SOME PROBLEMS OF ERDŐS
ON THE SUM-OF-DIVISORS FUNCTION

PAUL POLLACK AND CARL POMERANCE

For Richard Guy on his 99th birthday. May his sequence be unbounded.

ABSTRACT. Let $\sigma(n)$ denote the sum of all of the positive divisors of n, and let $s(n) = \sigma(n) - n$ denote the sum of the proper divisors of n. The functions $\sigma(\cdot)$ and $s(\cdot)$ were favorite subjects of investigation by the late Paul Erdős. Here we revisit three themes from Erdős's work on these functions. First, we improve the upper and lower bounds for the counting function of numbers n with n deficient but $s(n)$ abundant, or vice versa. Second, we describe a heuristic argument suggesting the precise asymptotic density of n not in the range of the function $s(\cdot)$; these are the so-called *nonaliquot numbers*. Finally, we prove new results on the distribution of friendly k-sets, where a *friendly k-set* is a collection of k distinct integers which share the same value of $\frac{\sigma(n)}{n}$.

1. INTRODUCTION

Let $\sigma(n)$ be the sum of the natural number divisors of n, so that $\sigma(n) = \sum_{d|n} d$. Let $s(n)$ be the sum of only the proper divisors of n, so that $s(n) = \sigma(n) - n$. Interest in these functions dates back to the ancient Greeks, who classified numbers as *deficient*, *perfect*, or *abundant* according to whether $s(n) < n$, $s(n) = n$, or $s(n) > n$, respectively. For example, 10 is deficient, 12 is abundant, and 6 is perfect. The ancient Greeks also found and remarked on 2-cycles, where $s(n) = m$ and $s(m) = n$, calling such a pair *amicable*; for example, 220 and 284.

A conjecture of Catalan in 1888 [8], as amended by Dickson in 1913 [12], claims that for every positive integer n, the *aliquot* sequence $n, s(n), s(s(n)), \dots$ either terminates at 0 or enters some cycle, such as a perfect number, an amicable pair, or some longer cycle. Despite much numerical work, there are still 12 numbers $n < 1000$ for which the conjecture is in doubt, the first being 276. This numerical work uncovered long stretches where some aliquot sequences grow monotonically and exponentially, and other stretches where they decay in like manner. This behavior was taken into account in the Guy–Selfridge [25] counter-conjecture that for asymptotically all even numbers n, the aliquot sequence starting with n is unbounded.

It was proved by Lenstra [29] that for each k there is an aliquot sequence that strictly increases for k terms. Erdős [20] showed that this is common and, in fact, if n is abundant, then but for a set of numbers n of asymptotic density 0, the aliquot sequence starting with n strictly increases for k terms. In the same paper, Erdős claimed the analogous result for deficient numbers (with "increases" replaced with

"decreases"), but this claim was retracted in [22]. In that paper, the claim was shown for $k = 2$. It is conjectured that the full claim holds, but this has never been shown.

These results assert that if n is abundant, it is likely that $s(n)$ is also abundant, and similarly for deficient. Say n is an *aliquot reversal* if this fails. More precisely, say n is an *up-down reversal* if n is nondeficient yet $s(n)$ is deficient, and say n is a *down-up reversal* if n is deficient and $s(n)$ is nondeficient.

Though the terminology is new to this paper, aliquot reversals have been studied for some time because of their connection with amicable pairs; in fact the lesser member of an amicable pair is an up-down reversal, while the larger member is a down-up reversal. Erdős was the first to prove, in [15], that the set of numbers involved in an amicable pair has vanishing asymptotic density, and he did this by showing that there are few up-down reversals. The count of up-down reversals was improved later by Rieger [40], by Erdős–Rieger [21], and by the second author (see [36]). In that last paper it was shown that the number of $n \le x$ giving an up-down reversal is at most

$$(1.1) \qquad x/\exp((c + o(1))(\log_3 x \log_4 x)^{1/2}), \text{ as } x \to \infty,$$

for a certain absolute constant $c > 0$. Here, the subscripts indicate iteration of the natural log function. The proof uses results of Erdős [13] for the count of primitive nondeficient numbers. Incorporating an improvement on these results due to Avidon [3], this argument gives $c = 1$. In [33, Corollary 1.5], the same upper estimate (1.1) was shown for the number of down-up reversals in $[1, x]$, but only with the smaller constant $c = \frac{1}{10}$.

Our first theorem is an improvement on these upper bounds.

Theorem 1.1. *Both the count of down-up reversals in $[1, x]$ and the count of up-down reversals in $[1, x]$ are bounded above by (1.1) with $c = \sqrt{3}$.*

The proof uses recent results from [2] on the number of solutions n to congruences of the form $\sigma(n) \equiv a \pmod{n}$.

The problem of obtaining lower bounds on the number of aliquot reversals does not seem to have been previously considered. Note that the upper bounds of the last paragraph just barely give density 0. In the following results we show there is a reason for this: reversals are fairly common. We prove the following.

Theorem 1.2. *The number of up-down reversals in $[1, x]$ is $\gg x/(\log_2 x)(\log_3 x)^3$.*

Theorem 1.3. *The number of down-up reversals in $[1, x]$ is $\gg x/(\log_2 x)(\log_3 x)^2$.*

One might wonder if these results lend support to the Catalan–Dickson conjecture or to the Guy–Selfridge counter-conjecture. It is helpful to think in terms of parity. In fact, n and $s(n)$ have opposite parity precisely when n is a square or a double of a square, and such n comprise a very sparse set. In addition, it follows from the proof in [20] that for each fixed k, almost all n have the same parity as $s_k(n)$ (the kth iterate of s at n). Since an aliquot sequence terminates only if it hits a prime, a perhaps reasonable conjecture is that aliquot sequences beginning with even numbers usually either are unbounded or enter some cycle of even numbers. Numbers involved in such cycles have density 0 (see [28]). Numerical evidence indicates that most numbers involved in these cycles are amicable (i.e., they are in a 2-cycle), and we know that amicable numbers are quite sparsely distributed

(see [38]). Thus, despite the fact that our results show nonmonotonicity is fairly common, the Guy–Selfridge point of view is believable. And some cases of non-monotonicity are quite rare. For example, take numbers $n \equiv 6 \pmod{12}$, $n > 6$. They are abundant, but the number of them to x with $s(n)$ deficient is $O(x^{1/2})$. Further, in a recent preprint of Bosma [6], it is shown that about $1/3$ of the even numbers to 1 million have their aliquot sequence still open beyond 10^{99}. This evidence seems to put a chill on the Catalan–Dickson conjecture! On the other hand, it is hard to align this numerical evidence with the recent paper of Bosma–Kane [7] where it is shown that the geometric mean of $s(2n)/2n$ for $n \leq x$ tends to a constant smaller than 1 as $x \to \infty$.

We remark that the lower bounds in Theorems 1.2, 1.3 are established for even numbers. It would be of interest to see what kind of lower bounds can be established for odd numbers.

We turn now to our second topic. Positive integers not in the range of $s(\cdot)$ are said to be *nonaliquot*, though the more colorful term *untouchable* (see Alanen [1]) has also been used. The concept dates back to at least ca. 1000 CE [41]. A useful initial observation is that if p and q are distinct primes, then $s(pq) = p + q + 1$. A slightly stronger form of Goldbach's conjecture is that every even integer $n \geq 8$ is a sum of two distinct primes. If so, then every odd $n \geq 9$ belongs to the range of $s(\cdot)$; noting that $s(2) = 1$, $s(4) = 3$, and $s(8) = 7$, we would have that 5 is the only odd nonaliquot number.

While the conjecture of the last paragraph remains out of reach, it follows from (independent) work of Chudakov, van der Corput, and Estermann that almost all even numbers can be written as a sum of two distinct primes. (See [46, Chapter 3] for a modern treatment.) Hence, the set of odd nonaliquot numbers has asymptotic density zero.

What about even numbers? In [18], Erdős proved that a positive proportion of even numbers are nonaliquot. By an elaboration of Erdős's methods, Chen and Zhao [9] have shown that the nonaliquot even integers make up a set of lower density greater than 0.06. (This improves earlier results reported in [5] and [44, Corollary 9.2, p. 64].) Erdős was unable to decide whether or not a positive proportion of even numbers do belong to the image of s, but this has been resolved affirmatively in very recent work of Luca and the second author [30].

In §3, refining some thoughts in [44], we propose a probabilistic model for $s(\cdot)$ that suggests a precise value for the density of nonaliquot numbers.

Conjecture 1.4. *The set of nonaliquot numbers has asymptotic density Δ, where*

$$\Delta = \lim_{y \to \infty} \frac{1}{\log y} \sum_{\substack{a \leq y \\ 2 \mid a}} \frac{1}{a} e^{-a/s(a)}.$$

At $y = 2 \cdot 10^{10}$, the expression inside the limit has value 0.171822. Table 1 collects counts of nonaliquot numbers up to 10^{10}; these counts are consistent with a limiting density of ≈ 0.17.

Finally, we consider the distribution of the rational numbers $\frac{\sigma(n)}{n}$. We are interested here in statistics for

$$F(n; x) = \# \left\{ m \leq x : m \neq n, \ \frac{\sigma(m)}{m} = \frac{\sigma(n)}{n} \right\},$$

TABLE 1. Counts $U(x)$ of nonaliquots to various heights x from 10^8 through 10^{10}, along with the corresponding densities $D(x)$.

x	$U(x)$	$D(x)$	x	$U(x)$	$D(x)$
100000000	16246940	0.1625	1000000000	165826606	0.1658
200000000	32721193	0.1636	2000000000	333261274	0.1666
300000000	49265355	0.1642	3000000000	501171681	0.1671
400000000	65855060	0.1646	4000000000	669372486	0.1673
500000000	82468000	0.1649	5000000000	837801755	0.1676
600000000	99107582	0.1652	6000000000	1006383348	0.1677
700000000	115764316	0.1654	7000000000	1175094232	0.1679
800000000	132438792	0.1655	8000000000	1343935989	0.1680
900000000	149128373	0.1657	9000000000	1512867678	0.1681
			10000000000	1681871718	0.1682

TABLE 2. Counts of primitive friendly pairs. Here "Limit" is L with the "Count" being the number of pairs to 10^L and "Exponent" being that E such that Count $= 10^{LE}$.

Limit	3	4	5	6	7	8	9	10	11
Count	9	19	38	66	148	292	548	966	1688
Exponent	.318	.320	.316	.303	.310	.308	.304	.298	.293

thought of as a random variable on the set of natural numbers $n \leq x$, equipped with the uniform measure. A theorem of Wirsing [47] asserts that an equation of the form $\frac{\sigma(n)}{n} = \alpha$ has at most $x^{o(1)}$ solutions $n \leq x$, as $x \to \infty$, uniformly in α. Thus, we have a pointwise bound $F(n;x) = x^{o(1)}$, as $x \to \infty$. The average behavior of $F(n;x)$ was determined by Erdős [17, Theorem 2]; for a certain constant $c > 0$,

$$(1.2) \qquad \sum_{n \leq x} F(n;x) \sim cx, \quad \text{as} \quad x \to \infty.$$

Note that $\sum_{n \leq x} F(n;x)$ counts the number of ordered pairs of distinct integers $(n,m) \in [1,x]^2$ with $\frac{\sigma(n)}{n} = \frac{\sigma(m)}{m}$, and this is the way Erdős states his result.

Call a pair (unordered) of distinct integers n and m a *friendly pair* if $\frac{\sigma(n)}{n} = \frac{\sigma(m)}{m}$. Evidently, if u is coprime to nm, then un, um is also a friendly pair. Call the friendly pair n, m *primitive friendly* if n and m have no nontrivial common unitary divisor. (We say d is a *unitary divisor* of n, and write $d \parallel n$, if $d \mid n$ and $\gcd(d, n/d) = 1$.) The key ingredient in Erdős's proof of (1.2) is the convergence of $\sum \frac{1}{m}$, where m runs over the larger members of the primitive friendly pairs (with multiplicities). We prove a substantially stronger result on the distribution of such pairs:

Theorem 1.5. *The number of primitive friendly pairs contained in $[1,x]$ is at most $x^{\frac{1}{2}+o(1)}$, as $x \to \infty$.*

Erdős's theorem on the convergence of $\sum \frac{1}{m}$ follows immediately by partial summation. The exponent $\frac{1}{2}$ appears to be the limit of our method, but the numerical data (see Table 2) suggest that this is not sharp. Perhaps the true count of primitive friendly pairs in $[1,x]$ is $x^{o(1)}$.

TABLE 3. $N_1(x), N_2(x), N_3(x)$, and $N_4(x)$ for $x = 10^j$, with $3 \leq j \leq 10$.

	10^3	10^4	10^5	10^6	10^7	10^8	10^9	10^{10}
$N_1(x)$	37	348	3475	34746	347471	3474564	34745605	347456117
$N_2(x)$	3	30	283	2816	28089	280938	2809813	28097701
$N_3(x)$	0	12	82	857	8517	85365	853513	8535154
$N_4(x)$	0	0	10	85	853	8457	84605	845674

By studying primitive friendly k-sets for integers $k \geq 2$, we are able to show that $F(n; x)$ has a limiting distribution. Set $N_k(x) = \#\{n \leq x : F(n; x) \geq k\}$.

Theorem 1.6. *For each fixed nonnegative integer k, we have $N_k(x) = (\alpha_k + o(1))x$, as $x \to \infty$, for some constant α_k. Moreover, $\alpha_k \to 0$ as $k \to \infty$.*

Clearly, $\alpha_0 = 1$. Though the constants α_k for $k \geq 1$ are effectively computable (via the proof of Theorem 1.6), we have not found good rigorous numerical estimates for them. However, explicit computations of $N_k(x)$ suggest that $\alpha_1 = 0.0347\ldots$, $\alpha_2 = 0.0028\ldots$, $\alpha_3 = 0.00085\ldots$, and $\alpha_4 = 0.000084\ldots$; see Table 3.

Notation. We use the Bachmann–Landau $\asymp$, O and o-notations, with their usual meanings, as well as the associated Vinogradov $\ll$ and $\gg$ notations. $P(n)$ denotes the largest prime factor of n, with the convention that $P(1) = 1$, and $\mathrm{rad}(n)$ denotes the largest squarefree divisor of n. We reserve the letter p, with or without subscripts or dashes, for primes. We write $v_p(n)$ for the exponent on the highest power of p dividing n. The notation $\gcud(m, n)$ stands for the greatest common unitary divisor of m and n. We use φ for Euler's totient function, ω for the number of prime divisors, Ω for the number of prime factors counted with multiplicity, and τ for the total number of divisors.

2. ALIQUOT REVERSALS

2.1. **Upper bounds.**

2.1.1. *A structural lemma on primitive nondeficient numbers.* The proof of Theorem 1.1 runs along similar lines as the proof of the main theorem of [36]. In particular, the primitive nondeficient numbers play a critical role. (These are numbers n with $\sigma(n)/n \geq 2$ and for each proper divisor d of n, $\sigma(d)/d < 2$.) The main new ingredient is a lemma that is perhaps of independent interest: Almost all primitive nondeficient numbers n have $P(n) \lessapprox n^{1/3}$.

We will deduce the lemma from the results of [2] concerning solutions n to the congruence $\sigma(n) \equiv a \pmod n$. In that paper, n is called a *regular* solution if

$$n = pm, \quad \text{where} \quad p \nmid m, \ m \mid \sigma(m), \text{ and } \sigma(m) = a.$$

(It is straightforward to check that these numbers n are solutions of the congruence.) In the remaining case, n is called *sporadic*. Note that regular solutions can exist only when $a > 0$. The following is a slightly less precise form of [2, Theorem 1].

Proposition 2.1. *The number of sporadic solutions $n \leq x$ to $\sigma(n) \equiv a \pmod n$ is at most $x^{1/2 + o(1)}$, as $x \to \infty$, uniformly for $|a| \leq x^{1/4}$.*

The restriction $|a| \leq x^{1/4}$ is inconvenient for our purposes. We can ameliorate this as follows. Call a sporadic solution n to $\sigma(n) \equiv a \pmod{n}$ *convenient* if all of the following hold:

(i) $k := \frac{\sigma(n)-a}{n}$ is an integer at least 2,

(ii) if d is a unitary divisor of n with $d < n$, then $\frac{\sigma(d)}{d} < k$,

(iii) there is a prime $q > P(n)$ with $\frac{\sigma(nq)}{nq} > k$.

Proposition 2.2. *The number of convenient solutions $n \leq x$ to $\sigma(n) \equiv a \pmod{n}$ is at most $x^{1/2+o(1)}$, as $x \to \infty$, uniformly in a with $|a| \leq x$.*

Sketch of the proof. Hornfeck and Wirsing [27, Satz 2] have shown that there are $x^{o(1)}$ multiply perfect numbers $n \leq x$; hence, we may assume that $a \neq 0$.

While not phrased this way in [2], the proof of [2, Theorem 1] when $a \neq 0$ is handled there by reduction to the convenient case. Indeed, Lemma 5 in [2] asserts that if n is a sporadic solution to $\sigma(n) \equiv a \pmod{n}$ and $n > 6a^2 \log(6|a|)$, then n is convenient. In our setup here, we start by assuming n is convenient. Thus, the conclusion of [2, Lemma 5] holds trivially. Moreover, only the convenience hypothesis is needed in the proofs of Lemmas 6 and 7 from [2]. That is, both of those lemmas hold with the assumption of convenience replacing the hypothesis "n is a sporadic solution satisfying $n > 6a^2 \log(6|a|)$".

With these lemmas in place, we can again run the proof of [2, Theorem 1]. It proceeds exactly as before, except that now there is no need to assume $n > 6a^2 \log(6|a|)$. That assumption was responsible for the restriction $|a| \leq x^{1/4}$, which can now also be dispensed with. Following the proof through, we deduce that there are at most $x^{1/2+o(1)}$ convenient solutions $n \leq x$, uniformly for $|a| \leq x$. $\qquad\square$

We now deduce the key lemma.

Lemma 2.3. *Fix $\epsilon > 0$. The number of primitive nondeficient $n \leq x$ with $P(n) > n^{1/3+\epsilon}$ is $O_\epsilon(x^{1-\epsilon})$, as $x \to \infty$.*

Proof. After a dyadic subdivision argument, it is enough to prove the claimed upper bound for the number of primitive nondeficient $n \in (x, 2x]$ with $P(n) > x^{1/3+\epsilon}$. By [27] or [47], the number of perfect $n \leq 2x$ is $x^{o(1)}$, as $x \to \infty$. Thus, we can assume that n is abundant. We can also assume $P(n)^2 \nmid n$, since otherwise n belongs to a set of size at most $2x \sum_{p > x^{1/3+\epsilon}} p^{-2} \ll x^{2/3-\epsilon}$. Write $n = pm$, where $p = P(m)$. Since $n = mp$ is primitive nondeficient,

$$2 > \frac{\sigma(m)}{m} > \frac{2}{\sigma(p)/p} > 2 - 2x^{-1/3-\epsilon}.$$

Hence,

$$-2mx^{-1/3-\epsilon} < a := \sigma(m) - 2m < 0$$

and

$$|a| < 2mx^{-1/3-\epsilon} \leq 4\frac{x}{p}x^{-1/3-\epsilon} = 4x^{2/3-\epsilon}/p.$$

The integer m is a sporadic solution to $\sigma(m) \equiv a \pmod{m}$. Moreover, conditions (i)–(iii) in the definition of a convenient solution are easily verified for m, with $k = 2$ and $q = p$.

Fixing a and p, Proposition 2.2 shows that the number of corresponding values of $m \leq 2x/p$ is $O_\epsilon((x/p)^{1/2+\epsilon/2})$. Summing on $a \in (-4x^{2/3-\epsilon}/p, 0)$, the number of

possibilities for m given p is

$$\ll_\epsilon \frac{x^{2/3-\epsilon}}{p} \cdot \left(\frac{x}{p}\right)^{1/2+\frac{1}{2}\epsilon} \ll \frac{x^{7/6-\frac{1}{2}\epsilon}}{p^{3/2}}.$$

Summing on $p > x^{1/3+\epsilon}$ gives a final upper bound of $O_\epsilon(x^{1-\epsilon})$, as desired.   □

Remark. The proof can be improved to achieve the upper bound $x^{1-\frac{3}{2}\epsilon+o(1)}$, as $x \to \infty$. For our purposes, any power-savings upper bound is sufficient.

It would be desirable to prove Lemma 2.3 with a smaller exponent than $\frac{1}{3}$. If the lemma holds with $\frac{1}{R}$ (say), then the arguments below show that Theorem 1.1 holds with $c = \sqrt{R}$. In particular, if R can be taken arbitrarily large, we would have a qualitative improvement over any bound of the shape (1.1). Such an improvement would follow from a conjecture posed in [2]:

Conjecture 2.4. *Let $x \geq 3$, and let a be an integer with $|a| \leq x/2$. Let k be an integer with $k \geq 2$. The number of solutions $n \leq x$ to $\sigma(n) = kn + a$ is*

$$\ll (\log x)^C,$$

where both the implied constant and C are absolute constants.

Remark. This is a corrected form of the conjecture as it appeared in [2]. The original formulation allowed $k = 0$ and $k = 1$; these cases must be excluded to account for popular values a of $\sigma(n)$ or $s(n)$.

2.1.2. *Proof of Theorem* 1.1. Let

$$y = y(x) = \frac{\log_2 x}{2 \log_3 x}.$$

We begin with a lemma.

Lemma 2.5. *The number of integers $n \leq x$ for which $\sigma(n)$ is not divisible by every integer $m \leq y$ is $O(x/\log_2 x)$.*

Proof. We use [36, Theorem 2], which asserts that for coprime integers ℓ, k with $k > 0$, the number of integers $n \leq x$ such that no prime $p \equiv \ell \pmod{k}$ has $p \parallel n$ is $O(x/(\log x)^{1/\varphi(k)})$, uniformly. Applied with $\ell = -1, k = m$ the result implies that the number of n counted in the lemma is

$$\ll \sum_{m \leq y} \frac{x}{(\log x)^{1/\varphi(m)}} \leq \sum_{m \leq y} \frac{x}{(\log x)^{1/m}} \leq \sum_{m \leq y} \frac{x}{(\log_2 x)^2} \ll \frac{x}{\log_2 x}.   □$$

For future reference we state the result of Avidon [3] mentioned in the introduction.

Lemma 2.6. *The number of primitive nondeficient numbers $a \leq t$ is at most*

$$t/\exp((1+o(1))\sqrt{\log t \log_2 t}), \quad t \to \infty.$$

We shall also use a result of Toulmonde [45, Théorème 1] (see the discussion in §10) on the modulus of continuity of the distribution function for $\sigma(n)/n$.

Lemma 2.7. *If α is in the range of $\sigma(n)/n$, then uniformly for $T \geq 3$, the number of integers $m \leq x$ with*

$$\alpha - \frac{1}{T} \leq \frac{\sigma(m)}{m} < \alpha$$

is $\ll x/\exp(\frac{1}{5}\sqrt{\log T \log_2 T})$.

Proof of Theorem 1.1. Let $\epsilon > 0$ be fixed. Suppose $n \leq x$ is nondeficient with $s(n)$ deficient, and let a be the least nondeficient number that divides n. If each prime power dividing a is at most y, then Lemma 2.5 implies that but for $O(x/\log_2 x)$ choices for n, we may assume that $a \mid \sigma(n)$. This in turn implies that $a \mid s(n)$, contradicting that $s(n)$ is deficient. Thus, a has a prime power divisor exceeding y. If $a > y^{3-\epsilon}$, then Lemma 2.6 implies that the number of $n \leq x$ divisible by such a primitive nondeficient number a is at most $x/\exp((1 + o(1))\sqrt{(3 - \epsilon)\log_3 x \log_4 x})$ as $x \to \infty$. Hence we may assume that $a \leq y^{3-\epsilon}$ and a is divisible by a prime power $q > y$. If q is a proper power, the number of $n \leq x$ divisible by such a large power is $O(x/\sqrt{\log_2 x})$, so we may assume that q is prime and $q > a^{1/3+\epsilon/9}$. But then, by Lemma 2.3, the number of $n \leq x$ divisible by such a primitive nondeficient a is $O_\epsilon(x/y^{\epsilon/9})$. Since ϵ may be taken arbitrarily close to 0, our result follows.

The argument for down-up reversals is similar, but more difficult. Suppose $n \leq x$ is deficient and $s(n)$ is not deficient, and let a be the least nondeficient number dividing $s(n)$. Again, by Lemma 2.5, if each prime power dividing a is at most y, then we may assume that $a \mid \sigma(n)$. This implies that $a \mid n$, contradicting n deficient. Thus, we may assume a has some prime power divisor greater than y.

Write $n = mp$ where $p = P(n)$. By standard results on the distribution of integers all of whose prime factors are small (see [10]), we may assume that $p > x^{1/\log_3 x}$ and $p \nmid m$. We have $s(n) = ps(m) + \sigma(m) \equiv 0 \pmod{a}$. Let $a' = \gcd(a, s(m))$. The congruence implies that $a' \mid \sigma(m)$, so that $a' \mid m$. The congruence also implies that p is in a residue class modulo a/a'.

We first assume that

$$a \leq Z := x^{1/2\log_3 x}.$$

Given a, m, a', by the Brun–Titchmarsh inequality, the number of choices for $p \leq x/m$ is

$$\ll \frac{x}{m\varphi(a/a')\log(x/m(a/a'))} \ll \frac{x\log_3 x}{m\varphi(a/a')\log x} \ll \frac{x\log_3 x}{m(a/a')\log x},$$

where for the last step we use that a is primitive nondeficient, so that $\varphi(a/a') \gg a/a'$. We have $a' \mid m$, so summing the displayed upper bound over m and then over a', we get that the number of choices is

$$\ll \frac{\tau(a)}{a} x\log_3 x.$$

We can restrict to a which have no proper prime power divisor in $(y, Z]$. Indeed, by an argument analogous to that just seen, the number of n as above for which $s(n)$ is divisible by a given prime power $\ell^v \leq Z$ is $O(x\tau(\ell^v)\ell^{-v}\log_3 x) = O(xv\ell^{-v}\log_3 x)$. Summing over proper prime powers $\ell^v > y$, we find that only a negligible set of n arise in this manner. Since $a \leq Z$, we may assume that a is divisible by a prime $\ell > y$.

We now wish to find a good upper bound for the sum of $\tau(a)/a$ as a runs over primitive nondeficient numbers divisible by a prime greater than y. We first assume that $a \leq y^{3-\epsilon}$. Then $P(a) > a^{1/3+\epsilon/9}$. Using $\tau(a) = a^{o(1)}$ as $a \to \infty$ and Lemma 2.3 we have $\sum \tau(a)/a \leq y^{-\epsilon/9+o(1)}$, which is sufficient for our needs. So now assume that $a > y^{3-\epsilon}$. If also $\tau(a) \leq \exp(\epsilon\sqrt{\log a \log_2 a})$, a brief calculation with Lemma 2.6 suffices.

Now assume that $y^{3-\epsilon} < a \leq Z$ and $\tau(a) > \exp(\epsilon\sqrt{\log a \log_2 a})$. We forget at this point that a is running over primitive nondeficient integers. For each fixed

positive integer j,

$$(2.1) \qquad \sum_{k \leq T} \frac{\tau(k)^j}{k} \leq \prod_{p \leq T} \left(1 + \frac{2^j}{p} + O_j \left(\frac{1}{p^2} \right) \right) \ll_j (\log T)^{2^j}.$$

Let $T \geq y^{3-\epsilon}$ and consider values of $k \in (T, T^2]$ with

$$\tau(k) > z(T) := \exp(\epsilon \sqrt{\log T \log_2 T}).$$

We have

$$\sum_{\substack{T < k \leq T^2 \\ \tau(k) > z(T)}} \frac{\tau(k)}{k} \leq \frac{1}{z(T)^{j-1}} \sum_{T < k \leq T^2} \frac{\tau(k)^j}{k} \ll_j \frac{(\log T)^{2^j}}{z(T)^{j-1}}.$$

Thus, letting T range over numbers of the form $(y^{3-\epsilon})^{2^i}$, for $i = 0, 1, 2, \ldots$,

$$\sum_{\substack{a > y^{3-\epsilon} \\ \tau(a) > z(a)}} \frac{\tau(a)}{a} \ll_j \frac{(\log y)^{2^j}}{z(y^{3-\epsilon})^{j-1}}.$$

If we choose $j = \lceil \epsilon^{-1} \rceil$, and let ϵ be arbitrarily small, this estimate completes the argument in the case $a \leq Z$.

Now we assume $a > Z$. Since $a \mid s(n)$ and $s(n) \leq n$, we have $a \leq x$. Consider the "Z-prefix" divisor b of a. This is the largest divisor of a in $[1, Z]$ with each prime dividing b bounded above by each prime dividing a/b. We may assume that

$$(2.2) \qquad\qquad\qquad b > Z^{1/3}.$$

Indeed, if $b \leq Z^{1/3}$, then every prime factor of a/b is greater than $Z^{2/3}$, so that

$$\frac{\sigma(a/b)}{a/b} = 1 + O \left(\frac{\log x}{Z^{2/3}} \right),$$

since a/b has fewer than $\log x$ distinct prime factors. Also, since b is deficient, we have

$$\frac{\sigma(b)}{b} \leq 2 - \frac{1}{b} \leq 2 - \frac{1}{Z^{1/3}}.$$

The last two displays imply that for x sufficiently large, we have a deficient, a contradiction. Thus, we have (2.2).

Let $\mathscr{B}$ denote the set of integers $b \in (Z^{1/3}, Z]$ which are the Z-prefix for some primitive nondeficient $a \in (Z, x]$. We will show that there are very few integers $n \leq x$ with $s(n)$ divisible by some member of $\mathscr{B}$. To do this, we show that the set $\mathscr{B}$ is rather sparse.

Let $Y := Z^{1/\log_2 x}$, and assume that $b \in \mathscr{B}$ and $P(b) > Y$. Then, as above,

$$\frac{\sigma(a/b)}{a/b} = 1 + O \left(\frac{\log x}{Y} \right),$$

so for a to be nondeficient, there is some absolute constant $c > 0$ such that

$$2 > \frac{\sigma(b)}{b} \geq 2 - \frac{c \log x}{Y}.$$

By Lemma 2.7, for $t \in (Z^{1/3}, Z]$, the number of integers $b \leq t$ which satisfy the last display is at most $t / \exp(\frac{1}{8} \sqrt{\log x / \log_3 x})$ for all sufficiently large x. (Here 8 could be replaced with any number exceeding $5\sqrt{2}$.)

We also wish to count the number of $b \in \mathscr{B}$, $b \leq t$ with $P(b) \leq Y$. For this we ignore the condition $b \in \mathscr{B}$ and use only a standard result in [10]: the number of such b is at most $t/(\log x)^{\frac{1}{4} \log_3 x}$ for all sufficiently large x.

As in the case $a \leq Z$, we find that the number of integers $n \leq x$ with $b \mid s(n)$ for some $b \in \mathscr{B}$ (and assuming $n = pm$ with $p > x^{1/\log_3 x}$ and $p \nmid m$) is

$$\ll x \log_3 x \sum_{b \in \mathscr{B}} \frac{\tau(b)}{b}.$$

Again as before we treat the cases of $\tau(b)$ small and large separately. If $\tau(b) \leq (\log x)^{\frac{1}{8} \log_3 x}$, then our estimates for the number of $b \leq t$ suffice to give an acceptable estimate. If $\tau(b) > (\log x)^{\frac{1}{8} \log_3 x}$, we use (2.1) with $j = 2$, getting

$$\sum_{\substack{Z^{1/3} < b \leq Z \\ \tau(b) > (\log x)^{\frac{1}{8} \log_3 x}}} \frac{\tau(b)}{b} \leq \frac{1}{(\log x)^{\frac{1}{8} \log_3 x}} \sum_{Z^{1/3} < b \leq Z} \frac{\tau(b)^2}{b} \ll (\log x)^{4 - \frac{1}{8} \log_3 x}.$$

This too is sufficiently small, so completing our proof. $\qquad\square$

2.2. **Lower bounds.** Here we prove Theorems 1.2, 1.3.

Proof of Theorem 1.2. Let x be large, let k be the least integer with $2^k > \log_2 x$, and let p be a prime with $2^k < p < 2^{k+1} - 1$. We shall consider numbers $n = 2^k pmq \leq x$ where $x^{1/6} < m \leq x^{1/3}$ and q is prime with $q > x^{1/3}$. Since $2^k p$ is abundant, such numbers n are abundant. With no further restrictions on the parameters m, q, the number of n's created in this way is $\gg x/\log_2 x \log_3 x$. We will show that with some mild restrictions on m, q we can be assured that $s(n)$ is deficient, and these restrictions will only reduce our count by a factor of $(\log_3 x)^2$.

We shall want the numbers m to satisfy:

(i) $2^k \mid \sigma(m)$,
(ii) $p \nmid \sigma(m)$,
(iii) m has no prime factors below $(\log_2 x)^2$.

We can ensure m satisfies (ii) and (iii) by eliminating those m divisible by any of

- a proper prime power $r^\alpha \geq p/2$,
- a prime $r \equiv -1 \pmod{p}$,
- a prime $r \leq (\log_2 x)^2$.

Let $t \in [x^{1/6}, x^{1/3}]$. By Brun's sieve [26, Theorem 2.2], there is an absolute constant $\eta > 0$ such that the number of $m \leq t$ divisible by no prime $r \leq t^\eta$ in the second or third bullet points is

$$\gg t \prod_{r \leq (\log_2 x)^2} \left(1 - \frac{1}{r}\right) \prod_{\substack{r \leq t^\eta \\ r \equiv -1 \pmod{p}}} \left(1 - \frac{1}{r}\right)$$

$$\gg \frac{t}{\log_3 x} \exp\left(- \sum_{\substack{r \leq t^\eta \\ r \equiv -1 \pmod{p}}} \frac{1}{r}\right) \gg \frac{t}{\log_3 x}.$$

(The sum on r here is $O(1)$, by the Brun–Titchmarsh inequality and the bound $p \gg \log_2 t$.) The number of $m \leq t$ divisible by some prime $r > t^\eta$ with $r \equiv -1 \pmod{p}$ is $\ll t \sum_{t^\eta < r \leq t, \ r \equiv -1 \pmod{p}} \frac{1}{r} \ll t/p \ll t/\log_2 x$, which is $o(t/\log_3 x)$, as $x \to \infty$. Finally, the number of $m \leq t$ divisible by some proper prime power

$r^\alpha \geq p/2$ is $\ll t/\sqrt{p} \ll t/\sqrt{\log_2 x}$, and this is also $o(t/\log_3 x)$. Thus, there are $\gg t/\log_3 x$ values of $m \leq t$ satisfying (ii) and (iii). In fact, this is the correct order of magnitude for this count, since (iii) by itself already implies an upper bound of the same form.

We now turn to (i). We apply the Turán–Kubilius inequality [43, Theorem 1, p. 302] to investigate the normal order of the additive function $v_2(\sigma(m))$. Uniformly for $x^{1/6} \leq t \leq x^{1/3}$,

$$\sum_{p^v \leq t} v_2(\sigma(p^v))p^{-v}(1-p^{-1}) \sim 2\log_2 x$$

and

$$\sum_{p^v \leq t} v_2(\sigma(p^v))p^{-v} \ll \log_2 x.$$

Hence, the number of positive integers $m \leq t$ where $v_2(\sigma(m)) < \log_2 x$ is $O(t/\log_2 x)$. Since $k \ll \log_3 x$, this is also an upper bound on the number of failures of (i).

We conclude that there are $\asymp t/\log_3 x$ values of $m \leq t$ satisfying all of (i)–(iii). By partial summation,

$$(2.3) \qquad \sum_m \frac{1}{m} \asymp \frac{\log x}{\log_3 x},$$

where the sum is over $m \in (x^{1/6}, x^{1/3}]$ satisfying (i)–(iii).

For each m, let $f_0(m) = \sum_{r|m} 1/r$, where r runs over primes. Using that m is free of prime factors up to $(\log_2 x)^2$ gives

$$\sum_m \frac{1}{m} f_0(m) = \sum_{r > (\log_2 x)^2} \frac{1}{r} \sum_{m:\, r|m} \frac{1}{m} \leq \sum_{r > (\log_2 x)^2} \frac{1}{r^2} \prod_{\substack{p \leq x \\ p > (\log_2 x)^2}} (1 - 1/p)^{-1},$$

which is $\ll \frac{\log x}{(\log_2 x)^2(\log_3 x)^2}$. In what follows, we will assume that

$$(2.4) \qquad f_0(m) \leq \frac{1}{2(\log_2 x)^2}.$$

Note that discarding contrary values of m changes $\sum_m \frac{1}{m}$ only by $O(\log x/(\log_3 x)^2)$ and so preserves the validity of (2.3).

We now complete the construction of a number $n = 2^k pmq$ by choosing a prime $q \in (x^{1/3}, x/2^k pm]$. Suppose an odd prime r divides $s(n)$. Then

$$s(n) = qs(2^k pm) + \sigma(2^k pm) \equiv 0 \pmod{r}.$$

If $r \mid s(2^k pm)$, then $r \mid \sigma(2^k pm)$, so that $r \mid 2^k pm$. Such a prime r cannot be 2 by assumption, and it cannot be p, since $p \nmid \sigma(2^k)$ and $p \nmid \sigma(m)$. Thus, $r \mid m$, and so $r > (\log_2 x)^2$. We conclude that if $2 < r \leq (\log_2 x)^2$ and $r \mid s(n)$, then $r \nmid s(2^k pm)$. We now restrict the q's that are available to us by removing for each odd prime $r \leq (\log_2 x)^2$ a single coprime residue class that will ensure that $r \nmid s(n)$. By another application of Brun's sieve, this cuts down the number of choices for q by a factor proportional to $\log_3 x$. For a surviving prime q, let

$$f_1(q) = \sum_{\substack{2 < r < x^{1/4} \\ r \nmid m \\ r | s(2^k pmq)}} \frac{1}{r}.$$

If $r \mid s(2^k pmq)$ and $r \nmid m$, then as we have seen, q is in a fixed residue class mod r. Using that q also avoids one residue class modulo each prime not exceeding $(\log_2 x)^2$, the Brun–Titchmarsh inequality implies that the average value of $f_1(q)$, over surviving primes q, is $\ll 1/((\log_2 x)^2 \log_3 x)$. Keeping (2.4) in mind, we conclude that there are $\gg \frac{\pi(x/2^k pm)}{\log_3 x} \gg \frac{x}{2^k pm \log x \log_3 x}$ values of q for which

$$(2.5) \qquad \sum_{\substack{r>2 \\ r \mid s(2^k pmq)}} \frac{1}{r} \le \frac{1}{(\log_2 x)^2}.$$

We finally note that for m, q chosen as above, we have $2^{k+1} \mid \sigma(2^k pmq)$, so that $2^k \parallel s(2^k pmq)$. Hence, $s(2^k pmq)/2^k$ has only odd prime factors. From (2.5), we conclude that $s(2^k pmq)$ is deficient. The theorem then follows with the estimate:

$$\sum_p \sum_m \sum_q 1 \gg \sum_p \sum_m \frac{x}{2^k pm \log x \log_3 x} \gg \sum_p \frac{x}{2^k p (\log_3 x)^2}$$
$$\gg \frac{x}{2^k (\log_3 x)^3} \gg \frac{x}{\log_2 x (\log_3 x)^3},$$

where we have used (2.3). $\qquad \square$

Proof of Theorem 1.3. We take k as the least integer with $2^k > 10 \log_2 x$. We consider numbers $n = 2^k mqr \le x$ with q, r primes, $x^{1/8} < m \le x^{1/4}$, $x^{1/4} < q \le x^{1/3}$, and $r > x^{1/3}$. As in the last proof, we impose some conditions on m. Namely, we insist that m satisfy all of

(i') m has no prime factors below $(\log_2 x)^2$,
(ii') $2^k \mid \sigma(m)$,
(iii') (2.4) holds,
(iv') for $I := [2^k, 2^{k+1} - 1)$, we have $\sum_{p \in I, \, p \nmid \sigma(m)} \frac{1}{p} > \frac{1}{10} \frac{1}{\log_3 x}$.

We will show we can construct a collection of such m that satisfies

$$(2.6) \qquad \sum_m \frac{1}{m} \gg \frac{\log x}{\log_3 x}.$$

It will turn out that the most stringent restriction is the first one.

By an elementary sieve, those $m \in (x^{1/8}, x^{1/4}]$ satisfying (i') have

$$(2.7) \qquad \sum_m \frac{1}{m} \sim \frac{1}{16 e^\gamma} \frac{\log x}{\log_3 x},$$

as $x \to \infty$. By arguments seen above, restricting m further to satisfy (ii') changes the value of $\sum_m \frac{1}{m}$ by $O(\log x / \log_2 x)$, and restricting m to satisfy (iii') affects the sum only by $O(\log x / (\log_3 x)^2)$. Hence, (2.7) still holds if m is restricted to satisfy all of (i')–(iii').

We now turn to (iv'). Changing $\sum_m \frac{1}{m}$ by $O(\log x / \sqrt{\log_2 x})$, we can assume that m is not divisible by any proper prime power exceeding $5 \log_2 x$. If $p \in I$ and $p \mid \sigma(m)$, then there is a prime power $\ell^\alpha \parallel m$ with $p \mid \sigma(\ell^\alpha)$; since $\ell^\alpha > p/2 > 5 \log_2 x$, it follows that $\alpha = 1$. Let

$$f_2(m) = \sum_{\substack{p \in I \\ p \mid \sigma(m)}} \frac{1}{p}.$$

Using that $m \leq x^{1/4}$ and that m is free of prime factors up to $(\log_2 x)^2$, we see that as $x \to \infty$,

$$\sum_m \frac{1}{m} f_2(m) \leq \sum_{p \in I} \frac{1}{p} \sum_{\ell \equiv 1 \ (\mathrm{mod}\ p)} \sum_{m: \ \ell | m} \frac{1}{m}$$

$$\leq \left(\frac{1}{8} + o(1) \right) \frac{\log x}{\log_3 x} \cdot \sum_{p \in I} \frac{1}{p} \sum_{\substack{\ell \leq x \\ \ell \equiv 1 \ (\mathrm{mod}\ p)}} \frac{1}{\ell}.$$

From [36, Theorem 1 and Remark 1], the final sum on ℓ is at most $(1 + o(1)) \frac{\log_2 x}{p}$, as $x \to \infty$, uniformly for $p \in I$. Inserting this estimate above,

$$\sum_m \frac{1}{m} f_2(m) \leq \left(\frac{1}{8} + o(1) \right) \frac{\log x}{\log_3 x} \log_2 x \cdot \sum_{p \in I} \frac{1}{p^2} \leq \left(\frac{1}{80} + o(1) \right) \frac{\log x}{(\log_3 x)^2}.$$

Hence, the contribution to the left-hand side of (2.7) from those values of m with $f_2(m) \geq \frac{1}{2 \log_3 x}$ is at most $(\frac{1}{40} + o(1)) \frac{\log x}{\log_3 x}$. Since $\frac{1}{16 e^\gamma} - \frac{1}{40} > \frac{1}{100}$, after discarding these m we may still assert (2.6). Moreover, since $\sum_{p \in I} \frac{1}{p} \sim \frac{\log 2}{\log_3 x}$ and $\log 2 - \frac{1}{2} > \frac{1}{10}$, we have

$$\sum_{\substack{p \in I \\ p \nmid \sigma(m)}} \frac{1}{p} > \frac{1}{10} \frac{1}{\log_3 x}$$

for all remaining m (and all large x).

Note that for any choice of distinct primes q, r greater than m, the number $2^k m q r$ is deficient; indeed, for large x,

$$\frac{\sigma(2^k m q r)}{2^k m q r} < 2 \left(1 - \frac{1}{2^{k+1}} \right) \left(\prod_{p | m} \left(1 + \frac{1}{p-1} \right) \right) \left(1 + \frac{1}{x^{1/8}} \right)^2$$

$$< 2 \left(1 - \frac{1}{40 \log_2 x} \right) \left(1 + O \left(\frac{1}{(\log_2 x)^2} \right) \right) \left(1 + O \left(\frac{1}{x^{1/8}} \right) \right),$$

and this is $< 2(1 - \frac{1}{80 \log_2 x}) < 2$. We shall show that with some mild restrictions on q, r, we will have $s(2^k m q r)$ abundant. For m satisfying the above restrictions, let $S(m)$ be the set of primes $p \in I$ with $p \nmid \sigma(m)$ and let S_m be the product of the members of $S(m)$.

We next consider primes $q \in (x^{1/4}, x^{1/3}]$. They are chosen with the requirements that both $q + 1$ and $s(2^k m q)$ are coprime to S_m. We have $s(2^k m q) = q s(2^k m) + \sigma(2^k m)$. By our choices of k, m, we have $\sigma(2^k m)$ coprime to S_m. So for a prime $p \mid \gcd(s(2^k m q), S_m)$, q is in a particular residue class mod p. Similarly this holds too for $p \mid q + 1$. Thus, the density of the primes q which avoid these classes is at least

$$\prod_{p | S_m} \frac{p - 3}{p - 1} \sim 1, \quad \text{as} \quad x \to \infty.$$

So, almost all of the primes $q \in (x^{1/4}, x^{1/3}]$ may be used, and for these primes q, we have $\gcd(s(2^k m q), S_m) = 1$ and $\gcd(q + 1, S_m) = 1$.

We now choose primes $r \in (x^{1/3}, x/(2^k m q)]$ with $\gcd(s(2^k m q r), S_m) > 1$. For such a prime r, $s(2^k m q r)$ is abundant. Let $p \mid S_m$. Note that for any prime $r \nmid 2^k m q$, we have $s(2^k m q r) = r s(2^k m q) + \sigma(2^k m q)$. Further, we have $p \nmid s(2^k m q)$

and $p \nmid \sigma(2^k mq)$. By the prime number theorem for progressions, the fraction of r with $p \mid s(2^k mqr)$ is asymptotically $1/(p-1)$. Similarly, if $pp' \mid S_m$, the fraction of choices for r with $pp' \mid s(2^k mqr)$ is asymptotically $O(1/pp')$. Since

$$\sum_{p \mid S_m} \frac{1}{p-1} \gg \frac{1}{\log_3 x}, \quad \sum_{pp' \mid S_m} \frac{1}{pp'} \ll \frac{1}{(\log_3 x)^2},$$

we have that the proportion of primes r with $\gcd(s(2^k mqr), S_m) > 1$ is $\gg 1/\log_3 x$.

Putting these thoughts together, we have that the number of choices for m, q, r is

$$\sum_m \sum_q \sum_r 1 \gg \sum_m \sum_q \frac{x}{2^k mq \log x \log_3 x} \gg \sum_m \frac{x}{2^k m \log x \log_3 x} \gg \frac{x}{2^k (\log_3 x)^2}.$$

Since $2^k \ll \log_2 x$, this completes the proof of the theorem. $\square$

3. The density of nonaliquot numbers

3.1. A heuristic argument. For a positive integer y, let

$$A_y = \mathrm{lcm}[1, 2, \ldots, y].$$

For a positive integer $a \mid A_y$, let

$$T_a = \{n : \gcd(n, A_y) = a\},$$

and let $T_a(x)$ denote the set of members of $T_a \cap [1, x]$. We assume that s maps T_a to T_a for each $a \mid A_y$. This is asymptotically true if $n > e^{e^y}$ and $y \to \infty$. Further, we assume that for $n \in T_a$ we have $\sigma(n)/n \approx \sigma(a)/a$. This is asymptotically true up to sets of vanishing density as $y \to \infty$. We assume in addition that s is a random map subject to the prior assumptions.

Now suppose a is even. If $s(m) \in T_a(x)$, then $m \in T_a$ and $m \le x/(\sigma(a)/a-1) = (a/s(a))x$, so that $m \in T_a((a/s(a))x)$. The probability that an integer $n \in T_a(x)$ is not in the range of s is about

$$\left(1 - \frac{1}{|T_a(x)|}\right)^{|T_a((a/s(a))x)|}.$$

Now $|T_a(x)| \sim (\varphi(A_y)/aA_y)x$, so the above probability is asymptotically

$$\exp\left(-\frac{a}{s(a)}\right).$$

We thus would have that the density Δ of even numbers not in the range of s is

$$(3.1) \qquad \Delta = \lim_{y \to \infty} \frac{\varphi(A_y)}{A_y} \sum_{a \mid A_y}^{*} \frac{1}{a} e^{-a/s(a)},$$

where the $*$ indicates that a runs over even numbers. Since almost all odd numbers are in the image of s, Δ is also the density of nonaliquot numbers.

We now prove that the limit (3.1) exists. Let

$$P_y = \frac{\varphi(A_y)}{A_y} = \prod_{p \le y}\left(1 - \frac{1}{p}\right), \quad S_y = \sum_{P(a) \le y}^{*} \frac{1}{a} e^{-a/s(a)}, \quad S_y' = \sum_{a \mid A_y}^{*} \frac{1}{a} e^{-a/s(a)}.$$

Note that $S'_y \leq S_y$. The difference comes from those a with $P(a) \leq y$ and $a \nmid A_y$. Such a number a is divisible by p^k for some prime power $p^k > y$ and $p \leq y$. Writing $a = p^k a'$, we have

$$(3.2) \qquad \sum_{\substack{P(a) \leq y \\ a \nmid A_y}}^* \frac{1}{a} e^{-a/s(a)} \leq \sum_{p \leq y} \sum_{k > \frac{\log y}{\log p}} \frac{1}{p^k} \sum_{P(a') \leq y} \frac{1}{a'} \ll \log y \sum_{p \leq y} \frac{1}{y} \ll 1.$$

Since, by Mertens' theorem, $P_y \sim e^{-\gamma}/\log y$ as $y \to \infty$, we have

$$P_y S_y - P_y S'_y \ll \frac{1}{\log y}.$$

Thus the limit in (3.1) exists if and only if $\lim_{y \to \infty} P_y S_y$ exists, in which case this latter limit is also Δ.

For a prime q, we have $P_q = P_{q-1}(1 - \frac{1}{q})$ and, for $q > 2$,

$$S_q - S_{q-1} = \sum_{P(a) < q}^* \sum_{i \geq 1} \frac{1}{aq^i} e^{-aq^i/s(aq^i)} > \sum_{P(a) < q}^* \sum_{i \geq 1} \frac{1}{aq^i} e^{-a/s(a)} = \frac{1}{q-1} S_{q-1}.$$

Thus,

$$P_q S_q > P_{q-1}\left(1 - \frac{1}{q}\right) S_{q-1}\left(1 + \frac{1}{q-1}\right) = P_{q-1} S_{q-1}.$$

Since $P_y S_y$ is always at most $\frac{1}{2}$, it follows that the limit in (3.1) exists and

$$(3.3) \qquad \Delta = \lim_{y \to \infty} \frac{e^{-\gamma}}{\log y} \sum_{a | A_y}^* \frac{1}{a} e^{-a/s(a)} = \lim_{y \to \infty} \frac{e^{-\gamma}}{\log y} \sum_{P(a) \leq y}^* \frac{1}{a} e^{-a/s(a)}.$$

We would like to compute a numerical approximation to Δ, and it may seem that the first expression in (3.3) is more easily computed than the second, since it involves a finite sum while the latter involves an infinite sum. However, we claim the infinite sum can be replaced with the sum over $a \leq y$, while changing $e^{-\gamma}$ to 1. That is,

$$(3.4) \qquad \Delta = \lim_{y \to \infty} \frac{1}{\log y} \sum_{a \leq y}^* \frac{1}{a} e^{-a/s(a)}.$$

The equation (3.4) may be proved as follows: Introduce a parameter $z \geq 2$, which will eventually be chosen as a slow-growing function of y. Define $h_z(n) = \sum_{d | n, \, 1 < d \leq z} 1/d$. For even a,

$$0 \leq e^{-\frac{a}{s(a)}} - e^{-\frac{1}{h_z(a)}} = e^{-\frac{a}{s(a)}}\left(1 - e^{\frac{1}{s(a)/a} - \frac{1}{h_z(a)}}\right) \ll e^{-\frac{a}{s(a)}}\left(\frac{s(a)}{a} - h_z(a)\right).$$

Thus,

$$0 \leq \sum_{a \leq y}^* \frac{1}{a} e^{-\frac{a}{s(a)}} - \sum_{a \leq y}^* \frac{1}{a} e^{-\frac{1}{h_z(a)}} \ll \sum_{a \leq y}^* \frac{1}{a} e^{-\frac{a}{s(a)}} \sum_{\substack{d | a \\ d > z}} \frac{1}{d}$$

$$(3.5) \qquad\qquad\qquad\qquad \leq \sum_{d > z} \frac{1}{d} \sum_{\substack{a \leq y \\ d | a}}^* \frac{1}{a} \leq \log y \sum_{d > z} \frac{1}{d^2} \ll \frac{1}{z} \log y.$$

A nearly identical computation gives

$$(3.6) \qquad \sum_{P(a) \leq y}^* \frac{1}{a} e^{-a/s(a)} - \sum_{P(a) \leq y}^* \frac{1}{a} e^{-\frac{1}{h_z(a)}} \ll \frac{1}{z} \log y.$$

Note that $h_z(a) = h_z(\gcd(a, A_z))$, and that when a is even, $d = \gcd(a, A_z)$ is also even. Hence, for all large y we have

$$\sideset{}{^*}\sum_{P(a)\leq y} \frac{1}{a} e^{-\frac{1}{h_z(a)}} = \sum_{\substack{d\mid A_z \\ 2\mid d}} \sum_{\substack{P(a)\leq y \\ \gcd(a,A_z)=d}} \frac{1}{a} e^{-\frac{1}{h_z(d)}} = \sum_{\substack{d\mid A_z \\ 2\mid d}} e^{-\frac{1}{h_z(d)}} \sum_{\substack{P(a)\leq y \\ d\mid a}} \frac{1}{a} \sum_{\substack{j\mid a/d \\ j\mid A_z/d}} \mu(j)$$

$$= \sum_{\substack{dj\mid A_z \\ 2\mid d}} \mu(j) e^{-\frac{1}{h_z(d)}} \sum_{\substack{P(a)\leq y \\ dj\mid a}} \frac{1}{a} = \sum_{\substack{dj\mid A_z \\ 2\mid d}} \frac{\mu(j) e^{-\frac{1}{h_z(d)}}}{dj} \sum_{P(a)\leq y} \frac{1}{a}.$$

Again almost identically, we have

$$\sideset{}{^*}\sum_{a\leq y} \frac{1}{a} e^{-\frac{1}{h_z(d)}} = \sum_{\substack{dj\mid A_z \\ 2\mid d}} \frac{\mu(j)}{dj} e^{-\frac{1}{h_z(d)}} \sum_{a\leq y/dj} \frac{1}{a}.$$

By Mertens' theorem (with the elementary error estimate shown in [43, Theorem 10, p. 17]),

$$\frac{1}{\log y} \sum_{P(a)\leq y} \frac{1}{a} = \frac{e^\gamma}{\log y} \sum_{a\leq y} \frac{1}{a} + O\left(\frac{1}{\log y}\right) = \frac{e^\gamma}{\log y} \sum_{a\leq y/dj} \frac{1}{a} + O\left(\frac{\log(dj)+1}{\log y}\right).$$

Making this substitution, we find that

$$(3.7) \quad \left| \frac{e^{-\gamma}}{\log y} \sideset{}{^*}\sum_{P(a)\leq y} \frac{1}{a} e^{-\frac{1}{h_z(a)}} - \frac{1}{\log y} \sideset{}{^*}\sum_{a\leq y} \frac{1}{a} e^{-\frac{1}{h_z(a)}} \right| \ll \frac{1}{\log y} \sum_{dj\mid A_z} \frac{\log dj}{dj} \ll \frac{(\log z)^3}{\log y}.$$

By choosing $z = \exp((\log y)^{1/4})$, say, from (3.2), (3.5), (3.6), and (3.7), we see that

$$\left| \frac{e^{-\gamma}}{\log y} \sideset{}{^*}\sum_{a\mid A_y} \frac{1}{a} e^{-a/s(a)} - \frac{1}{\log y} \sideset{}{^*}\sum_{a\leq y} \frac{1}{a} e^{-a/s(a)} \right| \ll \frac{1}{(\log y)^{1/4}}.$$

Thus, (3.4) follows. This completes the heuristic justification of Conjecture 1.4.

3.2. **Computations.** The following strategy was used to construct Table 1. Explicit computations on Goldbach's conjecture have shown that every even number in $[8, 4 \cdot 10^{18}]$ is a sum of two distinct primes p and q (see [31]). Hence, 5 is the only odd nonaliquot number up to 10^{18}. In [39, §5], an algorithm is described that enumerates the elements of $s(\mathbf{N}) \cap [1, N]$ in time $\tilde{O}(N)$. We ran the algorithm to $N = 10^{10}$. We made one modification, necessitated by memory constraints: Instead of using a 0-1 array to record which numbers had appeared so far, the even numbers in the range of s were written to disk as they were found by the algorithm, and sorted later with the GNU `sort` utility. The final data file required 36.1 GB.

On the computation of an approximation for Δ, this is straightforward using (3.4) for large values of y. Note that Δ is effectively computable, since the above calculations show that for each fixed integral value of $z \geq 2$,

$$\lim_{y\to\infty} \frac{e^{-\gamma}}{\log y} \sum_{\substack{dj\mid A_z \\ 2\mid d}} \frac{\mu(j)}{dj} e^{-\frac{1}{h_z(d)}} \sum_{P(a)\leq y} \frac{1}{a} = \sum_{\substack{dj\mid A_z \\ 2\mid d}} \frac{\mu(j)}{dj} e^{-\frac{1}{h_z(d)}}$$

TABLE 4. Counts $U_\varphi(x)$ of numbers not in the range of $s_\varphi(\cdot)$ to various heights x from 10^8 through 10^{10}, along with the corresponding densities $D_\varphi(x)$.

x	$U_\varphi(x)$	$D_\varphi(x)$	x	$U_\varphi(x)$	$D_\varphi(x)$
100000000	11355049	0.1136	1000000000	113482572	0.1135
200000000	22718595	0.1136	2000000000	226723208	0.1134
300000000	34079552	0.1136	3000000000	339811983	0.1133
400000000	45433178	0.1136	4000000000	452815609	0.1132
500000000	56783449	0.1136	5000000000	565739195	0.1131
600000000	68129835	0.1135	6000000000	678597689	0.1131
700000000	79472816	0.1135	7000000000	791405364	0.1131
800000000	90811869	0.1135	8000000000	904168542	0.1130
900000000	102146368	0.1135	9000000000	1016908189	0.1130
			10000000000	1129598504	0.1130

is within $4/z$ of Δ. We have not used this approach, but we record some intermediate values for the expression in (3.4) at $y = 10^j$ and $2 \cdot 10^j$ for $6 \le j \le 10$:

$10^6\colon 0.166923,\ 2{\cdot}10^6\colon 0.167484,\ 10^7\colon 0.168599,\ 2{\cdot}10^7\colon 0.169014,\ 10^8\colon 0.169857,$

$2{\cdot}10^8\colon 0.170176,\ 10^9\colon 0.170834,\ 2{\cdot}10^9\colon 0.171088,\ 10^{10}\colon 0.171617,\ 2{\cdot}10^{10}\colon 0.171822.$

3.3. **A φ analogue.** Let $s_\varphi(n) = n - \varphi(n)$. As with $s(n)$, partial results on Goldbach's conjecture imply that the range of $s_\varphi(\cdot)$ contains all odd numbers but for a set of asymptotic density 0. Little is known about even values. In [19] Erdős asks if a positive proportion of even numbers are of the form $s_\varphi(n)$ and if a positive proportion of even numbers are not of the form $s_\varphi(n)$. The proof in [30] does give a positive proportion of evens that are of the form $s_\varphi(n)$, but the even numbers not of this form remain an enigma. However, as with $s(n)$ the problem is amenable to calculation and heuristics. Let $U_\varphi(x)$ denote the number of integers in $[1, x]$ not of the form $s_\varphi(n)$. By an analogous algorithm from [39], we have extended the calculations of that paper 100-fold; see Table 4. The data file had 42.1 GB. (This file is larger than the one with nonaliquots since there are more even numbers n with $s_\varphi(n) \le 10^{10}$ than there are with $s(n) \le 10^{10}$.)

The conjectural density is

$$(3.8) \qquad \Delta_\varphi := \lim_{y \to \infty} \frac{1}{\log y} \sum_{a \le y}^* \frac{1}{a} e^{-a/s_\varphi(a)}.$$

At $y = 2 \cdot 10^{10}$ this expression rounds to 0.090595. It is difficult to say if the actual counts are supportive of the conjectural density Δ_φ; the evidence is somewhat less compelling than with the nonaliquots. At least the count densities seem to have peaked and are now slowly descending, perhaps to the limit Δ_φ. With both the nonaliquots and the φ analogues, the heuristic is very much an "at infinity" type of argument, while the counts are definitely very finite. It might be of some interest to try to do a statistical sample of random numbers that are much larger than our exhaustive counts, but it is not clear how one would decide if a given very large number n is of the required form.

Here are the intermediate calculations of the expression in (3.8) for various values of y:

10^6: 0.089193, $2 \cdot 10^6$: 0.089353, 10^7: 0.089673, $2 \cdot 10^7$: 0.089791, 10^8: 0.090033,

$2 \cdot 10^8$: 0.090124, 10^9: 0.090312, $2 \cdot 10^9$: 0.090385, 10^{10}: 0.090536, $2 \cdot 10^{10}$: 0.090595.

4. Primitive friendly pairs

We begin by collecting some results needed for the proof of Theorem 1.5. The following lemma says that we have a strong bound on the number of solutions to $\frac{\sigma(n)}{n} = \alpha$, even if we specify only the denominator of α. For the proof of this result, in a more precise form, see [32, Theorem 4.1].

Lemma 4.1. *The number of $n \le x$ for which $\sigma(n)/n$ can be written as a fraction with a given denominator d is at most $x^{o(1)}$, as $x \to \infty$, uniformly in $d \le x$.*

The next lemma is implicit in the proof of [23, Theorem 11]; for an explicit statement and proof, see [32, Lemma 4.2].

Lemma 4.2. *Let m be an integer in $[1, x]$. The number of integers $n \in [1, x]$ for which $\mathrm{rad}(n) \mid m$ is $x^{o(1)}$, as $x \to \infty$, uniformly in m.*

The next result, which was inspired by Erdős's method of proof in [16], appears as [35, Lemma 3.1].

Lemma 4.3. *Given a natural number m, the following algorithm outputs a unitary divisor a of m with $\gcd(a, \sigma(a)) = 1$. Moreover, at most $x^{o(1)}$ inputs $m \le x$ correspond to the same output a, as $x \to \infty$.*

Algorithm A:

Input: A natural number m
Output: A divisor a of m for which $a \parallel m$ and $\gcd(a, \sigma(a)) = 1$
Factor $m = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$, where $p_1 > p_2 > \cdots > p_k$.
$a \leftarrow 1$ // Initialize
for $i = 1$ **to** k **do**
$\quad$ **if** $\gcd(\sigma(p_i^{e_i} a), p_i^{e_i} a) = 1$ **then**
$\quad\quad$ $a \leftarrow p_i^{e_i} a$
$\quad$ **end**
end
return a

Proof of Theorem 1.5. Suppose that n and n' form a primitive friendly pair contained in $[1, x]$. Write $n = N_1 N_2$, where N_1 is squarefree, N_2 is squarefull, and $\gcd(N_1, N_2) = 1$. Define n_2 and n_3 by the equation $N_2 = n_2 n_3$, where n_2 is the output of Algorithm A when $m = N_2$. Note that n_2 and n_3 are coprime squarefull integers, since n_2 is a unitary divisor of N_2. Write

$$\frac{\sigma(n)}{n} = \frac{\sigma(n')}{n'} = \frac{A}{B}, \quad \text{where} \quad \gcd(A, B) = 1,$$

and let $U = \mathrm{rad}(\gcud(B, n_2))$. Define $N_1', N_2', n_2', n_3', U'$ analogously, but now with reference to n' instead of n. Since $\gcud(n_2, n_2') = 1$, it follows that $\gcd(U, U') = 1$. With $L = \lceil \log x \rceil$, choose $\theta_1, \theta_2, \theta_3, \delta \in \{\frac{1}{L}, \frac{2}{L}, \ldots, \frac{L-1}{L}, 1\}$ for which

$N_1 \in [\mathrm{e}^{-1} x^{\theta_1}, x^{\theta_1}], \quad n_2 \in [\mathrm{e}^{-1} x^{\theta_2}, x^{\theta_2}], \quad n_3 \in [\mathrm{e}^{-1} x^{\theta_3}, x^{\theta_3}], \text{ and } U \in [\mathrm{e}^{-1} x^\delta, x^\delta].$

Define $\theta_1', \theta_2', \theta_3', \delta'$ analogously. Finally, choose $\theta \in \{\frac{1}{L}, \frac{2}{L}, \ldots, \frac{L-1}{L}, 1\}$ with

$$B \in [\mathrm{e}^{-1} x^\theta, x^\theta].$$

So, running over all of the $L^9 = x^{o(1)}$ possibilities for the θ_i, the θ_i', δ, δ', and δ_0, we cover all possibilities for n, n'. Thus, it is enough to prove that the number of pairs n, n' corresponding to a given choice of these variables is bounded by $x^{1/2+o(1)}$, as $x \to \infty$. We will establish four upper bounds on the number of these pairs n, n'. The theorem will follow by combining these upper bounds.

Our first upper bound uses that

$$(4.1) \qquad B^2 \mid n_2 n_3 n_2' n_3'.$$

To prove this, write $B = B_1 B_2$, where B_1 is squarefree, B_2 is squarefull, and $\gcd(B_1, B_2) = 1$. Since B divides n and n', $B_2 \mid n_2 n_3$ and $B_2 \mid n_2' n_3'$, so that $B_2^2 \mid n_2 n_3 n_2' n_3'$. Now take a prime $p \mid B_1$. Since n and n' have no common unitary divisor, p cannot exactly divide both n and n'. So either $p^2 \mid n_2 n_3$ or $p^2 \mid n_2' n_3'$. Hence, $B_1^2 \mid n_2 n_3 n_2' n_3'$ and (4.1) holds. By Lemma 4.3, given n_2 there are only $x^{o(1)}$ possibilities for n_3. Similarly, n_2' determines n_3' in at most $x^{o(1)}$ ways. Thus, n_2 and n_2' determine the product $n_2 n_3 n_2' n_3'$ up to $x^{o(1)}$ possibilities. Now (4.1), along with the maximal order of the divisor function, shows that n_2 and n_2' determine B up to $x^{o(1)}$ possibilities. Invoking Lemma 4.1, n_2, n_2' determine the pair n, n' in at most $x^{o(1)}$ ways. Since the number of squarefull numbers in $[1, t]$ is $O(t^{1/2})$, the number of pairs is at most

$$(4.2) \qquad x^{\theta_2/2 + \theta_2'/2 + o(1)},$$

as $x \to \infty$.

For a second upper bound, expanding the equation $\frac{A}{B} = \frac{\sigma(N)}{N}$, we find that $B\sigma(N_1)\sigma(n_2)\sigma(n_3) = AN_1 n_2 n_3$. Since $\gcd(n_2, \sigma(n_2)) = 1$,

$$n_2 \mid B\sigma(N_1)\sigma(n_3), \quad \text{so that} \quad \mathrm{rad}(n_2) \mid U \cdot \sigma(N_1)\sigma(n_3).$$

By Lemma 4.2, the number of possibilities for $n = N_1 n_2 n_3$ is at most

$$(4.3) \qquad x^{\delta + \theta_1 + \theta_3/2 + o(1)}.$$

By Wirsing's theorem (or Lemma 4.1), this is also a bound on the number of possibilities for the pair n, n'. A symmetric argument gives our third upper bound,

$$(4.4) \qquad x^{\delta' + \theta_1' + \theta_3'/2 + o(1)}.$$

To prove the fourth upper bound, note that since $\gcd(U, U') = 1$, we have $U^2 U'^2 \mid B$. Consequently, the number of possibilities for B is

$$\ll \sum_{\substack{u_1 \geq \mathrm{e}^{-1} x^\delta \\ u_2 \geq \mathrm{e}^{-1} x^{\delta'}}} \frac{x^\theta}{u_1^2 u_2^2} \ll x^{\theta - \delta - \delta'}.$$

By Lemma 4.1, the number of possibilities for the pair n, n' is at most

$$(4.5) \qquad x^{\theta - \delta - \delta' + o(1)}.$$

Taking the geometric mean of the four upper bounds (4.2), (4.3), (4.4), and (4.5), the number of possibilities for the pair n, n' does not exceed

$$(4.6) \qquad x^{\frac{1}{4}(\theta_1 + \theta_1' + \frac{1}{2}(\theta_2 + \theta_2' + \theta_3 + \theta_3') + \theta) + o(1)}.$$

Since $n = N_1 n_2 n_3 \le x$, we have $\theta_1 \le 1 - \theta_2 - \theta_3 + o(1)$; similarly, $\theta_1' \le 1 - \theta_2' - \theta_3' + o(1)$. From (4.1), $2\theta \le \theta_2 + \theta_3 + \theta_2' + \theta_3' + o(1)$. Substituting these bounds for θ_1, θ_1' and θ into (4.6) gives a final upper bound of $x^{\frac{1}{2}+o(1)}$, as $x \to \infty$. $\qquad\square$

4.1. **Computations.** To compile Table 2, for each $n \le 10^L$, we found all solutions $m \le 10^L$ with $\frac{\sigma(n)}{n} = \frac{\sigma(m)}{m}$, recording only those solutions m with $\gcud(m,n) = 1$. To find the values of m, we used the `gp` script `solveBA` of Michel Marcus (see [42]), which was based on an earlier program by Jan Munch Pedersen. The script `solveBA` uses a recursive algorithm to find all solutions n up to a given limit satisfying an equation of the form $\frac{\sigma(n)}{n} = \frac{A}{B}$. The values of $\gcud(m,n)$ were computed using Algorithm B3 below.

Algorithm B1:

Input: A natural number n and a divisor d of n
Output: The smallest $D \parallel n$ with $d \mid D$
$D \leftarrow d$.
Iterate $D \mapsto D \cdot (D, n/D)$ until D stabilizes.
return D

Algorithm B2:

Input: A natural number n and a divisor d of n
Output: The largest $D \parallel n$ with $D \mid d$
$D \leftarrow d$.
Iterate $D \mapsto D/(D, n/D)$ until D stabilizes.
return D

Algorithm B3:

Input: Natural numbers m and n
Output: $\gcud(m,n)$
$d \leftarrow \gcd(m,n)$
$d_1 \leftarrow$ output of B1 with input d, m
$d_2 \leftarrow$ output of B1 with input d, n
$d_3 \leftarrow d_1 d_2 / d$
return the output of B2 with the pair d, d_3

---

5. The distribution function for $F(n;x)$

Define a *friendly k-set* as a set of k distinct integers all of which share the same value of $\frac{\sigma(n)}{n}$, and call such a set *primitive* if there is no $d > 1$ that is a unitary divisor of each element. The proof of Theorem 1.6 will use the following crude upper bound on the count of primitive friendly k-sets.

Lemma 5.1. *Fix $k \ge 2$. The number of primitive friendly k-sets contained in $[1, x]$ is at most $x^{1-\frac{1}{k+1}+o(1)}$, as $x \to \infty$.*

When $k = 2$, Lemma 5.1 gives an upper bound of $x^{2/3+o(1)}$, which is inferior to the result of Theorem 1.5. However, the proof of Lemma 5.1 is substantially simpler.

Proof. List the elements of the set as $n_1, \ldots, n_k$. Write the common value of $\frac{\sigma(n_i)}{n_i}$ as $\frac{A}{B}$ in lowest terms. If $\mathrm{rad}(B) \leq x^{1-\frac{1}{k+1}}$, then there are at most $x^{1-\frac{1}{k+1}+o(1)}$ possible values of B, by Lemma 4.2, and so also at most $x^{1-\frac{1}{k+1}+o(1)}$ possibilities for the k-set, by Lemma 4.1. So we suppose that $\mathrm{rad}(B) > x^{1-\frac{1}{k+1}}$.

Let p be a prime dividing B. Then p divides each n_i. Since $n_1, \ldots, n_k$ is primitive friendly, it is impossible for p to exactly divide each n_i. Thus, if $N = n_1 \cdots n_k$, then $\mathrm{rad}(B)^{k+1} \mid N$. Hence, the number of possibilities for the integer $N \leq x^k$ is at most

$$\sum_{r > x^{1-\frac{1}{k+1}}} \frac{x^k}{r^{k+1}} \ll x^{1-\frac{1}{k+1}}.$$

From the maximal order of the k-fold divisor function, N determines the set $\{n_1, \ldots, n_k\}$ in at most $x^{o(1)}$ ways. The result follows. $\qquad\square$

Proof of Theorem 1.6. We may assume $k \geq 1$, since when $k = 0$ the asymptotic claimed in the theorem holds with $\alpha_0 = 1$. We begin by placing each primitive $(k+1)$-set in increasing order, $n_0 < n_1 < \cdots < n_k$. We then list the primitive $(k+1)$-sets in increasing order of n_k, breaking ties arbitrarily. Say the jth tuple is labeled as $n_{0,j}, \ldots, n_{k,j}$. Put

$$\mathscr{M}_{i,j}(x) := \{dn_{i,j} : 1 \leq d \leq x/n_{k,j} \text{ and } \gcd(d, \prod_{\ell=0}^{k} n_{\ell,j}) = 1\},$$

and observe that a natural number $m \leq x$ has at least k friends in $[1, x]$ if and only if m belongs to one of the sets $\mathscr{M}_{i,j}(x)$. For each $J \geq 1$, put

$$\alpha_k^{(J)} = \lim_{x \to \infty} \frac{1}{x} \cdot \# \bigcup_{\substack{1 \leq j \leq J \\ 0 \leq i \leq k}} \mathscr{M}_{i,j}(x).$$

The existence of the limit follows from inclusion–exclusion. (Each finite intersection of the sets $\mathscr{M}_{i,j}(x)$ can be described as the set of numbers not exceeding a certain constant multiple of x that satisfy certain congruence conditions.) The $\alpha_k^{(J)}$ are weakly increasing in J and bounded by 1; thus, it makes sense to set

$$\alpha_k = \lim_{J \to \infty} \alpha_k^{(J)}.$$

Clearly, $\liminf_{x \to \infty} N_k(x)/x \geq \alpha_k$. We now show that $\limsup_{x \to \infty} N_k(x)/x \leq \alpha_k$.

If $m \leq x$ has at least k friends in $[1, x]$, but $m \notin \bigcup_{\substack{1 \leq j \leq J \\ 0 \leq i \leq k}} \mathscr{M}_{i,j}(x)$, then $m \in \mathscr{M}_{i,j}(x)$ for some $j > J$. Since $\#\mathscr{M}_{i,j}(x) \leq \frac{x}{n_{k,j}}$, the number of these $m \leq x$ is at most

$$(k+1)x \cdot \sum_{j > J} \frac{1}{n_{k,j}}.$$

We now study the sum on j. Without the restriction to $j > J$, the sum converges (or is a finite sum). Indeed, from Lemma 5.1, the contribution to the unrestricted sum from those j with $X \leq n_{k,j} < 2X$ is bounded by $X^{-\frac{1}{k+2}+o(1)}$, as $X \to \infty$; the desired result now follows upon summing X along the powers of 2. So with the restriction to $j > J$, the series decays to 0 as $J \to \infty$. Letting $x \to \infty$ and then $J \to \infty$, in that order, we get that $\limsup_{x \to \infty} N_k(x)/x \leq \alpha_k$.

It remains to prove that $\alpha_k \to 0$. Since $k \cdot N_k(x) \leq \sum_{n \leq x} F(n; x)$, Erdős's estimate (1.2) shows that $\alpha_k \ll 1/k$. □

Remark.

    (i) A similar argument to the one just seen will show that each positive integral moment $\frac{1}{x} \sum_{n \leq x} F(n; x)^k$ tends to a finite limit as $x \to \infty$. Here one uses that $F(n; x)^k$ can be viewed as the count of (ordered) friendly $(k+1)$-tuples in $[1, x]$ having final element n and all other elements distinct from n. As a consequence, the α_k tend to zero more rapidly than any power of k^{-1}.

    (ii) The computations of $N_i(x)$, $i = 1, \ldots, 4$, reported in Table 3 were made with the aforementioned `solveBA` script.

We do not know how to show that each α_k is nonzero. This is equivalent to the assertion that the count of natural numbers n with $\frac{\sigma(n)}{n} = \alpha$ is unbounded in α, a conjecture mentioned already in [19]. Note that this conjecture would follow from the infinitude of perfect numbers. Since the α_k may eventually vanish, we cannot show that the α_k are strictly decreasing. However, we can prove that the sequence is strictly decreasing until it hits zero.

Theorem 5.2. *If $\alpha_k > 0$, then $\alpha_k > \alpha_{k+1}$.*

Proof. Since $\alpha_k > 0$, there are examples of friendly $(k+1)$-sets. (When $k = 0$, we consider every singleton subset of the natural numbers to be a friendly $(k+1)$-set.) Fix such a set $\{n_0 < n_1 < \cdots < n_k\}$ once and for all, chosen to have n_k as small as possible. This set is automatically primitive. Every integer of the form $m = dn_k$, with $d \leq x/n_k$ and $\gcd(d, n_0 \cdots n_k) = 1$, has at least k friends in $[1, x]$. It suffices to show that $\gg x$ values of d give rise to an m with *exactly* k friends in $[1, x]$.

Sort all of the primitive friendly $(k+2)$-sets in increasing order, and then list the sets in order of their largest entry. Say the jth set consists of $n_{0,j} < n_{1,j} < \cdots < n_{k+1,j}$. With J a large, fixed integer to be chosen later, we restrict attention to d satisfying

$$(5.1) \qquad \frac{x}{n_k + 1} < d \leq \frac{x}{n_k} \quad \text{and} \quad d \equiv 1 \pmod{\prod_{i=0}^{k} n_i \prod_{\substack{0 \leq i \leq k+1 \\ 1 \leq j \leq J}} n_{i,j}}.$$

Let M_J be the modulus of this congruence. For large x, there are $\gg x/M_J$ values of d satisfying (5.1). (The implied constant also depends on k and the set $\{n_0, \ldots, n_k\}$, but since these parameters are fixed, we ignore this in our notation.) We now show that for most of these d, the integer $M = dn_k$ has exactly k friends in $[1, x]$.

If $m = dn_k$ has at least $k + 1$ friends in $[1, x]$, then for some j,

$$dn_k = d'n_{i,j}, \quad \text{where} \quad d' \leq x/n_{k+1,j} \text{ and } \gcd\left(d', \prod_{\ell=0}^{k+1} n_{\ell,j}\right) = 1.$$

In particular, $n_{i,j} \mid dn_k$. Suppose to begin with that $j \leq J$. Then $\gcd(n_{i,j}, d) = 1$ from (5.1), and so $n_{i,j} \mid n_k$. Putting $e = n_k/n_{i,j}$,

$$d' = de \leq \frac{x}{n_{k+1,j}}, \quad \text{so that} \quad d \leq \frac{x}{e \cdot n_{k+1,j}} \leq \frac{x}{n_{k+1,j}}.$$

Since $n_{0,j}, \ldots, n_{k,j}$ are in increasing order and make up a friendly $(k + 1)$-set, we have $n_{k+1,j} > n_{k,j} \geq n_k$, by the initial choice of n_k. Hence, $n_{k+1,j} \geq n_k + 1$, and

so this last upper bound on d contradicts (5.1). So we must have $j > J$. From $dn_k = d'n_{i,j}$ and $\gcd(d, M_J) = 1$, we see that $\gcd(n_{i,j}, M_J) \mid n_k$. In particular, $\gcd(n_{i,j}, M_J) \ll 1$. Using $dn_k = d'n_{i,j}$ and (5.1), we see that d' belongs to a fixed residue class modulo $n_K M_J / \gcd(n_{i,j}, n_K M_J)$. The number of these $d' \le x/n_{k+1,j}$ is at most

$$x\frac{\gcd(n_{i,j}, n_k M_J)}{n_k M_J \cdot n_{k+1,j}} + 1 \ll \frac{x}{M_J \cdot n_{k+1,j}} + 1.$$

Hence, the number of $m = dn_k$ with d satisfying (5.1) that have at least $k + 1$ friends in $[1, x]$ is

$$\ll \sum_{i=0}^{k+1} \sum_{\substack{j>J \\ n_{k+1,j} \le x}} \left(\frac{x}{M_J \cdot n_{k+1,j}} + 1 \right).$$

Since the sum on j is only over $j > J$, and since we can take J arbitrarily large, the main term here can be made smaller than any constant multiple of x/M_J, by the same argument used in the proof of Theorem 1.6. The error term is bounded by a power of x less than 1, by Lemma 5.1. So fixing J large enough, we have that for large x, there are $\gg x/M_J \gg x$ values of d satisfying (5.1) for which $m = dn_k$ has exactly k friends in $[1, x]$, as desired. $\qquad\square$

6. Some final remarks

In the paper [17], which inspired our work above on friendly pairs, Erdős writes "I would like to call attention to three simple problems which as far as I know are still unsolved". The problems are:

- Are there infinitely many solutions to the equation $\sigma(n) = \varphi(m)$?
- For each real number $c \ge 1$, is there an infinite sequence of pairs m_k, n_k with $\sigma(m_k) = \sigma(n_k)$ and $m_k/n_k \to c$ as $k \to \infty$?
- If $g(x)$ is the number of coprime pairs a, b with $a < b \le x$ and $\sigma(a) = \sigma(b)$, do we have $g(x)/x \to \infty$ as $x \to \infty$?

We note that the first two problems have been recently solved in the affirmative (see [24], [34]), and the third was "almost" solved in the affirmative (solved without the coprimality condition) in 1980 (see [37]). It seems fitting to close this paper with a complete proof of the third problem.

Following an idea of Erdős [14], say $\theta > 1$ is such that there are $y^\theta/(\log y)^{O(1)}$ primes $p \le y^\theta$ with $P(p+1) \le y$. Let x be large, let $y = \log x$, let $\mathscr{S}$ be the set of primes $p \le y^\theta$ with $P(p+1) \le y$, let $n = \#\mathscr{S}$, and let $k = \lfloor \log x/\log(y^\theta) \rfloor$. Consider integers a which are the product of k distinct primes in $\mathscr{S}$. Since $\binom{n}{k} = x^{1-1/\theta+o(1)}$, as $x \to \infty$, there are $x^{1-1/\theta+o(1)}$ of these integers $a \le x$, each with $P(\sigma(a)) \le y$. But, as was shown by Erdős (see [11]), the number of integers to x not divisible by any prime $> y$ is $x^{o(1)}$. Hence, there is one value v such that $\sigma(a) = v$ has $x^{1-1/\theta+o(1)}$ solutions $a \le x$, as $x \to \infty$. It follows that there are $x^{2-2/\theta+o(1)}$ pairs $a < b \le x$ with $\sigma(a) = \sigma(b)$.

This would solve the third problem above if we had $\theta > 2$ and if the pairs a, b created were coprime. We know from [37] that $\theta > 2$, the current record being a

number slightly larger than 3.377; see [4]. Let $\epsilon > 0$ be fixed but arbitrarily small, and let $k_0 = \lfloor (1-\epsilon)k \rfloor$. The number of pairs a, b which reduce to a coprime pair a', b' with $\omega(a') = \omega(b') = j$ is at most

$$\binom{n}{j}^2 \binom{n}{k-j}.$$

Summing this for $j \le k_0$ gives $x^{2-2/\theta-(1-1/\theta)\epsilon+o(1)}$. Thus, almost all of our pairs a, b reduce to a pair a', b' where $\omega(a') = \omega(b') > k_0$. In fact, there is some $j_0 \in (k_0, k]$ where $x^{2-2/\theta+o(1)}$ pairs a, b reduce to coprime pairs a', b' with $\omega(a') = \omega(b') = j_0$. Each of these pairs comes from at most $\binom{n}{k-j_0} \le x^{(1-1/\theta)\epsilon+o(1)}$ pairs a, b, so there must be at least $x^{(2-\epsilon)(1-1/\theta)+o(1)}$ coprime pairs $a', b' \le x$ with $\sigma(a') = \sigma(b')$. If ϵ is sufficiently small, we see that $g(x) > x^{1.4}$ for all sufficiently large values of x.

References

[1] J. D. Alanen, *Empirical study of aliquot sequences*, Ph.D. thesis, Yale University, 1972.

[2] Aria Anavi, Paul Pollack, and Carl Pomerance, *On congruences of the form $\sigma(n) \equiv a$ (mod n)*, Int. J. Number Theory **9** (2013), no. 1, 115–124, DOI 10.1142/S1793042112501266. MR2997493

[3] Michael R. Avidon, *On the distribution of primitive abundant numbers*, Acta Arith. **77** (1996), no. 2, 195–205. MR1411032

[4] R. C. Baker and G. Harman, *Shifted primes without large prime factors*, Acta Arith. **83** (1998), no. 4, 331–361. MR1610553

[5] William D. Banks and Florian Luca, *Nonaliquots and Robbins numbers*, Colloq. Math. **103** (2005), no. 1, 27–32, DOI 10.4064/cm103-1-4. MR2148946

[6] W. Bosma, *Aliquot sequences with small starting values*, preprint available from `http://www.math.ru.nl/~bosma/`.

[7] Wieb Bosma and Ben Kane, *The aliquot constant*, Q. J. Math. **63** (2012), no. 2, 309–323, DOI 10.1093/qmath/haq050. MR2925292

[8] E. Catalan, *Propositions et questions diverses* (French), Bull. Soc. Math. France **16** (1888), 128–129. MR1504036

[9] Yong-Gao Chen and Qing-Qing Zhao, *Nonaliquot numbers*, Publ. Math. Debrecen **78** (2011), no. 2, 439–442, DOI 10.5486/PMD.2011.4820. MR2796778

[10] N. G. de Bruijn, *On the number of positive integers $\le x$ and free of prime factors $> y$*, Nederl. Acad. Wetensch. Proc. Ser. A. **54** (1951), 50–60. MR0046375

[11] N. G. de Bruijn, *On the number of positive integers $\le x$ and free prime factors $> y$. II*, Nederl. Akad. Wetensch. Proc. Ser. A 69=Indag. Math. **28** (1966), 239–247. MR0205945

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

[13] P. Erdős, *On primitive abundant numbers*, J. London Math. Soc. **10** (1935), 49–58.

[14] P. Erdős, *On the normal number of prime factors of $p-1$ and some related problems concerning Euler's φ-function*, Q. J. Math., Oxford Ser. **6** (1935), 205–213.

[15] P. Erdős, *On amicable numbers*, Publ. Math. Debrecen **4** (1955), 108–111. MR0069198

[16] P. Erdös, *On perfect and multiply perfect numbers*, Ann. Mat. Pura Appl. (4) **42** (1956), 253–258. MR0082516

[17] P. Erdős, *Remarks on number theory. II. Some problems on the σ function*, Acta Arith. **5** (1959), 171–177. MR0107623

[18] P. Erdős, *Über die Zahlen der Form $\sigma(n) - n$ und $n - \varphi(n)$*, Elem. Math. **28** (1973), 83–86.

[19] P. Erdős, *On the distribution of numbers of the form $\sigma(n)/n$ and on some related questions*, Pacific J. Math. **52** (1974), 59–65. MR0354601

[20] P. Erdős, *On asymptotic properties of aliquot sequences*, Math. Comp. **30** (1976), no. 135, 641–645. MR0404115

[21] P. Erdős and G. J. Rieger, *Ein Nachtrag über befreundete Zahlen* (German), J. Reine Angew. Math. **273** (1975), 220. MR0364157

[22] P. Erdős, A. Granville, C. Pomerance, and C. Spiro, *On the normal behavior of the iterates of some arithmetic functions*, Analytic number theory (Allerton Park, IL, 1989), Progr. Math., vol. 85, Birkhäuser Boston, Boston, MA, 1990, pp. 165–204, DOI 10.1007/978-1-4612-3464-7_13. MR1084181

[23] Paul Erdős, Florian Luca, and Carl Pomerance, *On the proportion of numbers coprime to a given integer*, Anatomy of integers, CRM Proc. Lecture Notes, vol. 46, Amer. Math. Soc., Providence, RI, 2008, pp. 47–64. MR2437964

[24] Kevin Ford, Florian Luca, and Carl Pomerance, *Common values of the arithmetic functions ϕ and σ*, Bull. Lond. Math. Soc. **42** (2010), no. 3, 478–488, DOI 10.1112/blms/bdq014. MR2651943

[25] Richard K. Guy and J. L. Selfridge, *What drives an aliquot sequence?*, Math. Comput. **29** (1975), 101–107. Collection of articles dedicated to Derrick Henry Lehmer on the occasion of his seventieth birthday. MR0384669

[26] H. Halberstam and H.-E. Richert, *Sieve methods*, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], London-New York, 1974. London Mathematical Society Monographs, No. 4. MR0424730

[27] Bernhard Hornfeck and Eduard Wirsing, *Über die Häufigkeit vollkommener Zahlen* (German), Math. Ann. **133** (1957), 431–438. MR0090600

[28] Mitsuo Kobayashi, Paul Pollack, and Carl Pomerance, *On the distribution of sociable numbers*, J. Number Theory **129** (2009), no. 8, 1990–2009, DOI 10.1016/j.jnt.2008.10.011. MR2522719

[29] H. W. Lenstra, Jr., *Problem 6064*, Amer. Math. Monthly **82** (1975), 1016. *Solution by the proposer*, op. cit. **84** (1977), 580.

[30] Florian Luca and Carl Pomerance, *The range of the sum-of-proper-divisors function*, Acta Arith. **168** (2015), no. 2, 187–199, DOI 10.4064/aa168-2-6. MR3339454

[31] Tomás Oliveira e Silva, Siegfried Herzog, and Silvio Pardi, *Empirical verification of the even Goldbach conjecture and computation of prime gaps up to $4 \cdot 10^{18}$*, Math. Comp. **83** (2014), no. 288, 2033–2060, DOI 10.1090/S0025-5718-2013-02787-1. MR3194140

[32] Paul Pollack, *On the greatest common divisor of a number and its sum of divisors*, Michigan Math. J. **60** (2011), no. 1, 199–214, DOI 10.1307/mmj/1301586311. MR2785871

[33] Paul Pollack, *Some arithmetic properties of the sum of proper divisors and the sum of prime divisors*, Illinois J. Math. **58** (2014), no. 1, 125–147. MR3331844

[34] Paul Pollack, *Remarks on fibers of the sum-of-divisors function*, in Analytic number theory (in honor of Helmut Maier's 60th birthday), M. Rassias and C. Pomerance, eds., Springer, Cham, Switzerland, 2015, pp. 305–320.

[35] Paul Pollack and Carl Pomerance, *Prime-perfect numbers*, Integers **12** (2012), no. 6, 1417–1437, DOI 10.1515/integers-2012-0044. MR3011565

[36] Carl Pomerance, *On the distribution of amicable numbers*, J. Reine Angew. Math. **293/294** (1977), 217–222. MR0447087

[37] Carl Pomerance, *Popular values of Euler's function*, Mathematika **27** (1980), no. 1, 84–89, DOI 10.1112/S0025579300009967. MR581999

[38] Carl Pomerance, *On amicable numbers*, in Analytic number theory (in honor of Helmut Maier's 60th birthday), M. Rassias and C. Pomerance, eds., Springer, Cham, Switzerland, 2015, pp. 321–327.

[39] Carl Pomerance and Hee-Sung Yang, *Variant of a theorem of Erdős on the sum-of-proper-divisors function*, Math. Comp. **83** (2014), no. 288, 1903–1913, DOI 10.1090/S0025-5718-2013-02775-5. MR3194134

[40] G. J. Rieger, *Bemerkung zu einem Ergebnis von Erdős über befreundete Zahlen* (German), J. Reine Angew. Math. **261** (1973), 157–163. MR0319882

[41] J. Sesiano, *Two problems of number theory in Islamic times*, Arch. Hist. Exact Sci. **41** (1991), no. 3, 235–238, DOI 10.1007/BF00348408. MR1107382

[42] N. J. A. Sloane, *On-line encyclopedia of integer sequences*, http://oeis.org. Sequence A001970; gp script by Michael Marcus online at http://oeis.org/A246827/a246827.gp.txt.

[43] Gérald Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambridge Studies in Advanced Mathematics, vol. 46, Cambridge University Press, Cambridge, 1995. Translated from the second French edition (1995) by C. B. Thomas. MR1342300

[44] H. J. J. te Riele, *A theoretical and computational study of generalized aliquot sequences*, Mathematisch Centrum, Amsterdam, 1976. Mathematical Centre Tracts, No. 74. MR0556033

[45] Vincent Toulmonde, *Module de continuité de la fonction de répartition de $\phi(n)/n$* (French), Acta Arith. **121** (2006), no. 4, 367–402, DOI 10.4064/aa121-4-6. MR2224402

[46] R. C. Vaughan, *The Hardy-Littlewood method*, 2nd ed., Cambridge Tracts in Mathematics, vol. 125, Cambridge University Press, Cambridge, 1997. MR1435742

[47] Eduard Wirsing, *Bemerkung zu der Arbeit über vollkommene Zahlen* (German), Math. Ann. **137** (1959), 316–318. MR0104636

Department of Mathematics, University of Georgia, Athens, Georgia 30602
E-mail address: pollack@uga.edu

Department of Mathematics, Dartmouth College, Hanover, New Hampshire 03755
E-mail address: carl.pomerance@dartmouth.edu