PaNTS VI\textsuperscript{II} in memory of Kevin James
Clemson University, October 21–22, 2023

The shifted-prime divisor function

Kai (Steve) Fan and Carl Pomerance
Dartmouth College
A shifted prime is a number \( p + a \), where \( p \) is prime and \( a \) is any nonzero integer. The view is that \( a \) is fixed and \( p \) runs over primes, so we shift the primes by \( a \). Most questions are not too sensitive about the choice of \( a \), except for some obvious things. Like: are there infinitely many shifted primes \( p + a \) that are also prime? (Conjecturally YES if \( a \) is even. Obviously NO if \( a \) is odd!)

In this talk we will focus on the case \( a = -1 \), that is, numbers of the form \( p - 1 \), since this case has some interesting applications.
Hardy and Ramanujan had shown that a typical integer $n$ has about $\log \log n$ prime factors. In the paper below of Erdős, submitted when he was 21, he showed that most shifted primes $p - 1$ are like typical integers in this regard.

ON THE NORMAL NUMBER OF PRIME FACTORS OF $p - 1$ AND SOME RELATED PROBLEMS CONCERNING EULER'S $\phi$-FUNCTION

By Paul Erdős (Manchester)

[Received 13 November 1934]
Erdős also gave an argument that there exist integers with an extraordinarily large number of shifted-prime divisors. He used this to show a fantastic result about Euler’s function \( \varphi \):

There is a positive number \( c \) such that for all large \( x \) there is a number \( n \leq x \) such that the equation \( \varphi(m) = n \) holds for at least \( x^c \) numbers \( m \).

He conjectured this holds for every fixed \( c < 1 \). There’s been a long history of getting larger values of \( c \), the current record holder is Lichtman, who got \( c = .7156 \).
Let $\omega(n)$ denote the number of primes that divide $n$. For example, $\omega(11) = 1$, $\omega(12) = 2$, etc. We know, essentially from Mertens that

$$\sum_{n \leq x} \omega(n) = \sum_{p \leq x} \left\lfloor \frac{x}{p} \right\rfloor = x \log \log x + O(x).$$

Further, let $\omega^*(n)$ denote the number of shifted-prime divisors $p - 1$ of $n$. For example, $\omega^*(11) = 1$, $\omega^*(12) = 5$, etc. It might seem perhaps that $\omega^*$ more resembles $\tau(n)$, the total number of divisors of $n$. However, the same argument using the Mertens theorem gets us

$$\sum_{n \leq x} \omega^*(n) = \sum_{p \leq x+1} \left\lfloor \frac{x}{p-1} \right\rfloor = x \log \log x + O(x).$$

And of course:

$$\sum_{n \leq x} \tau(n) = \sum_{d \leq x} \left\lfloor \frac{x}{d} \right\rfloor = x \log x + O(x).$$
As mentioned, Hardy and Ramanujan proved that the normal order of \( \omega(n) \) is \( \log \log n \). This result was greatly improved by Erdős and Kac, who showed there is a Gaussian distribution:

\[
\# \{ n \leq x : \omega(n) \leq \log \log n + u \sqrt{\log \log n} \} \sim \frac{x}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2} dt
\]
as \( x \to \infty \). Remarkably, Halberstam proved the same result for shifted primes:

\[
\# \{ p \leq x : \omega(p - 1) \leq \log \log p + u \sqrt{\log \log p} \} \sim \frac{\pi(x)}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2} dt
\]
as \( x \to \infty \).

Predating Erdős and Kac, Turán worked out the second moment of \( \omega(n) \), later generalized by Kubilius to general additive functions:

\[
\sum_{n \leq x} \omega(n)^2 = x(\log \log x)^2 + O(x \log \log x).
\]
The function $\omega^*(n)$, which counts the number of shifted-prime divisors $p - 1$ of $n$, is not additive, but perhaps we have the same theorems as with $\omega(n)$? In particular, what can be said about $\omega^*(n)^2$ on average?

(On the number of divisors of a natural number which have the form $p - 1$.)
Prachar showed here that

$$\sum_{n \leq x} \omega^*(n)^2 = O(x \log^2 x),$$

contrasting this with

$$\sum_{n \leq x} \tau(n)^2 \sim \frac{1}{\pi^2} x \log^3 x,$$

an old result of Ramanujan. Also, in comparison with \( \tau(n) \), he showed that there is a positive constant \( c \) and infinitely many \( n \) with

$$\omega^*(n) > \exp(c \log n/(\log \log n)^2),$$

and that assuming the RH, "\((\log \log n)^2\)" could be replaced with \( \log \log n \). In fact we know unconditionally after Wigert that the maximal order of \( \tau(n) \) is \( \exp((\log 2 + o(1)) \log n/\log \log n) \).
Here a key step in proving the complexity of our algorithm was in proving the Prachar conjecture that

$$\omega^*(n) > \exp(c \log n / \log \log n)$$

for infinitely many $n$. The Prachar argument was also a key step in my paper with Alford and Granville in proving there are infinitely many Carmichael numbers.
But what of the second moment? Here in a brief note added in the journal by Erdős, jointly with Prachar, the exponent 2 on \( \log x \) was sort of reduced to 1.

\[
\text{Über die Anzahl der Lösungen von } [p-1, q-1] \leq x.
\]

\((\text{Aus einem Brief von P. Erdös an K. Prachar})\)

\((\text{Eingelangt am 18. Mai 1955.})\)

(On the number of solutions of \([p-1, q-1] \leq x\).)

Why “sort of”? 
Well, they proved that the number of solutions to \([p - 1, q - 1] \leq x\) is \(O(x(\log \log x)^3)\) and they indicated even more briefly in a footnote how the log log factor could be removed, using some standard tricks and the \textbf{Titchmarsh} divisor problem.

What is the connection to the average of \(\omega^*(n)^2\)? Well, assume that the number of solutions to \([p - 1, q - 1] \leq t\) is \(O(t)\). Then

\[
\sum_{n \leq x} \omega^*(n)^2 = \sum_{n \leq x} \left( \sum_{p - 1 \mid n} 1 \right)^2 = \sum_{n \leq x} \sum_{[p - 1, q - 1] \mid n} 1 \\
= \sum_{[p - 1, q - 1] \leq x} \left( \frac{x}{[p - 1, q - 1]} \right) \leq x \sum_{[p - 1, q - 1] \leq x} \frac{1}{[p - 1, q - 1]}.
\]

And partial summation shows the last sum is \(O(\log x)\).
What is the **Titchmarsh** divisor problem? This is the assertion that (recall that $\tau$ counts the total number of divisors)

$$\sum_{p \leq x} \tau(p - 1) \sim cx, \quad x \to \infty$$

for a certain constant $c > 0$. So, on average, $\tau(p - 1)$ is $\sim c \log x$ for $p \leq x$, which contrasts nicely with all integers $n \leq x$, where on average it is $\sim \log x$. **Titchmarsh** gave a GRH-conditional proof of this in 1931, and **Linnik** gave a difficult, but unconditional proof in 1963. Using the duality between a divisor $d$ of $p - 1$ and its co-divisor $(p - 1)/d$, a quick proof can be had using the **Bombieri–Vinogradov** theorem, as noted by **Rodriguez** in 1965 and **Halberstam** in 1967.

It would be nice to have a clean, self-contained proof that $\omega^*(n)^2$, on average for $n \leq x$, is $O(\log x)$. Or possibly better?
These issues were discussed in the recent paper:

In addition to proving that \( \sum_{n \leq x} \omega^*(n)^2 = O(x \log x) \), they conjectured that this is best possible. And they proved a nontrivial lower bound of \( cx(\log \log x)^3 \).
In a very recent paper, Ding addresses this conjecture, claiming a proof.

However, there’s a small problem with these lower bounds. In the Murty–Murty paper we find the equation

\[
\sum_{n \leq x} \omega^*(n)^2 = \sum_{[p-1, q-1] \leq x} \left[ \frac{x}{[p-1, q-1]} \right] = \sum_{p, q \leq x} \frac{x}{[p-1, q-1]} + O(x).
\]
Repeating:

\[ \sum_{n \leq x} \omega^*(n)^2 = \sum_{[p-1, q-1] \leq x} \left\lfloor \frac{x}{[p-1, q-1]} \right\rfloor = \sum_{p, q \leq x} \frac{x}{[p-1, q-1]} + O(x). \]

This is justified by the earlier result that the number of \( p, q \) with \( [p-1, q-1] \leq x \) is \( O(x) \). Yes, removing the floor symbol for terms with \( [p-1, q-1] \leq x \) creates an error of \( O(x) \), but the number of pairs \( p, q \) that are now considered in this last sum is expanded. And it is not clear if the extra terms have a sum that is \( O(x) \).

In fact, this is not just an error in the proof, the assertion is incorrect as well:

**Theorem (Fan, P).** We have

\[ \sum_{p, q \leq x} \frac{x}{[p-1, q-1]} \gg x \log x. \]
Fortunately, there’s an easy fix that saves the lower bound arguments of Murty–Murty and Ding:

$$\sum_{n \leq x} \omega^*(n)^2 \geq \sum_{p,q \leq \sqrt{x}} \left| \frac{x}{[p-1, q-1]} \right| = \sum_{p,q \leq \sqrt{x}} \frac{x}{[p-1, q-1]} + O(x).$$

And from the Ding proof, this is $\gg x \log x$.

However, more recently, Ding (arXiv:2209.01087v1) gives a heuristic argument for a certain constant $c$ with

$$\sum_{n \leq x} \omega(n)^2 \sim cx \log x,$$

namely $c = 2\zeta(2)\zeta(3)/\zeta(6) \approx 3.8$. This is almost certainly wrong in view of our theorem. We believe the correct constant should be about 3.1 and we’re working on trying to show this, at least heuristically.
We have tried some computing and here is what we get up to various limits $x$ for the sum of $\omega^*(n)^2$, and then the sum divided by $x \log x$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>sum to $10^k$</th>
<th>$\div 10^k \log 10^k$</th>
<th>$3 \times 10^k \log 10^k - 6 \times 10^k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>971</td>
<td>2.11</td>
<td>782</td>
</tr>
<tr>
<td>3</td>
<td>15,530</td>
<td>2.25</td>
<td>14,723</td>
</tr>
<tr>
<td>4</td>
<td>219,128</td>
<td>2.38</td>
<td>216,310</td>
</tr>
<tr>
<td>5</td>
<td>2,849,311</td>
<td>2.47</td>
<td>2,853,878</td>
</tr>
<tr>
<td>6</td>
<td>35,261,891</td>
<td>2.55</td>
<td>35,446,532</td>
</tr>
<tr>
<td>7</td>
<td>421,296,839</td>
<td>2.61</td>
<td>423,542,870</td>
</tr>
<tr>
<td>8</td>
<td>4,902,181,351</td>
<td>2.66</td>
<td>4,926,204,223</td>
</tr>
<tr>
<td>9</td>
<td>56,067,311,859</td>
<td>2.71</td>
<td>56,169,797,511</td>
</tr>
<tr>
<td>10</td>
<td>631,033,824,202</td>
<td>2.74</td>
<td>630,775,527,898</td>
</tr>
</tbody>
</table>

The last column is a pure guess that seems to fit fairly well!
Some other natural questions about $\omega^*(n)$ is how it behaves for “most” values of $n$. Strangely, even though for very special values of $n$, $\omega^*(n)$ is much bigger than $\omega(n)$, for most values of $n$, it is quite a bit smaller; in fact it is essentially $O(1)$.

That something like this is true can be seen already from the obvious fact that $\omega^*(n) = 1$ whenever $n$ is odd (and conversely). Beyond odd and even though, it is less clear that $\omega^*(n)$ is usually small.

An important ingredient in showing this is a shocking (at first sight) theorem of Erdős–Wagstaff from 1980: There is a positive constant $c$ such that the number of $n \leq x$ divisible by a shifted prime $p-1 > y$ is $O(x/\log^c y)$. That is, most integers are not divisible by any large shifted primes! In a more recent paper, McNew, Pollack, P (2017) show that the “correct” value of $c$ is $1 - (1 + \log \log 2)/\log 2 \approx .086$. 
Already from the Erdős–Wagstaff paper we know that for each fixed $k$, the set $\{n : \omega^*(n) = k\}$ has an asymptotic density. In fact, this density is positive, a fact that is perhaps not so easy to prove (we used Chen’s theorem).

**Theorem (Fan, P).** For each positive integer $k$, $\{n : \omega^*(n) = k\}$ has a positive asymptotic density $\delta_k$ and $\sum \delta_k = 1$.

Note that if $\omega^*(n) \geq y$, then the largest shifted-prime divisor of $n$ is $\gg y \log y$, and is in particular $\geq y$. So, we have the dichotomy:

For each $y > 1$, the set $\{n : \omega(n) \geq y\}$ has asymptotic density 1 while the set $\{n : \omega^*(n) \geq y\}$ has asymptotic density $o(1)$ as $y \to \infty$. 

18
**Theorem (Fan, P).** For $y \geq 2$, $N(x, y) = O(x(\log y)/y)$, where $N(x, y) = \#\{n \leq x : \omega^*(n) \geq y\}$.

For the proof, we may assume that $y$ is large. Note that from the average of $\omega^*(n)$ being $O(\log \log x)$ for $n \leq x$, it follows that $N(x, y) = O(x(\log \log x)/y)$. Thus, the theorem follows when $y > (\log x)^{1/20}$, say.

Now assume that $y \leq (\log x)^{1/20}$ and let $z = \exp(y^{19}) = x^{o(1)}$. If $n$ is divisible by a shifted prime $p - 1 > z$, then by the **McNew, Pollack, P** theorem, the count of such numbers $n$ is $x/(\log z)^{\beta + o(1)}$, where $\beta = 1 - (1 + \log \log 2)/\log 2$. Since $19\beta > 1$, we have $(\log z)^{\beta + o(1)} = (\log y)^{19\beta + o(1)} > \log y$ when $y$ is large. So, these numbers $n$ are negligible.
So assume that both $\omega^*(n) \geq y$ and that each shifted-prime divisor of $n$ is at most $z$.

Let $\omega^*_z(n)$ denote the number of shifted-prime divisors $p - 1$ of $n$ with $p - 1 \leq z$. The average order argument (Mertens’ theorem) shows that $\omega^*_z(n)$ is $O(\log \log z)$ on average, so that the number of $n$ left in our count is $O\left(x(\log \log z)/y\right)$. But since $z = \exp(y^{19})$, we have $\log \log z = O(\log y)$ and we’re done.

We can also ask for a lower bound for $N(x, y)$, the number of $n \leq x$ with $\omega^*(n) \geq y$. Here, by considering the first integer $m$ with $\omega^*(m) \geq y$, we can show

$$N(x, y) > \frac{x}{y^c \log \log y}$$

for a positive constant $c$ and all large $y$.  

20
Other problems to consider are estimates of the densities $\delta_k$ (the density of the set of numbers $n$ with $\omega^*(n) = k$). Though $\delta_k \to 0$ as $k \to \infty$, the sequence is likely not monotone.

On the next slide are exact counts of integers $n \leq 10^6, 10^8, 10^{10}$ with $\omega^*(n) = k$ for $k \leq 11$. (For each $k \geq 12$ there were fewer than 1% of the $n$’s with $\omega^*(n) = k$.)
<table>
<thead>
<tr>
<th>$k$</th>
<th>$10^6$</th>
<th>$10^8$</th>
<th>$10^{10}$</th>
<th>$\approx \delta_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>500,000</td>
<td>50,000,000</td>
<td>5,000,000,000</td>
<td>.5</td>
</tr>
<tr>
<td>2</td>
<td>77,696</td>
<td>7,436,825</td>
<td>720,726,912</td>
<td>.070</td>
</tr>
<tr>
<td>3</td>
<td>91,602</td>
<td>8,826,498</td>
<td>859,002,140</td>
<td>.084</td>
</tr>
<tr>
<td>4</td>
<td>79,986</td>
<td>7,691,971</td>
<td>748,412,490</td>
<td>.074</td>
</tr>
<tr>
<td>5</td>
<td>59,518</td>
<td>5,684,323</td>
<td>555,900,984</td>
<td>.055</td>
</tr>
<tr>
<td>6</td>
<td>40,641</td>
<td>4,031,009</td>
<td>401,146,301</td>
<td>.040</td>
</tr>
<tr>
<td>7</td>
<td>29,565</td>
<td>3,016,881</td>
<td>300,330,932</td>
<td>.030</td>
</tr>
<tr>
<td>8</td>
<td>23,190</td>
<td>2,324,769</td>
<td>233,611,502</td>
<td>.023</td>
</tr>
<tr>
<td>9</td>
<td>17,914</td>
<td>1,800,298</td>
<td>182,793,491</td>
<td>.018</td>
</tr>
<tr>
<td>10</td>
<td>13,899</td>
<td>1,401,307</td>
<td>144,740,573</td>
<td>.015</td>
</tr>
<tr>
<td>11</td>
<td>10,487</td>
<td>1,131,836</td>
<td>118,302,267</td>
<td>.012</td>
</tr>
<tr>
<td>$\geq 12$</td>
<td>55,682</td>
<td>6,654,283</td>
<td>735,032,408</td>
<td></td>
</tr>
</tbody>
</table>

Exact counts of $n \leq 10^6, 10^8, 10^{10}$ with $\omega^*(n) = k$

(Largest $k$ values $10^6$: 86, $10^8$: 231, $10^{10}$: 519)
Thank you