimate here comes from Brun–Titchmarsh and partial summation.) Summing on squarefull $d > Y^{E_i}$ yields

$$\sum_{p_i}' \frac{1}{p_i} \ll (\log_2 x)^2 Y^{-E_i/2}.$$

Referring back to (10), we find that the number of n as above is at most

$$x \cdot C^k (\log_2 x)^{2k} Y^{-\frac{1}{2}\sum_{i=1}^k E_i}$$

for some absolute constant C > 0. As each $n \in [1, x]$ has at most $(1+o(1)) \log x / \log_2 x$ distinct prime factors, we must have $k \leq (1+o(1)) \log x / \log_2 x$, so that

$$C^k (\log_2 x)^{2k} \le L(x)^{2+o(1)}.$$

Since $\sum_{i=1}^{k} E_i = S \ge Z$, we conclude that the number of *n* corresponding to the partition with parts E_1, \ldots, E_k is at most $xL(x)^{2+o(1)}Y^{-Z/2}$.

It remains to sum over all possible partitions. But the number S being partitioned satisfies $S < 2 \log x = L(x)^{o(1)}$, while the number of partitions of each integer smaller than $2 \log x$ is at most $\exp(O(\sqrt{\log x})) = L(x)^{o(1)}$. (For a short proof of the crude upper bound on p(m) employed here, see for instance [17, pp. 89–90].) The lemma follows.

Lemma 6. Let $Z \ge 1$. The number of $n \le x$ with

$$\omega(n) \le \frac{\log x \cdot \log_3 x}{(\log_2 x)^2}$$

satisfying

$$\sum_{p|n} \Omega(p-1) \geq Z$$

is at most

$$xL(x)^{2+o(1)}2^{-Z/2}$$

as $x \to \infty$ (uniformly in Z).

Proof. The proof is similar to that of the last lemma. For each $p \mid n$, put $E_p = \Omega(p-1)$. We consider the nonzero E_p as the components of a partition of S, where $Z \leq S < 2 \log x$.

Suppose now that n corresponds to a given partition E_1, \ldots, E_k . Then there are distinct primes p_1, \ldots, p_k dividing n with each $E_i = \Omega(p_i - 1)$. Thus, the number of corresponding n is at most

(11)
$$x \prod_{i=1}^{k} \left(\sum_{\substack{p_i \le x \\ \Omega(p_i-1) = E_i}} \frac{1}{p_i} \right)$$

To estimate the sums on the p_i , recall that the number of integers $n \leq T$ with $\Omega(n) = E$ is $\ll \frac{E}{2^E} \cdot T \log T$ (for all $E \geq 1$ and $T \geq 2$); see Exercise 05 in [15, p. 12]. Applying partial summation,

$$\sum_{\substack{p \le x \\ \Omega(p-1)=E}} \frac{1}{p} \le \sum_{\substack{n \le x \\ \Omega(n)=E}} \frac{1}{n} \ll \frac{E}{2^E} (\log x)^2 \ll \frac{1}{2^{E/2}} (\log x)^2.$$

Inserting this in (11) bounds the number of n by

$$x(C(\log x)^2)^k \cdot 2^{-\frac{1}{2}\sum_{i=1}^k E_i} \le x(C(\log x)^2)^k 2^{-Z/2}.$$

The assumed upper bound on $\omega(n)$ implies that $k \leq \frac{\log x \cdot \log_3 x}{(\log_2 x)^2}$, and so

$$(C(\log x)^2)^k \le L(x)^{2+o(1)}$$

We complete the proof by summing on the $L(x)^{o(1)}$ possible partitions.

The final result we need is a variant of [18, Lemma 2.1] (see also [23, Lemma 3.3]).

Recall that a *multiplicative partition* of a positive integer n is a way of writing n as a product of integers larger than 1, where two multiplicative partitions are considered the same if they differ only in the order of the factors.

Lemma 7. The following statement holds for a certain constant C > 0: Let $x \ge 3$, and let d be a positive integer. Then the number of positive integers $n \le x$ for which $d \mid \varphi(n)$ is at most

$$\frac{x}{d} (C(\log_2 x)^2)^{\Omega(d)} \cdot M(d),$$

where M(d) denotes the number of multiplicative partitions of d.

Remark. Much is known about the behavior of M(n) as a function of n; see, for instance, [20, 21] and [6]. Our application requires only the crude bound

(12)
$$M(n) \le \Omega(n)^{\Omega(n)}.$$

One way of seeing (12) is to observe that, with $w := \Omega(n)$, there is always a surjection from the collection of set partitions of $\{1, 2, 3, \ldots, w\}$ onto the collection of multiplicative partitions of n.¹ The number of such set partitions is easily seen to be at most w^w (see [5] for sharper bounds on the number of set partitions).

Proof. Write $n = \prod_{p} p^{e_p}$, where the product is over the primes p dividing n. Assume that

$$d \mid \varphi(n) = \prod_{p} (p-1)p^{e_p-1}.$$

Let $w := \Omega(n)$, and view the above right-hand side as a product of $\sum_{p|n}(1+(e_p-1)) = w$ terms, say $P_1 \cdots P_w$, where each $P_i = p - 1$ or p. Let p_i denote the prime p associated to P_i in this way. Then $n = \prod_{i=1}^w p_i$. Since $d \mid \prod_{i=1}^w P_i$, there is a decomposition $d = d_1 \cdots d_w$ where each d_i divides P_i . Renumbering, we can assume that $d_i > 1$ precisely for $i = 1, \ldots, k$. For later use, we observe that

(13)
$$k \le \sum_{i=1}^{k} \Omega(d_i) = \Omega(d).$$

Now suppose we start from a given list d_1, \ldots, d_k . Since each $p_i \equiv 0$ or 1 (mod d_i) and n is divisible by $\prod_{i=1}^k p_i$, we can bound the number of $n \leq x$ that give rise to this list of d_i by

$$x \prod_{i=1}^{k} \bigg(\sum_{\substack{p_i \le x \\ p_i \equiv 0, 1 \pmod{d_i}}} \frac{1}{p_i} \bigg).$$

For each *i*, the inner sum is $\ll \frac{\log_2 x}{\varphi(d_i)} + \frac{1}{d_i} \ll \frac{(\log_2 x)^2}{d_i}$. Thus, for a certain absolute constant C > 0, the number of *n* as above is at most

$$x\prod_{i=1}^{k} (C(\log_2 x)^2/d_i) = \frac{x}{d} (C(\log_2 x)^2)^k \le \frac{x}{d} (C(\log_2 x)^2)^{\Omega(d)}.$$

(We use here that $k \leq \Omega(d)$.)

It remains to sum over all possible lists d_1, \ldots, d_k , keeping in mind that the order of this list is irrelevant above. Since the d_i are larger than 1 and their product is d, the number of possibilities for d_1, \ldots, d_k (up to ordering) is no more than M(d).

3. Proof of Theorem 1

We use z with the same meaning as in Lemma 4, namely $z = \exp((\log_2 x)^{1/2})$.

¹Explicitly: Label the prime factors of n as p_1, \ldots, p_w . If $\mathscr{S}_1, \ldots, \mathscr{S}_k$ form a set partition of $\{1, 2, 3, \ldots, w\}$, then $s_1 \cdots s_k$ is a multiplicative partition of n, where each $s_I := \prod_{i \in \mathscr{S}_I} p_i$.

Suppose that $n \leq x$ and that $m := \varphi(n) \in \mathscr{S}$. If m < x/L(x), then $n \ll x \log_2 x/L(x)$. The number of such n is negligible compared to our target upper bound, and so we can assume that

(14)
$$m \ge x/L(x).$$

Similarly, we can assume that the largest squarefull divisor of n is at most

$$\exp(\log x/(\log_2 x)^{3/4}),$$

and that

$$\omega(n) \le \frac{\log x \cdot \log_3 x}{(\log_2 x)^2}.$$

Indeed, the number of n with a squarefull part larger than the above is

$$\ll x / \exp\left(\frac{1}{2}\log x / (\log_2 x)^{3/4}\right) < x/L(x),$$

while by a 1917 theorem of Hardy–Ramanujan [16] the number of n with $\omega(n)$ larger than specified is

$$\ll \frac{x}{\log x} \sum_{k > \log x \log_3 x/(\log_2 x)^2} \frac{(\log_2 x + O(1))^{k-1}}{(k-1)!} \le x/L(x)^{1+o(1)}.$$

With c a constant in (0, 1) whose precise value will be chosen later in terms of α , we assume that

$$W_z(n) < c \frac{\log x}{\log_2 x}.$$

Note that by Lemma 4, for any fixed $\eta \in (0, 1)$, the number of $n \leq x$ for which this inequality fails is at most

(15)
$$x/L(x)^{(1-\eta)c+o(1)},$$

as $x \to \infty$.

At the cost of excluding another set of n of size at most x/L(x), we may assume that

(16)
$$\Omega_{>z}(m) < (1+\eta)c\frac{\log x}{\log_2 x}$$

Indeed, suppose this inequality fails. Then recalling the definition of $W_z(\cdot)$, we see that

$$\eta c \frac{\log x}{\log_2 x} < \Omega_{>z}(m) - W_z(n) = \sum_{p^e \parallel n} (\Omega_{>z}(\varphi(p^e)) - \omega_{>z}(\varphi(p^e))).$$

Since $\varphi(p^e) = p^{e-1}(p-1)$, it follows that

$$\sum_{p|n} (\Omega_{>z}(p-1) - \omega_{>z}(p-1)) > \eta c \frac{\log x}{\log_2 x} - \sum_{\substack{p^e \parallel n \\ p > z, \ e \ge 2}} (e-2).$$

If

$$\sum_{\substack{p^e \parallel n \\ p > z, \ e \ge 2}} (e-2) \ge \frac{1}{2} \eta c \frac{\log x}{\log_2 x},$$

then n is divisible by a squarefull number exceeding

$$z^{\frac{1}{2}\eta c \frac{\log x}{\log_2 x}} = \exp\left(\frac{1}{2}\eta c \frac{\log x}{(\log_2 x)^{1/2}}\right);$$

this contradicts our assumption on the squarefull part of n. Hence,

$$\sum_{p|n} (\Omega_{>z}(p-1) - \omega_{>z}(p-1)) > \frac{1}{2} \eta c \frac{\log x}{\log_2 x}.$$

Lemma 5 (with Y = z and $Z = \frac{1}{2}\eta c \frac{\log x}{\log_2 x}$) then implies that n falls into a set of size at most $xL(x)^{2+o(1)} \exp(-\frac{1}{4}\eta c \frac{\log x}{(\log_2 x)^{1/2}})$; but this last quantity is < x/L(x) for large x.

Finally, excluding another set of size at most x/L(x), we may assume that

(17)
$$\Omega(m) < \frac{\log x}{(\log_2 x)^{2/3}}.$$

The argument is similar to the one just seen. Supposing the inequality fails,

$$\sum_{p|n} \Omega(p-1) \ge \frac{\log x}{(\log_2 x)^{2/3}} - \sum_{\substack{p^e \parallel n \\ e \ge 2}} (e-1).$$

If

$$\sum_{\substack{p^e \parallel n \\ e \ge 2}} (e-1) \ge \frac{1}{2} \frac{\log x}{(\log_2 x)^{2/3}},$$

then n is divisible by a squarefull number exceeding

$$2^{\frac{1}{2}\log x/(\log_2 x)^{2/3}} = \exp\left(\frac{\log 2}{2}\frac{\log x}{(\log_2 x)^{2/3}}\right),$$

contrary to our assumption on n. Thus,

$$\sum_{p|n} \Omega(p-1) \ge \frac{1}{2} \frac{\log x}{(\log_2 x)^{2/3}}$$

Applying Lemma 6 shows that the number of possibilities for n is at most $xL(x)^{2+o(1)} \cdot \exp(-\frac{\log 2}{4}\frac{\log x}{(\log_2 x)^{2/3}})$, which is eventually < x/L(x).

Suppose now that $m = \varphi(n) \in S$, where m, n satisfy (14), (16), and (17). Write m = m'd, where m' is the z-smooth component of m. Notice that

$$m' \le z^{\Omega(m)} \le \exp(\log x / (\log_2 x)^{1/6}) = x^{o(1)}$$

Thus,

$$d = m/m' \ge (x/L(x))/m' \ge x^{1-o(1)}$$

Moreover,

$$\Omega(d) = \Omega_{>z}(m) < (1+\eta)c\frac{\log x}{\log_2 x}.$$

So by Lemma 7 and the remark following it,

$$\begin{split} \#\{n \leq x : \varphi(n) = m\} &\leq \#\{n \leq x : d \mid \varphi(n)\} \\ &\leq \frac{x}{d} (C(\log_2 x)^2)^{\Omega(d)} \Omega(d)^{\Omega(d)} \\ &\leq x^{(1+\eta)c+o(1)}, \end{split}$$

as $x \to \infty$. Since $\#\mathscr{S} \leq x^{1-\alpha}$, summing on *m* yields at most $x^{1-\alpha+(1+\eta)c+o(1)}$ values of *n*, which is negligible if we choose $c = (1 - \eta)\alpha$.

Looking back through the argument, we find that the total number of $n \leq x$ with $\varphi(n) \in \mathscr{S}$ is at most

$$x/L(x)^{(1-\eta)c+o(1)} = x/L(x)^{(1-\eta)^2\alpha+o(1)}$$

(The dominant contribution here comes from (15).) Given $\epsilon > 0$, we fix $\eta > 0$ small enough to ensure that $(1 - \eta)^2 \alpha > \alpha - \epsilon$. The theorem follows.

4. Application to a problem of Davenport–Heilbronn: Proof of Corollary 2

Put $X = 2x \log_2 x$, and observe that if $\varphi(n) \in [1, x]$, then $n \leq X$ (once x is large). To bound $\sum_{m \leq x} N(m)^2$, we split the positive integers $m \leq x$ into two sets: \mathscr{S}_1 , consisting of those m with $N(m) \leq x/L(x)^2$, and \mathscr{S}_2 consisting of all other m. For large x,

(18)
$$\sum_{m \in \mathscr{S}_1} N(m)^2 \le \frac{x}{L(x)^2} \sum_{m \le x} N(m) \le \frac{x \cdot X}{L(x)^2} \le \frac{x^2}{L(x)^{2-\epsilon}}.$$

To handle the contribution from \mathscr{S}_2 , notice that

$$\#\mathscr{S}_2 \cdot \frac{x}{L(x)^2} \le \sum_{m \in \mathscr{S}_2} N(m) \le X.$$

Thus, $\#\mathscr{S}_2 \leq L(x)^3$ (say), for large x. Since $L(x)^3 = X^{o(1)}$, Theorem 1 (with x replaced by X) implies that

$$\sum_{m \in \mathscr{S}_2} N(m) = \#\{n \le X : \varphi(n) \in \mathscr{S}_2\}$$
$$\le \frac{X}{L(X)^{1-\epsilon}}$$

for large x. Hence,

(19)
$$\sum_{m \in \mathscr{S}_2} N(m)^2 \leq (\max_{m \leq x} N(m)) \cdot \frac{X}{L(X)^{1-\epsilon}}$$
$$\leq \frac{x}{L(x)^{1-\epsilon}} \cdot \frac{X}{L(X)^{1-\epsilon}} \leq \frac{x^2}{L(x)^{2-3\epsilon}}.$$

The claimed upper bound (2) follows from combining (18) and (19), after replacing ϵ by $\epsilon/4$ (say).

5. The typical number of solutions to
$$\varphi(n) = \varphi(m)$$
:
Proof of Corollary 3

We need a version of Lemma 7 incorporating a restriction on the number of prime factors of n. Write $\binom{N}{K}$ for the number of ways of partitioning an N-element set into K nonempty subsets (a Stirling number of the second kind).

Lemma 8. The following statement holds for a certain constant C > 0: Let $x \ge 3$. Let d be a squarefree positive integer, and let $K \ge 0$. The number of positive integers $n \le x$ for which $d \mid \varphi(n)$ and $\Omega(n) \le K$ is at most

$$\frac{x}{d} (C(\log_2 x)^2)^K \sum_{0 \le k \le K} \left\{ \frac{\omega(d)}{k} \right\}.$$

Proof. We refer back to the proof of Lemma 7. Rather than bound k as in (13), we use that $k \leq w = \Omega(n) \leq K$. This allows us to bound the number of n corresponding to a particular multiplicative partition $d = d_1 \cdots d_k$ by

$$\frac{x}{d}(C(\log_2 x)^2)^K.$$

Moreover, instead of multiplying by M(d) at the end of the proof, it suffices to multiply by the number of multiplicative partitions of d into at most K parts. Since d is squarefree, this number is exactly $\sum_{0 \le k \le K} {\omega(d) \choose k}$.

We can now prove that (4) holds almost always. It is enough to show that as $x \to \infty$, the inequality

$$C(n) \le L^*(x)^{\frac{1}{2} + \frac{1}{2}\epsilon}$$

holds for all but o(x) values of $n \in (x/2, x]$.

Excluding o(x) values of n in (x/2, x], we may assume that

$$\Omega(n) \le 2\log_2 x.$$

Excluding another o(x) values of n, we may also assume that

$$\Omega(\varphi(n)) \le \left(\frac{1}{2} + \epsilon\right) (\log_2 x)^2;$$

see [12], where Erdős and Pomerance show that $\Omega(\varphi(n))$ has normal order $\frac{1}{2}(\log_2 n)^2$ (in fact, their result is stronger: an analogue of the Erdős-Kac theorem for $\Omega(\varphi(n))$). On p. 350 of the same paper [12], it is shown that excluding o(x) values of $n \leq x$, we may assume that every prime p for which $p^2 \mid \varphi(n)$ satisfies $p \leq (\log_2 x)^2$, and that

$$\sum_{\substack{p^e \|\varphi(n)\\ p \le (\log_2 x)^2}} e \le 2\log_2 x \log_4 x.$$

Note that these assumptions imply that apart from at most o(x) exceptional $n \in (x/2, x]$, the squarefull part of $\varphi(n)$ is bounded above by

$$((\log_2 x)^2)^{2\log_2 x \log_4 x} < \exp((\log_2 x)^2)$$

As a consequence, the largest squarefree divisor of $\varphi(n)$ is bounded below by

(20)
$$\varphi(n) / \exp((\log_2 x)^2) > x / \exp(2(\log_2 x)^2)$$

Let U be the least integer with $2^U \ge (\log_2 x)^2$. We can clearly assume, at the cost of excluding o(x) values of $n \le x$, that $2^U \nmid n$. We can also assume that n is not divisible by any prime $p \equiv 1 \pmod{2^U}$, since the number of exceptional $n \le x$ is

$$\ll x \sum_{\substack{p \le x \\ p \equiv 1 \pmod{2^U}}} \frac{1}{p} \ll x \frac{\log_2 x}{\varphi(2^U)} \ll \frac{x}{\log_2 x}.$$

We proceed to bound C(n) for all $n \in (x/2, x]$ not belonging to any of the exceptional sets just described.

Note that with $v_2(\cdot)$ denoting the 2-adic valuation,

$$v_2(\varphi(n)) = \sum_{p^e ||\varphi(n)|} v_2(\varphi(p^e)) \le U + \sum_{\substack{p|n\\p>2}} v_2(p-1) \le U + U\omega(n) < 10\log_2 x \log_3 x$$

for large x, since $U < 3\log_3 x$ and $\omega(n) \le \Omega(n) \le 2\log_2 x$.

Suppose now that $\varphi(n') = \varphi(n)$. Then

$$v_2(\varphi(n)) = v_2(\varphi(n')) \ge \omega(n') - 1$$

(since $2 \mid p-1$ for all p > 2). Hence,

$$\omega(n') < 11 \log_2 x \log_3 x.$$

To apply Lemma 8 we require an upper bound on the larger quantity $\Omega(n')$. If $p^3 \mid n'$, then $p^2 \mid \varphi(n') = \varphi(n)$, and so our assumptions on n force $p \leq (\log_2 x)^2$. Thus, if N'denotes the $(\log_2 x)^2$ -smooth component of n', then

$$\Omega(n') \le \Omega(N') + 2\omega(n').$$

To bound $\Omega(N')$, notice that $\varphi(N')$ is a $(\log_2 x)^2$ -smooth divisor of $\varphi(n') = \varphi(n)$. Hence,

$$\Omega(\varphi(N')) \le \sum_{\substack{p^e \| \varphi(n) \\ p \le (\log_2 x)^2}} e \le 2\log_2 x \log_4 x.$$

On the other hand, $\Omega(\varphi(N')) \ge \Omega(N') - 1$. (This follows easily from the observation that $\Omega(\varphi(p^e)) = e - 1 + \Omega(p - 1)$ for each prime power p^e .) Therefore,

$$\Omega(N') \le 1 + \Omega(\varphi(N')) < 3\log_2 x \log_4 x.$$

Piecing together the results of this paragraph,

$$\Omega(n') < 25 \log_2 x \log_3 x.$$

Let d denote the largest squarefree divisor of $\varphi(n)$, which we know to be bounded below by (20). Letting $X = 2x \log_2 x$ and $K = 25 \log_2 x \log_3 x$, we see that

$$C(n) = \#\{n' : \varphi(n') = \varphi(n)\} = \#\{n' \le X : \varphi(n') = \varphi(n), \Omega(n') \le K\}$$

$$\le \#\{n' \le X : d \mid \varphi(n'), \Omega(n') \le K\}.$$

By Lemma 8, the number of these n' does not exceed

$$\frac{X}{d} (C(\log_2 X)^2)^K \sum_{0 \le k \le K} \left\{ \frac{\omega(d)}{k} \right\} \le \exp(O((\log_2 x)^2)) \sum_{0 \le k \le K} \left\{ \frac{\omega(d)}{k} \right\}$$

Bounding the Stirling numbers trivially, and keeping in mind that $\omega(d) \leq \Omega(\varphi(n)) \leq (\frac{1}{2} + \epsilon) (\log_2 x)^2$, we see that

$$\sum_{0 \le k \le K} \left\{ \begin{matrix} \omega(d) \\ k \end{matrix} \right\} \le \sum_{0 \le k \le K} k^{\omega(d)} \le (K+1) \cdot K^{\omega(d)} \le L^*(x)^{\frac{1}{2} + 2\epsilon}$$

Since $\exp(O((\log_2 x)^2)) = L^*(x)^{o(1)}$, we conclude that $C(n) \leq L^*(x)^{\frac{1}{2}+3\epsilon}$ for all but o(x) values of $n \in (x/2, x]$, as $x \to \infty$. Replacing ϵ by $\epsilon/6$ finishes the proof.

Acknowledgments. The author is supported by NSF award DMS-1402268. He thanks the referee for a careful reading of the manuscript and helpful comments.

References

- R. C. Baker and G. Harman, Shifted primes without large prime factors, Acta Arith. 83 (1998), 331–361.
- [2] W. D. Banks, K. Ford, F. Luca, F. Pappalardi, and I. E. Shparlinski, Values of the Euler function in various sequences, Monatsh. Math. 146 (2005), 1–19.
- [3] W. D. Banks, J. B. Friedlander, C. Pomerance, and I. E. Shparlinski, *Counting integers with a smooth totient*, submitted; arXiv:1809.01214 [math.NT].
- [4] _____, Multiplicative structure of values of the Euler function, High primes and misdemeanours: lectures in honour of the 60th birthday of Hugh Cowie Williams, Fields Inst. Commun., vol. 41, Amer. Math. Soc., Providence, RI, 2004, pp. 29–47.
- [5] D. Berend and T. Tassa, Improved bounds on Bell numbers and on moments of sums of random variables, Probab. Math. Statist. 30 (2010), 185–205.
- [6] E. R. Canfield, P. Erdős, and C. Pomerance, On a problem of Oppenheim concerning "factorisatio numerorum", J. Number Theory 17 (1983), 1–28.
- [7] R. D. Carmichael, Note on Euler's φ-function, Bull. Amer. Math. Soc. 28 (1922), 109–110.
- [8] P. Erdős, On the normal number of prime factors of p-1 and some related problems concerning Euler's φ -function, Quart. J. Math. 6 (1935), 205–213.
- [9] _____, On pseudoprimes and Carmichael numbers, Publ. Math. Debrecen 4 (1956), 201–206.
- [10] _____, Solution of two problems of Jankowska, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astr. Phys. 6 (1958), 545–547.
- [11] _____, Some remarks on Euler's φ function, Acta Arith. 4 (1958), 10–19.
- [12] P. Erdős and C. Pomerance, On the normal number of prime factors of φ(n), Rocky Mountain J. Math. 15 (1985), 343–352.
- [13] K. Ford, The distribution of totients, Ramanujan J. 2 (1998), 67-151; see arXiv:1104.3264 [math.NT] for updates and corrections.
- [14] _____, The number of solutions of $\varphi(x) = m$, Ann. of Math. (2) **150** (1999), 283–311.
- [15] R. R. Hall and G. Tenenbaum, *Divisors*, Cambridge Tracts in Mathematics, vol. 90, Cambridge University Press, Cambridge, 1988.
- [16] G. H. Hardy and S. Ramanujan, The normal number of prime factors of a number n, Quart. J. Math. 48 (1917), 76–92.
- [17] M.I. Knopp, Modular functions in analytic number theory, Markham, Chicago, 1970.
- [18] F. Luca and P. Pollack, An arithmetic function arising from Carmichael's conjecture, J. Théor. Nombres Bordeaux 23 (2011), 697–714.
- [19] F. Luca and C. Pomerance, On the average number of divisors of the Euler function, Publ. Math. Debrecen 70 (2007), 125–148, errata in 89 (2016), 257–260.
- [20] A. Oppenheim, On an arithmetic function, J. London Math. Soc. 1 (1926), 205–211.
- [21] _____, On an arithmetic function (II), J. London Math. Soc. 2 (1927), 123–130.
- [22] P. Pollack, How often is Euler's totient a perfect power?, J. Number Theory 197 (2019), 1–12.
- [23] P. Pollack and J. Vandehey, Some normal numbers generated by arithmetic functions, Canad. Math. Bull. 58 (2015), 160–173.
- [24] C. Pomerance, Popular values of Euler's function, Mathematika 27 (1980), 84–89.
- [25] _____, On the distribution of pseudoprimes, Math. Comp. **37** (1981), 587–593.
- [26] _____, Two methods in elementary analytic number theory, Number theory and applications (Banff, AB, 1988), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 265, Kluwer Acad. Publ., Dordrecht, 1989, pp. 135–161.
- [27] G. Tenenbaum, Introduction to analytic and probabilistic number theory, third ed., Graduate Studies in Mathematics, vol. 163, American Mathematical Society, Providence, RI, 2015.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF GEORGIA, ATHENS, GA 30602

E-mail address: pollack@uga.edu