Popular values and popular subsets of Euler's φ-function

Paul Pollack

Max Planck Institute Seminar

November 2019

MPI Mathematik

We let $\varphi(n)$ denote Euler's totient function. That is, $\varphi(n)$ is the number of integers in $[1, n]$ that are relatively prime to n. Equivalently,

$$\varphi(n) = \#(\mathbb{Z}/n\mathbb{Z})^{\times}.$$

The chief object of study in this talk is the arithmetic function

$$N(m) = \#\{n : \varphi(n) = m\}.$$

We let $\varphi(n)$ denote Euler's totient function. That is, $\varphi(n)$ is the number of integers in $[1, n]$ that are relatively prime to n. Equivalently,

$$\varphi(n) = \#(\mathbb{Z}/n\mathbb{Z})^\times.$$

The chief object of study in this talk is the arithmetic function

$$N(m) = \#\{n : \varphi(n) = m\}.$$

Here are its first several values:

m	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$N(m)$	2	3	0	4	0	4	0	5	0	2	0	6	0	0

As an example, $N(12) = 6$, corresponding to $\varphi^{-1}(12) = \{13, 21, 26, 28, 36, 42\}$.

How large might we expect $N(m)$ to be?

To get a feel for this, one might look at the first moment (or average) of $N(m)$. Note that

$$\sum_{m \leq x} N(m) = \#\{n : \varphi(n) \leq x\}.$$

Clearly, $\varphi(n)$ is never larger than n. It is also not that much smaller: For all $n \geq 3$, we have $\varphi(n) \gg n/\log\log n$. Hence, any n with $\varphi(n) \leq x$ satisfies $n \ll x \log\log x$, and so

$$\sum_{m \leq x} N(m) \ll x \log\log x.$$

The above argument is crude.

While $\varphi(n)/n$ is occasionally as small as $O(1/\log\log n)$, such n are quite rare. Quantifying this, Erdős and Turán showed that there is a constant $C > 0$ with

$$\sum_{m \le x} N(m) \sim Cx, \quad x \to \infty.$$

It was later noticed that C could be given in closed form: $C = \zeta(2)\zeta(3)/\zeta(6)$.

Proof sketch.
Apply the Wiener–Ikehara theorem to

$$\sum_{m=1}^{\infty} N(m)/m^s = \sum_{n=1}^{\infty} 1/\varphi(n)^s = \zeta(s) \prod_p (1 + (p-1)^{-s} - p^{-s}). \quad \blacksquare$$

So $N(m)$ behaves like a constant, on average.

So $N(m)$ behaves like a constant, on average.

Question

For how many $m \leq x$ is $N(m) > 0$? In other words, how does the count of φ-values up to x grow, as a function of x?

So $N(m)$ behaves like a constant, on average.

Question

For how many $m \leq x$ is $N(m) > 0$? In other words, how does the count of φ-values up to x grow, as a function of x?

Question

How large can $N(m)$ be as a function of m? In other words, what is the maximal order of $N(m)$.

Starting from about 1930, the 1st question was the subject of several papers, by Pillai, Erdős, Erdős–Hall, Pomerance, Maier–Pomerance, and most recently Ford (1998).

Ford obtains the correct order of the counting function. It behaves roughly like $\frac{x}{\log x} \exp(C(\log\log\log x)^2)$, where $C \approx 0.82$.

The first and second questions are, of course, related.

Since φ maps $[1, x]$ into a subset of $[1, x]$ of size $x/(\log x)^{1+o(1)}$, one sees immediately that

$$\max_{m \le x} N(m) \ge (\log x)^{1+o(1)}.$$

The first and second questions are, of course, related.

Since φ maps $[1, x]$ into a subset of $[1, x]$ of size $x/(\log x)^{1+o(1)}$, one sees immediately that

$$\max_{m \le x} N(m) \ge (\log x)^{1+o(1)}.$$

Erdős saw in 1935 how to do better. His idea was to construct a large $\mathcal{N} \subset [1, x]$ such that $\varphi(\mathcal{N})$ is entirely contained in the set of numbers that are $(\log x)$-smooth, meaning having no prime factors $> \log x$. Erdős knew that there were only $x^{o(1)}$ numbers up to x that are $(\log x)$-smooth. Hence, by the Pigeonhole principle,

$$\max_{m \le x} N(m) \ge \frac{\#\mathcal{N}}{\#\{(\log x)\text{-smooths } [1, x]\}} \ge x^{o(1)} \#\mathcal{N}.$$

If $\varphi(n)$ is $(\log x)$-smooth, one needs $p-1$ to be $(\log x)$-smooth for all $p \mid n$. So to construct these n, one needs to know that there are primes p with $p-1$ having only small prime factors.

Theorem (Baker–Harman 1998)

For large T, there are "many" primes $p \leq T$ with $p-1$ having all prime factors at most $T^{0.2961}$.

Erdős in 1935 had a much weaker version of the theorem, with 0.2961 replaced by an exponent slightly smaller than 1.

Using B–H, Erdős's construction allows $\#\mathcal{N} \geq x^{0.7039-o(1)}$, each member of $\varphi(\mathcal{N})$ being $(\log x)$-smooth; hence,

$$\max_{m \leq x} N(m) \geq x^{0.7038}.$$

Erdős had this result with 0.7038 replaced by some positive constant.

Probably the exponent 0.2961 can be replaced with any exponent > 0. This would show that for any $\epsilon > 0$ and all large x,

$$\max_{m \le x} N(m) \ge x^{1-\epsilon}.$$

This is already beyond reach, but we can be bold and see what happens if we assume something like the Baker–Harman theorem for still smaller values of T, of the size $T^{o(1)}$. The "right" conjecture can be guessed based on assuming that smooth $p-1$'s are distributed similarly to smooth n's of the same size. This suggests:

Conjecture

Let $L(x) = \exp(\log x \cdot \frac{\log\log\log x}{\log\log x})$. *Then*

$$\max_{m \le x} N(m) \ge x/L(x)^{1+o(1)}, \quad (\text{as } x \to \infty).$$

This conjecture was first proposed by Pomerance (1980).

In the same paper, Pomerance proved (unconditionally) that

$$\max_{m \le x} N(m) \le x/L(x)^{1+o(1)} \quad (\text{as } x \to \infty).$$

Thus, subject to plausible conjectures on the distribution of smooth shifted primes, we understand the upper order of $N(m)$.

From popular values to popular subsets

We are concerned with the following question: Suppose $\mathcal{S}$ is a subset of $[1, x]$ with $\#\mathcal{S} \approx x^{\alpha}$? What can one say about $\#\varphi^{-1}(S)$? In other words, how large is

$$\sum_{m \in \mathcal{S}} N(m) \quad ?$$

Of course, one can put in Pomerance's pointwise upper bound for $N(m)$, but the result is worse than trivial if $\alpha > 0$. Remember that the sum is $O(x)$, while our worst case upper bound for a single term is $x/L(x)^{1+o(1)}$.

Theorem (P., 2018)

Fix $\alpha \in (0,1)$. Then as $x \to \infty$,

$$\#\varphi^{-1}(\mathcal{S}) \le x/L(x)^{1-\alpha+o(1)},$$

uniformly in the choice of subsets $\mathcal{S} \subset [1,x]$ with $\#\mathcal{S} \le x^{\alpha}$.

Choosing $\alpha \approx 0$ recovers Pomerance's pointwise bound on $N(m)$. In fact, we get the same upper bound of $x/L(x)^{1+o(1)}$ whenever $\#\mathcal{S} \le x^{o(1)}$.

Theorem (P., 2018)

Fix $\alpha \in (0,1)$. Then as $x \to \infty$,

$$\#\varphi^{-1}(\mathcal{S}) \leq x/L(x)^{1-\alpha+o(1)},$$

uniformly in the choice of subsets $\mathcal{S} \subset [1,x]$ with $\#\mathcal{S} \leq x^{\alpha}$.

Choosing $\alpha \approx 0$ recovers Pomerance's pointwise bound on $N(m)$. In fact, we get the same upper bound of $x/L(x)^{1+o(1)}$ whenever $\#\mathcal{S} \leq x^{o(1)}$.

The result is probably best possible for every α. Let $\mathcal{S}$ be the set of $(\log x)^{1/(1-\alpha)}$-smooths in $[1,x]$. Then $\#\mathcal{S} = x^{\alpha+o(1)}$. Banks, Friedlander, Pomerance, Shparlinski have shown — conditional on the same conjectures alluded to before — that $\#\varphi^{-1}(\mathcal{S}) = x/L(x)^{1-\alpha+o(1)}$.

A problem of Davenport–Heilbronnn

According to Erdős (1958), Davenport and Heilbronn corresponded about the second moment of $N(m)$, i.e., the behavior of

$$\frac{1}{x}\sum_{m\le x} N(m)^2. \tag{*}$$

Note that the sum counts the number of pairs n, n' with $\varphi(n) = \varphi(n') \le x$.

Heilbronn proved that (*) tends to infinity as $x \to \infty$.

A problem of Davenport–Heilbronnn

According to Erdős (1958), Davenport and Heilbronn corresponded about the second moment of $N(m)$, i.e., the behavior of

$$\frac{1}{x}\sum_{m\le x} N(m)^2. \qquad (*)$$

Note that the sum counts the number of pairs n, n' with $\varphi(n) = \varphi(n') \le x$.

Heilbronn proved that (*) tends to infinity as $x \to \infty$.

Taking a single term is enough to show unconditionally that

$$(*) > (x^{0.7038})^2/x > x^{0.4}$$

for large x and, conjecturally, that

$$(*) > x/L(x)^{2+o(1)}, \quad \text{as } x \to \infty.$$

There's also a fairly easy upper bound. Let $A = \max_{m \le x} N(m)$. Then

$$\sum_{m \le x} N(m)^2 \le A \sum_{m \le x} N(m) \ll Ax.$$

Since $A \le x/L(x)^{1+o(1)}$, we get that

$$(*) \le x/L(x)^{1+o(1)}, \quad \text{as } x \to \infty.$$

Question: What is the correct exponent on $L(x)$? Is it 1, 2 or something inbetween?

There's also a fairly easy upper bound. Let $A = \max_{m \leq x} N(m)$. Then

$$\sum_{m \leq x} N(m)^2 \leq A \sum_{m \leq x} N(m) \ll Ax.$$

Since $A \leq x/L(x)^{1+o(1)}$, we get that

$$(*) \leq x/L(x)^{1+o(1)}, \quad \text{as } x \to \infty.$$

Question: What is the correct exponent on $L(x)$? Is it 1, 2 or something inbetween?

Theorem (P., 2018)

As $x \to \infty$,

$$(*) \leq x/L(x)^{2+o(1)}.$$

The number of solutions to $\varphi(n') = \varphi(n)$, with n given

Let $C(n) = N(\varphi(n))$. In other words, $C(n)$ is the number of n' with $\varphi(n') = \varphi(n)$. Clearly, $C(n) \geq 1$ for all n.

Conjecture (Carmichael, 1907)

$C(n) > 1$ *for all* n.

Carmichael's conjecture remains open.

Theorem (Ford, 1999)

$C(n)$ *assumes each integer value* > 1 *infinitely often.*

One could also consider the average and typical values of $C(n)$. Studying the average of $C(n)$ is more or less equivalent to understanding the second moment of $N(m)$ (the Davenport–Heilbronn question).

What about the typical size of $C(n)$? Erdős and Pomerance showed that for asymptotically 100% of $n \le x$ (as $x \to \infty$), the number $\varphi(n)$ has $(\frac{1}{2} + o(1))(\log\log x)^2$ prime factors.

Now the number of integers up to x with more than $(\frac{1}{2} + o(1))(\log\log x)^2$ prime factors has size

$$x/\exp((1/2 + o(1))(\log\log x)^2 \log\log\log x).$$

Using this, one can show — as Florian Luca and I did in 2011 – that for 100% of $n \le x$ (as $x \to \infty$), we have

$$C(n) > \exp((1/2 + o(1))(\log\log x)^2 \log\log\log x).$$

Theorem (P., 2018)

For 100% of $n \le x$ (as $x \to \infty$), we have

$$C(n) < \exp((1/2 + o(1))(\log\log x)^2 \log\log\log x).$$

Power values of Euler's function

Question: Is $\varphi(n)$ a square for infinitely many n?

Power values of Euler's function

Question: Is $\varphi(n)$ a square for infinitely many n?

YES, since $\varphi(2^{2k+1}) = (2^k)^2$, or $\varphi(5^{2k+1}) = (2 \cdot 5^k)^2$.

OK, but what if we restrict to **squarefree** n? The answer is still yes.

Here is a sketch of a proof. Consider the numbers $\varphi(p) = p - 1$ for odd primes $p \leq x$. Each $p - 1$ is even and smaller than x, we can write

$$p - 1 = \prod_{\ell \leq x/2} \ell^{v_\ell(p-1)}.$$

To each p, we assign the *exponent vector* $(v_\ell(p-1) \bmod 2)_{\ell \leq x/2}$. This lives vector in $\mathbb{F}_2^k$, where $k = \pi(x/2)$.

To each p, we assign the *exponent vector* $(v_\ell(p-1) \bmod 2)_{\ell \le x/2}$. This lives vector in $\mathbb{F}_2^k$, where $k = \pi(x/2)$.

Suppose that some subset of our p's is such that the corresponding collection of exponent vectors sums to 0 in $\mathbb{F}_2^k$. Then the product of those p's — call this n — has $\varphi(n)$ a square.

How do we know we can find such a subset? Linear algebra to the rescue! The number of vectors is $\pi(x) - 1$, while the dimension of $\mathbb{F}_2^k$ is $k = \pi(x/2)$. So a dependence relation is forced.

So there is at least **one** squarefree n with $\varphi(n) = \square$. But we can re-do the construction after removing the finitely many p's dividing n to produce a new n, etc.

Question: Can we count the number of $n \leq x$ for which $\varphi(n)$ is a square? Call this $V_\square(x)$.

To produce many n by the above construction, one can restrict the primes p in the construction to ones for which $p-1$ is smooth. Note that this puts the exponent vector in an $\mathbb{F}_2$-vector space of small dimension.

Using the Baker–Harman theorem on smooth $p-1$'s, Banks–Friedlander–Pomerance–Shparlinski showed that

$$V_\square(x) > x^{0.7038}$$

for large x.

Theorem (Banks–Friedlander–Pomerance–Shparlinski)

Assuming the aforementioned conjectures on smooth shfited primes,

$$V_{\square}(x) \geq x/L(x)^{1+o(1)}, \quad (x \to \infty).$$

In fact, the same lower bound holds for the number of $n \leq x$ *for which* $\varphi(n)$ *is a* k*th power, for any fixed* k*.*

Theorem (Banks–Friedlander–Pomerance–Shparlinski)

Assuming the aforementioned conjectures on smooth shfited primes,

$$V_{\square}(x) \geq x/L(x)^{1+o(1)}, \quad (x \to \infty).$$

In fact, the same lower bound holds for the number of $n \leq x$ for which $\varphi(n)$ is a kth power, for any fixed k.

Theorem (P., 2018)

The number of $n \leq x$ for which $\varphi(n)$ is squarefull is at most $x/L(x)^{1+o(1)}$, as $x \to \infty$.

PART II: PROOFS

WARNING: Actual proofs may be longer than they appear.

Back to the problem of Davenport–Heilbronn

Theorem

As $x \to \infty$, $\displaystyle\sum_{m \le x} N(m)^2 \le x/L(x)^{2+o(1)}$.

Let $\mathcal{S} = \{m \le x : N(m) > x/L(x)^{100}\}$, and write

$$\sum_{m \le x} N(m)^2 = \sum_{\substack{m \le x \\ m \notin \mathcal{S}}} N(m)^2 + \sum_{m \in \mathcal{S}} N(m)^2.$$

Back to the problem of Davenport–Heilbronn

Theorem

As $x \to \infty$, $\displaystyle\sum_{m \le x} N(m)^2 \le x/L(x)^{2+o(1)}.$

Let $\mathcal{S} = \{m \le x : N(m) > x/L(x)^{100}\}$, and write

$$\sum_{m \le x} N(m)^2 = \sum_{\substack{m \le x \\ m \notin \mathcal{S}}} N(m)^2 + \sum_{m \in \mathcal{S}} N(m)^2.$$

We see easily that

$$\sum_{\substack{m \le x \\ m \notin \mathcal{S}}} N(m)^2 \le \frac{x}{L(x)^{100}} \sum_{\substack{m \le x \\ m \notin S}} N(m) \le \frac{x}{L(x)^{100}} \sum_{m \le x} N(m) \ll \frac{x^2}{L(x)^{100}}.$$

To handle the sum over $m \in \mathcal{S}$, we use our popular subsets theorem. First, since each individual value $N(m) \leq x/L(x)^{1+o(1)}$ we have

$$\sum_{m \in \mathcal{S}} N(m)^2 \leq \frac{x}{L(x)^{1+o(1)}} \sum_{m \in \mathcal{S}} N(m).$$

How large (or small) is $\mathcal{S}$? By definition of $\mathcal{S}$,

$$\frac{x}{L(x)^{100}} \cdot \#\mathcal{S} \leq \sum_{m \in \mathcal{S}} N(m) \leq \sum_{m \leq x} N(m) \ll x,$$

and so

$$\#\mathcal{S} \ll L(x)^{100}.$$

In particular, $\#\mathcal{S} = x^{o(1)}$. Thus, $\sum_{m \in \mathcal{S}} N(m) \leq x/L(x)^{1+o(1)}$. Inserting this above finishes the proof.

Popular subsets

Fix $\alpha \in (0, 1)$.

Fix $\epsilon > 0$. Let $\mathcal{S}$ be a subset of $[1, x]$ with $\#\mathcal{S} \le x^{1-\alpha}$. We want to show that, as long as $x > x_0(\epsilon, \alpha)$,

$$\underbrace{\sum_{m \in \mathcal{S}} N(m)}_{\#\varphi^{-1}(\mathcal{S})} \le x/L(x)^{\alpha-\epsilon}.$$

Key lemma

Lemma

Let $x \geq 3$, and let $d \in \mathbb{Z}^+$. Then for some absolute constant $C > 0$,

$$\#\{n \leq x : d \mid \varphi(n)\} \leq \frac{x}{d}(C(\log\log x)^2)^{\Omega(d)} f(d)$$

where $f(d)$ is the number of unordered factorizations of d. Since $f(d) \leq \Omega(d)^{\Omega(d)}$, this last expression is

$$\leq \frac{x}{d}(C\Omega(d)(\log\log x)^2)^{\Omega(d)}.$$

Proof

With p running over the primes dividing n, write $n = \prod_p p^{e_p}$, so that

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

We think of the right-hand product as a factorization of $\varphi(n)$ into w terms $P_1, \ldots, P_w$, where each $P_i = p$ or $p - 1$ for a $p \mid n$. Here

$$w = \sum_{p \mid n} (1 + (e_p - 1)) = \sum_{p \mid n} e_p = \Omega(n).$$

Let p_i be the prime p associated to P_i in this way.

Since $d \mid \varphi(n) = P_1 \cdots P_w$, we can factor

$$d = d_1 \cdots d_w$$

in such a way that each $d_i \mid P_i$. Since $P_i = p_i$ or $p_i - 1$, we see that

$$p_i \equiv 0 \text{ or } 1 \pmod{d_i}.$$

It could be that some of the $d_i = 1$; renumbering, say that the $d_i > 1$ are precisely $d_1, \cdots d_k$. so

$$d = d_1 \cdots d_k.$$

We now count the number of number of $n \leq x$ that can give rise to the list $d_1, \cdots, d_k$.

We now count the number of number of $n \leq x$ that can give rise to the list $d_1, \cdots, d_k$. The integer n is divisible by $p_1 p_2 \cdots p_k$, where each $p_i \equiv 0$ or $1 \pmod{d_i}$. The number of such $n \leq x$ is at most

$$\sum_{\substack{p_1, \ldots, p_k \\ \text{each } p_i \equiv 1 \pmod{d_i}}} \frac{x}{p_1 \cdots p_k} \leq x \prod_{i=1}^{k} \left(\sum_{\substack{p \leq x \\ p \equiv 1 \pmod{d_i}}} \frac{1}{p} \right).$$

The inner sum is $\ll (\log \log x)/\varphi(d_i)$ (via Brun–Titchmarsh), which is $\ll (\log \log x)^2/d_i$. Thus,

$$\mathrm{RHS} \leq \frac{x}{d}(C(\log \log x)^2)^k.$$

Notice that $k \leq \sum_{i=1}^{k} \Omega(d_i) = \Omega(d)$. We conclude by summing all possible factorizations $d_1, \ldots, d_k$ of d.

How could we apply the key lemma?

Suppose $m \in \mathcal{S}$. Then for an appropriate universal constant C,

$$\begin{aligned} N(m) &= \#\{n : \varphi(n) = m\} \\ &= \#\{n \le Cm \log\log m : \varphi(n) = m\} \\ &\le \#\{n \le Cm \log\log m : m \mid \varphi(n)\}, \end{aligned}$$

which by the Key Lemma is

$$\lessapprox \Omega(m)^{\Omega(m)}.$$

[We suppress factors of size $(\log\log x)^{O(1+\Omega(m))}$, to draw attention to what turns out to be the important term.]

We might try to use this estimate for each $m \in \mathcal{S}$, then sum over the $\le x^{\alpha}$ values of $m \in \mathcal{S}$.

This can't work, because there could be values of $m \in \mathcal{S}$ with $\Omega(m)$ very large, say $\asymp \log x$. Then $\Omega(m)^{\Omega(m)} > x$.

We can try to weed out these m beforehand. Remember, we are trying to count n with $\varphi(n) = m$, so if $\Omega(m)$ is large, then $\Omega(\varphi(n))$ is large.

Maybe one can prove **first** that there are few n with $\Omega(\varphi(n))$ large! Such an argument would allow us to dispense of all m with $\Omega(m)$ large, in one fell swoop. And we would be allowed to assume, in the last slide, that $\Omega(m)$ is small.

Maybe one can prove **first** that there are few n with $\Omega(\varphi(n))$ large!

How few is "few n"? Want $\lessapprox x/L(x)^{1-\alpha}$ such n.

What does it mean for $\Omega(\varphi(n))$ to be "large"? We try to arrange that

$$\Omega(\varphi(n)) \le (1-c)\log x/\log\log x,$$

with c slightly larger than α. Then with $m = \varphi(n)$,

$$\Omega(m)^{\Omega(m)} \le x^{1-c}.$$

Then summing on the $\le x^\alpha$ values of $m \in \mathcal{S}$, we pick up a contribution of $\le x^{1-c+\alpha}$, which is negligible compared to $x/L(x)^{1-\alpha}$.

Unfortunately, this doesn't work! There are too many n with $\Omega(\varphi(n))$ larger than we want.

Solution: Majorize $N(m)$ by

$$\#\{n \leq Cm \log\log m : m' \mid \varphi(n)\},$$

where m' is the component of m consisting only of prime factors larger than a convenient number of z.

If $z = \exp((\log\log x)^{1/2})$, can show that the z-rough part of $\varphi(n)$ usually has small Ω (in the sense we needed). We also use that the z-rough part of $\varphi(n)$ is usually not much smaller than $\varphi(n)$.

Factorizations, revisited

Recall that a *factorization* of an integer is an unordered representation of n as a product of integers > 1.

We saw in the proof of the Key Lemma that every φ-preimage of m induces a factorization of m: If $n = \prod_p p^{e_p}$, then

$$\varphi(n) = \prod_p (p-1)p^{e_p - 1}.$$

Actually, the above representation is not quite a 'factorization' in the technical sense, because the term 1 (=$2-1$) may appear, although if it appears it appears at most once. Call this kind of representation an 'extended factorization'.

It is easy to see that the number of extended factorizations of n is $2f(n)$.

Exercise.(Erdős, Pomerance, Sarközy).
For every m, at most two different φ-preimages of n induce the same factorization of m.

Corollary

$N(m) \leq 4f(m)$.

This makes one wonder how many of our theorems about $N(m)$ are really about $f(m)$.

Theorem (P., 2019)

Fix $\alpha \in (0,1)$. If $\mathcal{S}$ is any subset of $[1,x]$ with $\#\mathcal{S} \le x^{1-\alpha}$, then

$$\sum_{n \le x} f(n) \le x/L(x)^{\alpha+o(1)},$$

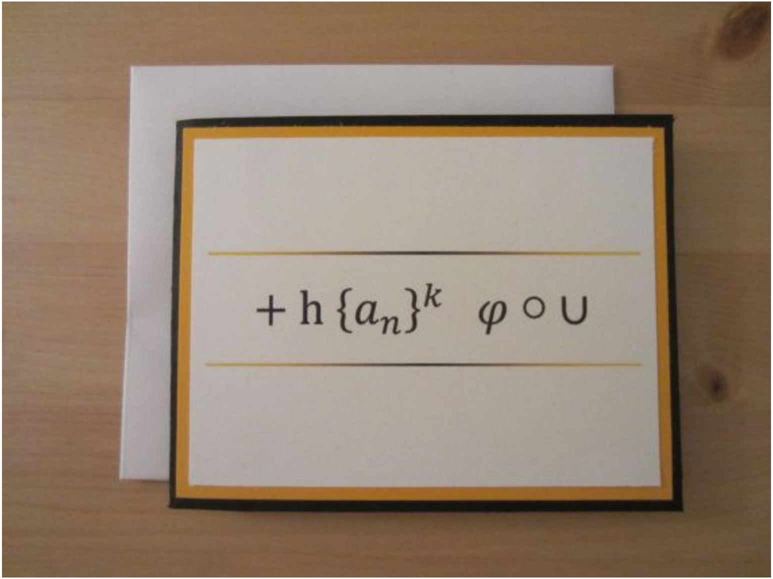
as $x \to \infty$, uniformly in the choice of $\mathcal{S}$.

Theorem (P., 2019)

For each fixed $\beta > 1$,

$$\sum_{m \le x} f(m)^{\beta} = x^{\beta}/L(x)^{\beta+o(1)},$$

as $x \to \infty$.


$+ \mathrm{h} \{a_n\}^k \quad \varphi \circ \cup$