

https://dspace.org/

Composite numbers n whose sum of prime factors is prime

Moodley, Damon

2021

https://hdl.handle.net/10539/32451

Moodley, Damon (2021) Composite numbers n whose sum of prime factors is prime, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/32451>

Downloaded from a DSpace repository. https://dspace.org/



Composite numbers n whose sum of prime factors is prime

Damon Moodley
School of Mathematics, University of the Witwatersrand
1 Jan Smuts Avenue, Braamfontein 2000, Johannesburg, South Africa
747011@students.wits.ac.za

April 25, 2021

1 Declaration

I declare that this thesis is my own, unaided work. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

Signature: _____

____day of ____20____at _____

2 Acknowledgement

A special acknowledgement to Florian Luca for the supervision that he has given me, in writing this thesis.

Contents

1 Declaration ii

2 Acknowledgement iii

1 Introduction 1

2 Background 2
- 2.1 Preliminaries . . . 2
- 2.2 The Abel summation formula . . . 5
- 2.3 The Prime Number Theorem . . . 16
- 2.4 The Divisor function . . . 19
- 2.5 Sieves . . . 23
- 2.6 The Bateman–Horn Conjecture . . . 37
- 2.7 Chudakov . . . 38
- 2.8 Smooth numbers and the Dickman–de Brujin function . . . 39
- 2.9 Motivation for this Thesis . . . 45
  - 2.9.1 Composite numbers n such that the sum of its distinct divisors divide n. . . . 45
  - 2.9.2 Product of twin primes . . . 45
  - 2.9.3 Vinogradov's theorem . . . 46
  - 2.9.4 $\beta(n)$ being prime . . . 46

3 Composite numbers n whose sum of prime factors is prime: The upper bound 47
- 3.1 Eliminating smooth numbers from $\mathcal{B}(x)$ . . . 47
- 3.2 Eliminating numbers divisible by the square of a large prime 48
- 3.3 Applying the sieve . . . 48
- 3.4 Eliminating smooth m's . . . 50
- 3.5 The case $m \leq x^{9/10}$ . . . 51
- 3.6 The case m large . . . 56

4 Composite numbers n whose sum of prime factors is prime: The lower bound 60
- 4.1 A conditional lower bound . . . 60
- 4.2 An unconditional lower bound . . . 61
- 4.3 A further conjecture on the count of $\mathcal{B}(x)$ . . . 63

5 Remarks 66

Composite numbers n whose sum of prime factors is prime

Abstract

In this thesis, we estimate the number of composite positive integers $n \leq x$ such that $\beta(n) = \sum_{p|n} p$ is a prime.

1 Introduction

In this dissertation, we will look for positive integers n such that $\beta(n) = \sum_{p|n} p = P$, where P is prime. We write p, P to indicate prime numbers. The function $\beta(n)$ has already been studied by a few authors. Its first appearance was in the papers of Pomerance and his co–authors, including Erdős (see [9], [13], [14]), where the positive integers n with $\beta(n) = \beta(n+1)$ were investigated. An example of such n is 714, a number which at that time appeared in the context of a baseball game, and since then such numbers have been called Ruth–Aaron numbers. De Koninck and Luca studied positive integers n with $\beta(n) \mid n$ (see [5], [6]).

This thesis is based on a paper by Luca and Moodley, which estimated the number of composite integers $n \leq x$ such that $\beta(n)$ is a prime. An example of this is 210 where $\beta(210) = 17$.

We shall now proceed in starting this thesis formally.

The main result of the paper which this thesis is based on is as follows:

Let

$$\mathcal{B} = \{n \text{ composite} : \beta(n) \text{ is prime}\}.$$

For a subset $\mathcal{A}$ of positive integers and a real number $x \geq 1$, we write

$$\mathcal{A}(x) = \mathcal{A} \cap [1, x].$$

We have the following theorem.

Theorem 1.1. *The estimates*

$$\frac{x}{(\log x)^3} \ll \#\mathcal{B}(x) \ll \frac{x}{\log x}$$

hold [12].

We believe the upper bound is closer to the truth. We can prove a lower bound of the same order of magnitude as the upper bound assuming a uniform version of the Bateman–Horn Conjecture.

We will first provide some definitions needed to read the proof of the main result, motivation for the study of this question, and methods used in answering this question.

2 Background

Most of the results which will be stated if no other citation is given can be found using the following reference [7].

Throughout this thesis, the Landau symbols O and o as well as the Vinogradov symbols $\ll$ and $\gg$ are used. The constants implied by them are absolute. We shall provide definitions below for the notation mentioned above. We shall denote the natural logarithm of x as $\log x$.

2.1 Preliminaries

Definition 2.1. *A prime number denoted p, is a positive integer larger than 1, which is not divisible by any other natural number besides itself and 1.*

Definition 2.2 (Fundamental Theorem of Arithmetic)**.** *Every integer $n > 1$ can be written as a product of primes. This factorization is unique, up to the ordering of the factors.*

Definition 2.3. *Given two functions f and g defined on $[a, \infty)$, where $a \geq 0$ and $g(x) > 0$, we shall write that $f(x) = O(g(x))$ (Or $f(x) \ll g(x)$) if there exist two constants $M > 0$ and x_0 such that $|f(x)| < Mg(x)$ for all $x \geq x_0$. In particular, $f(x) = O(1)$ if $f(x)$ is a bounded function.*

It is easy to prove that

$$\begin{aligned} x = O(x^2), \qquad & (x+k)^2 = O(x^2) = x^2 + O(x) \quad (k \text{ is fixed}), \\ \log x = O(x), \qquad & x = O(e^x). \end{aligned}$$

Definition 2.4. *Given two functions f and g defined on $[a,\infty)$ with $a \geq 0$, we write $f(x) \sim g(x)$ to mean that* $\lim_{x\to\infty} \dfrac{f(x)}{g(x)} = 1$.

Finally, if we have $g(x) \ll f(x)$ and $f(x) \ll g(x)$ simultaneously then we write $f(x) \asymp g(x)$.

Definition 2.5. *Given two functions f and g defined on $[a,\infty)$ with $a \geq 0$, and $g(x) \geq 0$, we write $f(x) = o(g(x))$ as $x \to \infty$ if, for each $\varepsilon > 0$, there exists a constant $x_0 = x_0(\varepsilon)$ such that $|f(x)| < \varepsilon g(x)$ for all $x \geq x_0$. We can then write*

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0.$$

Well known results are:

$$\begin{aligned} \frac{1}{x} &= o(1), \qquad & x &= o(e^x), \\ \log x &= o(x), \qquad & \sin x &= o(e^x). \end{aligned}$$

Here we provide a small exercise, which will be applied extensively in the main proof.

Example 2.1. *For positive real numbers a and b, the following holds:*

$$(\log x)^a = o(x^b).$$

Proof. For any fixed a, b let $c = b/a$, let $y = x^c$ and let $\varepsilon > 0$. Now let us take the series expansion for $e^{\varepsilon y}$

$$e^{\varepsilon y} = 1 + \varepsilon y + \frac{(\varepsilon y)^2}{2!} + \cdots .$$

Thus we get

$$\begin{aligned} \varepsilon y &< e^{\varepsilon y} \\ \log\varepsilon + \log y &< \varepsilon y \\ \log y &< \varepsilon y - \log\varepsilon \\ \log y &< 2\varepsilon y. \end{aligned}$$

The last inequality follows from the previous one provided that $y > (1/\varepsilon)^{(1/\varepsilon)}$. If this is so, then we will have

$$\begin{aligned} \log y &< 2\varepsilon y \quad (\varepsilon > 0 \text{ is arbitrary}), \text{ so} \\ \log y &= o(y). \end{aligned}$$

Using $y = x^c$,

$$\begin{aligned} c \log x &= o(x^c) \\ \log x &= o(x^c). \end{aligned}$$

Recalling $c = b/a$, the result follows:

$$(\log x)^a = o(x^b).$$

□

2.2 The Abel summation formula

Proposition 1 (Abel summation formula)**.** *Let $\{a_n\}_{n\geq 1}$ be a sequence of complex numbers and let $f : [1,\infty) \to \mathbb{C}$. For each real number $x \geq 1$, let $A(x) = \sum_{n\leq x} a_n$ and assume that $f(x)$ has a continuous derivative for $x \geq 1$. Then*

$$\sum_{n\leq x} a_n f(n) = A(x)f(x) - \int_1^x A(t)f'(t)dt.$$

Proof. Note we can write $a_1 = A_1$ and $a(n) = A(n) - A(n-1)$, for all $n \geq 2$. We will give the proof when x is an integer. We put $x = N$.
Let us expand this summation

$$\begin{aligned}
\sum_{n\leq N} a_n f(n) &= a_1 f(1) + \cdots + a_N f(N) \\
&= A(1)f(1) + (A(2) - A(1))f(2) + \cdots \\
&\quad \cdots + (A(N) - A(N-1))f(N) \\
&= A(N)f(N) - A(N-1)f(N) + A(N-1)f(N-1) + \cdots \\
&\quad \cdots + A(2)f(2) - A(1)f(2) + A(1)f(1) \\
&= A(N)f(N) - A(1)(f(2) - f(1)) - \cdots \\
&\quad - \cdots A(N-1)(f(N) - f(N-1)).
\end{aligned}$$

We see that $f(i+1) - f(i) = \int_i^{i+1} f'(t)dt$, for $i = 1, \ldots, N-1$, and that $A(t)$ is constant on the interval $[i, i+1)$. Therefore,

$$\begin{aligned}
\sum_{n\leq N} a_n f(n) &= A(N)f(N) - \sum_{i=1}^{N-1} A(i) \int_i^{i+1} f'(t)dt \\
&= A(x)f(x) - \int_1^x A(t)f'(t)dt.
\end{aligned}$$

Now we shall prove the result when x is not an integer. We set $N = \lfloor x \rfloor$. $A(t)$ then is constant on the interval $[N, x]$. We can now express

$$A(x)f(x) - \int_1^x A(t)f'(t)dt,$$

as the following:

$$\begin{aligned} A(x)f(x) &- \int_1^x A(t)f'(t)dt = A(x)f(x) - \int_N^x A(t)f'(t)dt - \int_1^N A(t)f'(t)dt \\ &= A(x)f(x) - A(N)\int_N^x f'(t)dt - \int_1^N A(t)f'(t)dt \\ &= A(x)f(x) - A(N)(f(x) - f(N)) - \int_1^N A(t)f'(t)dt \\ &= A(N)f(N) - \int_1^N A(t)f'(t)dt \\ &= \sum_{n\leq N} a_n f(n), \end{aligned}$$

where the last equality holds since we proved that the result holds when $x = N$ is an integer. We have $\sum_{n\leq x} a_n f(n) = \sum_{n\leq N} a_n f(n)$, thus we can conclude the proof. □

We use Abel's summation formula to evaluate a number of sums of continuous differentiable functions in a certain interval. The following results make use of Abel's summation formula. We shall provide a quick proof, which shall demonstrate why this tool is so powerful.

Definition 2.6.

$$\gamma = 1 - \int_1^\infty \frac{t - \lfloor t \rfloor}{t^2} dt.$$

This is Euler's constant and is approximately $0.57721\ldots$

Theorem 2.1. *For all $x \geq 1$,*

$$\sum_{n \leq x} \frac{1}{n} = \log x + \gamma + O\left(\frac{1}{x}\right).$$

Proof. Let us choose our $a_n = 1$ and $f(n) = 1/n$.
So $A(x) = \sum_{1 \leq n \leq x} 1 = \lfloor x \rfloor$, and $f'(t) = -1/t^2$. Applying Proposition 1, we get

$$\begin{aligned}
\sum_{n \leq x} \frac{1}{n} &= \frac{\lfloor x \rfloor}{x} + \int_1^x \frac{\lfloor t \rfloor}{t^2} dt \\
&= \frac{x - (x - \lfloor x \rfloor)}{x} + \int_1^x \frac{t - (t - \lfloor t \rfloor)}{t^2} dt \\
&= 1 + O\left(\frac{1}{x}\right) + \log x - \left(\int_1^\infty \frac{t - \lfloor t \rfloor}{t^2} - \int_x^\infty \frac{t - \lfloor t \rfloor}{t^2}\right) \\
&= \log x + \left(1 - \int_1^\infty \frac{t - \lfloor t \rfloor}{t^2}\right) + O\left(\frac{1}{x}\right) + \int_x^\infty \frac{t - \lfloor t \rfloor}{t^2} \\
&= \log x + \gamma + O\left(\frac{1}{x}\right) + \int_x^\infty \frac{t - \lfloor t \rfloor}{t^2} \\
&= \log x + \gamma + O\left(\frac{1}{x}\right) + O\left(\int_x^\infty \frac{1}{t^2}\right) \\
&= \log x + \gamma + O\left(\frac{1}{x}\right).
\end{aligned}$$

□

Definition 2.7. *The Von Mangoldt function is defined as:*

$$\Lambda(n) = \begin{cases} \log p \text{ if } n = p^k \text{ for some } k \geq 1, \\ 0 \quad \text{otherwise.} \end{cases}$$

Furthermore, the following functions shall be defined, for $x > 0$,

$$\theta(x) = \sum_{p \leq x} \log p,$$

and

$$\psi(x) = \sum_{n \leq x} \Lambda(n).$$

Theorem 2.2. *As $x \to \infty$, then we have,*

$$\theta(x) \asymp x, \text{ and } \psi(x) \sim \theta(x).$$

A stronger result is

$$\pi(x) \sim \frac{\theta(x)}{\log x}$$

and thus, $\theta(x) \sim x$ as $x \to \infty$.

Lemma 1. *Let $e_p(n!)$ be the exponent of p that appears in the factorization of $n!$. Then*

$$e_p(n!) = \sum_{k \geq 1} \left\lfloor \frac{n}{p^k} \right\rfloor.$$

Proof. The proof will be done by induction on n. For $n = 1$, this holds as no prime appears in this factorization. Assume that it holds for n. We now show that it is true for $n + 1$. First we write $n + 1 = p^u m$, where $p \nmid m$. Then, by the induction hypothesis,

$$e_p((n+1)!) = e_p(n!) + u = \sum_{k=1}^{u} \left(\left\lfloor \frac{n}{p^k} \right\rfloor + 1 \right) + \sum_{k > u} \left\lfloor \frac{n}{p^k} \right\rfloor.$$

All that is left to observe is that

$$\left\lfloor \frac{n+1}{p^k} \right\rfloor = \left\lfloor \frac{n}{p^k} \right\rfloor + 1 \text{ if } 1 \leq k \leq u \quad \text{and} \quad \left\lfloor \frac{n+1}{p^k} \right\rfloor = \left\lfloor \frac{n}{p^k} \right\rfloor \text{ if } k > u.$$

□

Theorem 2.3. *As* $x \to \infty$,

$$\sum_{n \le x} \frac{\Lambda(n)}{n} = \log x + O(1).$$

Proof. We shall apply the Abel summation formula, letting $a_n = 1$, and $f(t) = \log t$. Then we get that,

$$\begin{aligned}
\sum_{n \le x} \log n &= \lfloor x \rfloor \log x - \int_1^x \frac{\lfloor t \rfloor}{t} dt \\
&= (x - (x - \lfloor x \rfloor)) \log x - \int \frac{t - (t - \lfloor t \rfloor)}{t} dt \\
&= x \log x + O(\log x) - \int_1^x \frac{t}{t} dt + O\left(\int_1^x \frac{dt}{t} \right) \\
&= x \log x + O(\log x) - x + 1 + O(\log x) \\
&= x \log x - x + O(\log x).
\end{aligned}$$

On the other hand, using the Lemma 1, we also have that,

$$\begin{aligned}
\sum_{n \le x} \log n &= \log(\lfloor x \rfloor !) \\
&= \sum_{p \le x} \left(\sum_{k=1}^{\infty} \left\lfloor \frac{x}{p^k} \right\rfloor \right) \log p \\
&= \sum_{\substack{p^k \le x \\ k \ge 1}} \left\lfloor \frac{x}{p^k} \right\rfloor \log p \\
&= \sum_{n \le x} \left\lfloor \frac{x}{n} \right\rfloor \Lambda(n) \\
&= \sum_{n \le x} \frac{x}{n} \Lambda(n) - \sum_{n \le x} \left(\frac{x}{n} - \left\lfloor \frac{x}{n} \right\rfloor \right) \Lambda(n) \\
&= x \sum_{n \le x} \frac{\Lambda(n)}{n} + O\left(\sum_{n \le x} \Lambda(n) \right).
\end{aligned}$$

Since

$$\sum_{n \le x} \Lambda(n) = \psi(x) = O(x),$$

we get that

$$\sum_{n\leq x}\log n = x\sum_{n\leq x}\frac{\Lambda(n)}{n}+O(x).$$

So we see that

$$x\log x - x + O(\log x) = x\sum_{n\leq x}\frac{\Lambda(n)}{n}+O(x),$$

and we now divide out by x, and get the following:

$$\log x + O\left(\frac{\log x}{x}\right) = \log x + O(1) = \sum_{n\leq x}\frac{\Lambda(n)}{n}+O(1).$$

□

This next result will aid us in proving a crucial result.

Theorem 2.4. *The estimate*

$$\sum_{p\leq x}\frac{\log p}{p} = \log x + O(1),$$

holds as $x\to\infty$.

Proof. We observe that,

$$\begin{aligned}\sum_{p\leq x}\frac{\log p}{p} &= \sum_{n\leq x}\frac{\Lambda(n)}{n} - \sum_{k\geq 2}\sum_{p^k\leq x}\frac{\log p}{p^k}\\ &= \log x + O(1) - \sum_{k\geq 2}\sum_{p^k\leq x}\frac{\log p}{p^k}.\end{aligned}$$

We wish to now show that the double sum is convergent and then we are done.

$$\begin{aligned}\sum_{k\geq 2}\sum_{p^k\leq x}\frac{\log p}{p^k} &\leq \sum_{p}\left(\frac{1}{p^2}+\frac{1}{p^3}+\cdots\right)\log p\\ &= \sum_{p}\left(\frac{1}{1-\frac{1}{p}}-1-\frac{1}{p}\right)\log p\\ &= \sum_{p}\frac{\log p}{p(p-1)} = O(1).\end{aligned}$$

□

Theorem 2.5 (Mertens' estimate)**.** *There exists a constant β such that the estimate*

$$\sum_{p\leq x}\frac{1}{p}=\log\log x+\beta+O\left(\frac{1}{\log x}\right)$$

holds as $x\to\infty$.
Note: The above estimate implies that as $x\to\infty$,

$$\sum_{p\leq x}\frac{1}{p}\sim\log\log x.$$

Proof. Here we will sketch the proof as this is a very important estimate. We will define

$$a_n\begin{cases}\log p/p & \text{if } n=p,\\ 0 & \text{otherwise,}\end{cases}$$

and let $f(t)=1/\log t$. Then $f'(t)=-1/(t(\log t)^2)$. Then, by Theorem 2.4,

$$A(x)=\sum_{n\leq x}a_n=\sum_{p\leq x}\frac{\log p}{p}=\log x+O(1).$$

So, here if we take

$$\sum_{p\leq x}\frac{1}{p}=\sum_{p\leq x}\frac{\log p}{p\log p},$$

then choose $a_n=\log p/p$ and $f(t)=1/\log t$, we can then apply Abel's summation formula.

$$\begin{aligned}\sum_{p\leq x}\frac{\log p}{p\log p} &= \frac{\log x+O(1)}{\log x}-\int_2^x\frac{\log t+O(1)}{-t(\log t)^2}dt\\ &= 1+O\left(\frac{1}{\log x}\right)+\int_2^x\frac{1}{t\log t}dt+O\left(\int_2^x\frac{1}{t(\log t)^2}dt\right)\\ &= 1+O\left(\frac{1}{\log x}\right)+\log\log x-\log\log 2+O\left(\int_2^\infty\frac{1}{t(\log t)^2}dt\right)\\ &\quad +O\left(\int_x^\infty\frac{1}{t(\log t)^2}dt\right)\\ &= \log\log x+\beta+O\left(\frac{1}{\log x}\right)+O\left(\int_x^\infty\frac{1}{t(\log t)^2}dt\right)\\ &= \log\log x+\beta+O\left(\frac{1}{\log x}\right)+O\left(\frac{1}{\log t}\Big|_x^\infty\right)\\ &= \log\log x+\beta+O\left(\frac{1}{\log x}\right),\end{aligned}$$

where

$$\beta = 1 - \log\log 2 + O\left(\int_2^\infty \frac{1}{t(\log t)^2}dt\right).$$

□

Theorem 2.6. *Let $2 \leq n \leq x$. Then the following holds:*

$$\sum_{2\leq n\leq x} \frac{1}{n\log n} = \log\log x + \beta + \frac{\gamma}{\log 2} + O\left(\frac{1}{\log x}\right).$$

Proof. Let us take note of

$$\frac{1}{(t\log t)^2} = O\left(\frac{1}{t(\log t)^2}\right).$$

Using Theorem 2.1, we have that

$$\sum_{2\leq n\leq x} \frac{1}{n} = \log x + \gamma + O\left(\frac{1}{x}\right),$$

and thus we apply the summation formula again, choosing $A(t) = \sum_{2\leq t\leq n} 1/t$, and $f(t) = 1/\log t$.

$$\begin{aligned}
\sum_{2\leq n\leq x} \frac{1}{n\log n} &= \frac{\log x}{\log x} + \frac{\gamma}{\log x} + O\left(\frac{1}{x\log x}\right) + \int_2^x \frac{\log t}{t(\log t)^2}dt \\
&\quad + \int_2^x \frac{\gamma}{t(\log t)^2}dt + O\left(\int_2^x \frac{1}{t^2(\log t)^2}dt\right) \\
&= 1 + \frac{\gamma}{\log x} + O\left(\frac{1}{x\log x}\right) + \log\log x - \log\log 2 \\
&\quad - \frac{\gamma}{\log t}\Big|_2^\infty + \frac{\gamma}{\log t}\Big|_x^\infty + O\left(\int_2^x \frac{1}{t^2(\log t)^2}dt\right) \\
&= 1 + \frac{\gamma}{\log x} + O\left(\frac{1}{x\log x}\right) + \log\log x - \log\log 2 \\
&\quad + \frac{\gamma}{\log 2} - \frac{\gamma}{\log x} + O\left(\int_2^\infty \frac{1}{t(\log t)^2}dt\right) \\
&\quad + O\left(\int_x^\infty \frac{1}{t(\log t)^2}dt\right) \\
&= \log\log x + O\left(\frac{1}{x\log x}\right) + \beta + \frac{\gamma}{\log 2} + O\left(\frac{1}{\log x}\right) \\
&= \log\log x + \beta + \frac{\gamma}{\log 2} + O\left(\frac{1}{\log x}\right).
\end{aligned}$$

□

Note 2.1. *It can be shown that the constant β appearing in the estimate Theorem 2.5 is given by*

$$\beta = \gamma + \sum_{p} \left(\log\left(1 - \frac{1}{p}\right) + \frac{1}{p} \right) \approx 0.26149.$$

Note 2.2. *Let's take a set $A(x)$ and let us say that $\#A(x) \ll x/\log x$ for $2 \leq n \leq x$. Now we wish to evaluate the following*

$$\sum_{\substack{2\leq n\leq x\\ n\in A(x)}} \frac{1}{n}.$$

Proof. We shall make use of the Abel summation formula Proposition 1. Letting $a_n = 1$ and $f(t) = 1/t$, we have

$$\begin{aligned}
\sum_{\substack{2\leq n\leq x\\ n\in A(x)}} \frac{1}{n} &= \frac{\#A(x)}{x} + \int_2^x \frac{\#A(t)}{t^2}dt \\
&\ll \frac{x}{x\log x} + \int_2^x \frac{t}{t^2\log t}dt \\
&= \frac{1}{\log x} + \int_2^x \frac{1}{t\log t}dt \\
&= \frac{1}{\log x} - \log\log 2 + \log\log x \ll \left(\frac{1}{\log\log x\log x} + 1\right)\log\log x.
\end{aligned}$$

Taking the limit when x tends to infinity. We get

$$\sum_{\substack{2\leq n\leq x\\ n\in A(x)}} \frac{1}{n} \ll \log\log x.$$

□

Theorem 2.7. *We have the following equality, referred to at times as Mertens' formula (a very useful result).*

$$\prod_{p\leq x}\left(1-\frac{1}{p}\right)=\frac{e^{-\gamma}}{\log x}\left(1+O\left(\frac{1}{\log x}\right)\right)\quad(x\to\infty).$$

Proof. Let

$$U(x) \;=\; \prod_{p\leq x}\left(1-\frac{1}{p}\right).$$

Then

$$\begin{aligned}\log U(x) &= \log\prod_{p\leq x}\left(1-\frac{1}{p}\right)=\sum_{p\leq x}\log\left(1-\frac{1}{p}\right)\\ &= -\sum_{p\leq x}\frac{1}{p}+\sum_{p\leq x}\left(\log\left(1-\frac{1}{p}\right)+\frac{1}{p}\right)\ \text{(now apply Theorem 2.5)}\\ &= -\log\log x-\beta+O\left(\frac{1}{\log x}\right)+\sum_{p\leq x}\left(\log\left(1-\frac{1}{p}\right)+\frac{1}{p}\right).\end{aligned}$$

We now make use of the result

$$\left|\log\left(1-\frac{1}{p}\right)+\frac{1}{p}\right|\ll\frac{1}{p^2},$$

so let us sum through all these primes so we get

$$S=\sum_{p}\left(\log\left(1-\frac{1}{p}\right)+\frac{1}{p}\right).$$

This then results in

$$\begin{aligned}\sum_{p\leq x}\left(\log\left(1-\frac{1}{p}\right)+\frac{1}{p}\right) &= S-\sum_{p>x}\left(\log\left(1-\frac{1}{p}\right)+\frac{1}{p}\right)\\ &= S+O\left(\sum_{p>x}\frac{1}{p^2}\right)=S+O\left(\sum_{p>x}\frac{1}{n^2}\right)\\ &= S+O\left(\int_{\lfloor x\rfloor-1}^{\infty}\frac{dt}{t^2}\right)\\ &= S+O\left(\frac{1}{\lfloor x\rfloor-1}\right)=S+O\left(\frac{1}{x}\right).\end{aligned}$$

Combining these and then applying exponentiation, we will get the following

$$\log U(x) = -\log\log x - \beta + S + O\left(\frac{1}{\log x}\right),$$

now exponentiate

$$U(x) = e^{-\log\log x} e^{-\beta+S} e^{O(1/\log x)},$$

having $\beta - S = \gamma$, we can conclude the proof. □

2.3 The Prime Number Theorem

Definition 2.8. *Let $\pi(x)$ be the number of primes $p \leq x$ that is given by the following notation:*

$$\pi(x) = \sum_{p \leq x} 1.$$

Theorem 2.8 (Prime Number Theorem)**.** *As $x \to \infty$,*

$$\pi(x) \sim \frac{x}{\log x}.$$

The Li–function denoted $\mathrm{Li}(x)$, and defined as the following

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log t}.$$

The Li–function is a good approximation for $\pi(x)$. We have

$$\pi(x) \sim \mathrm{Li}(x), \quad \text{as } x \to \infty.$$

The French mathematician Legendre conjectured, in the year of 1798, a relatively accurate approximation for $\pi(x)$. It was conjectured that

$$\pi(x) \sim \frac{x}{A\log x + B}.$$

Later in 1808 he then proposed that $A = 1$ and $B = -1.08366$. It was later found, that he was not far off as it can be shown that $\pi(x) \sim x/(\log x - 1)$. It is also true that $\pi(x) \sim x/(\log x - 100)$.

A proof to show that $\pi(x) \sim x/(\log x - 1)$ will be given.

Theorem 2.9.

$$\pi(x) \sim \mathrm{Li}(x) \sim \frac{x}{\log x - 1}.$$

Proof. We first take

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log t},$$

integrating by parts we get:

$$\begin{aligned}
\int_2^x \frac{dt}{\log t} &= \frac{t}{\log t}\Big|_2^x + \int_2^x \frac{t}{t(\log t)^2}dt \\
&= \frac{x}{\log x} + O(1) + \frac{t}{(\log t)^2}\Big|_2^x + \int_2^x \frac{t(2\log t)}{t(\log t)^4}dt \\
&= \frac{x}{\log x} + \frac{x}{(\log x)^2} + O(1) + O\left(\int_2^{\sqrt{x}} \frac{1}{(\log t)^3}dt\right) \\
&\quad + O\left(\int_{\sqrt{x}}^x \frac{1}{(\log t)^3}dt\right)
\end{aligned}$$

In the smaller range $2 \le t \le \sqrt{x}$, we have that $\log t \ge \log 2$, so that

$$\int_2^{\sqrt{x}} \frac{1}{(\log t)^3}dt \le \frac{1}{(\log 2)^3}\int_2^{\sqrt{x}} dt \ll \sqrt{x} \ll \frac{x}{(\log x)^3}.$$

In the larger range $\sqrt{x} \le t \le x$, we have that $\log t \ge \log(\sqrt{x}) = (\log x)/2$, so that

$$\int_{\sqrt{x}}^x \frac{1}{(\log t)^3}dt \le \frac{8}{(\log x)^3}\int_{\sqrt{x}}^x dt < \frac{8x}{(\log x)^3} = O\left(\frac{x}{(\log x)^3}\right).$$

Hence, we have that,

$$\int_2^x \frac{dt}{\log t} = \frac{x}{\log x} + \frac{x}{(\log x)^2} + O\left(\frac{x}{(\log x)^3}\right).$$

Now we take

$$\begin{aligned}
\frac{x}{\log x - 1} &= \frac{x}{\log x}\left(\frac{1}{1 - (1/\log x)}\right) \\
&= \frac{x}{\log x}\left(1 + \frac{1}{\log x} + O\left(\frac{1}{(\log x)^2}\right)\right) \\
&= \frac{x}{\log x} + \frac{x}{(\log x)^2} + O\left(\frac{x}{(\log x)^3}\right).
\end{aligned}$$

Comparing the two results and making use of $\pi(x) \sim \mathrm{Li}(x)$, we conclude that $\mathrm{Li}(x)$ and $x/(\log x - 1)$ coincide at the first two terms. □

Some remarks in showing that $\pi(x) \sim x/(\log x - 1)$. The result holds for $\pi(x) \sim x/(\log x - k)$ with k as a constant s.t $1/(\log x - k) \ll 1/(\log x)$.

Chebyshev proved that if the function $\pi(x)/(x/\log x)$ has a limit as $x \to \infty$, then this limit has to be equal to 1. He also proved very accurate lower and upper bounds namely

$$c_1 \frac{x}{\log x} \leq \pi(x) \leq c_2 \frac{x}{\log x},$$

where $c_1 = 0.9212\ldots$ and $c_2 = 1.1055\ldots$

The year of 1896, Hadamard and De la Vallée Poussin independantly proved the Prime Number Theorem using complex analysis. They showed that

$$\lim_{x \to \infty} \frac{\pi(x)}{x/\log x} = 1.$$

2.4 The Divisor function

Definition 2.9. *Let $\tau(n)$ be the number of divisors of n. Then,*

$$\tau(n) = \sum_{d|n} 1.$$

Theorem 2.10.

$$\sum_{m=1}^{n} \tau(m) = \sum_{m=1}^{n} \left\lfloor \frac{n}{m} \right\rfloor = n \log n + (2\gamma - 1)n + O(n^{1/2}).$$

Proof. Let D_n be the region in the upper right hand quadrant not containing the x and y axis and under (and including) the hyperbola $xy = n$. That is

$$D_n = \{(x, y) \in \mathbb{R}^2 : x > 0, y > 0, xy \leq n\}.$$

Every point with integer coordinates in D_n is contained in a hyperbola $xy = m$ for some postive integer $m \leq n$. Thus, $\sum_{m=1}^{n} \tau(m)$ is the number of points with integer coordinates in D_n. Note that if $x = m$ is fixed, then the number of points in D_n having this value of x is $\left\lfloor \frac{n}{m} \right\rfloor$. Hence, the number of points in D_n is also $\sum_{m=1}^{n} \left\lfloor \frac{n}{m} \right\rfloor$.

To estimate how many points are in D_n, we first observe that the number of points above the line $y = x$ is the same as the number of points below it. Thus,

$$\begin{aligned}
\sum_{m=1}^{n} \left\lfloor \frac{n}{m} \right\rfloor &= 2 \sum_{x=1}^{\lfloor \sqrt{n} \rfloor} \left(\left\lfloor \frac{n}{x} \right\rfloor - \lfloor x \rfloor \right) + \lfloor \sqrt{n} \rfloor \\
&= 2 \sum_{k=1}^{\lfloor \sqrt{n} \rfloor} \left(\frac{n}{k} - k + O(1) \right) + \lfloor \sqrt{n} \rfloor \\
&= \left(2n \sum_{k=1}^{\lfloor \sqrt{n} \rfloor} \frac{1}{k} \right) - 2 \left(\frac{\lfloor \sqrt{n} \rfloor (\lfloor \sqrt{n} \rfloor + 1)}{2} \right) + O(\sqrt{n}).
\end{aligned}$$

We apply Theorem 2.1 to the sum in the above step,

$$\sum_{k=1}^{\lfloor \sqrt{n} \rfloor} \frac{1}{k} = \log \lfloor \sqrt{n} \rfloor + \gamma + O\left(\frac{1}{\sqrt{n}} \right).$$

Plugging this back into our original workings we get,

$$\begin{aligned}\sum_{m=1}^{n}\left\lfloor\frac{n}{m}\right\rfloor &= \left(2n\sum_{k=1}^{\lfloor\sqrt{n}\rfloor}\frac{1}{k}\right)-2\left(\frac{\lfloor\sqrt{n}\rfloor(\lfloor\sqrt{n}\rfloor+1)}{2}\right)+O(\sqrt{n})\\ &= 2n\left(\log\lfloor\sqrt{n}\rfloor+\gamma+O\left(\frac{1}{\sqrt{n}}\right)\right)-(n+\lfloor\sqrt{n}\rfloor)+O(\sqrt{n}).\\ &= 2n\left(\log\lfloor\sqrt{n}\rfloor+\gamma+O\left(\frac{1}{\sqrt{n}}\right)\right)-(n+O(\sqrt{n}))+O(\sqrt{n}).\end{aligned}$$

We rework the $\log\lfloor\sqrt{n}\rfloor$.

$$\begin{aligned}\log(\lfloor\sqrt{n}\rfloor) &= \log(\sqrt{n}+O(1))\\ &= \log\left(\sqrt{n}\left(1+O\left(\frac{1}{\sqrt{n}}\right)\right)\right)\\ &= \log(\sqrt{n})+\log\left(1+O\left(\frac{1}{\sqrt{n}}\right)\right)\\ &= \log(\sqrt{n})+O\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Plugging this back into

$$\sum_{m=1}^{n}\left\lfloor\frac{n}{m}\right\rfloor = 2n\left(\log\lfloor\sqrt{n}\rfloor+\gamma+O\left(\frac{1}{\sqrt{n}}\right)\right)-(n+O(\sqrt{n}))+O(\sqrt{n}).$$

We get

$$\begin{aligned}\sum_{m=1}^{n}\left\lfloor\frac{n}{m}\right\rfloor &= 2n\left(\log(\sqrt{n})+O\left(\frac{1}{\sqrt{n}}\right)+\gamma+O\left(\frac{1}{\sqrt{n}}\right)\right)-(n+O(\sqrt{n}))\\ &= n\log n+(2\gamma-1)n+O(\sqrt{n}).\end{aligned}$$

Thus concluding the proof. □

Definition 2.10. *Let $\omega(n)$ be the number of distinct prime factors of n.*

$$\omega(n) = \sum_{p|n} 1.$$

Definition 2.11. *Let $\sigma(n)$ be the function which sums the divisors of n.*

$$\sigma(n) = \sum_{d|n} d.$$

Note 2.3. *The maximal growth rate of σ can be expressed by*

$$\limsup_{n\to\infty} \left(\frac{\sigma(n)}{n \log\log n} \right) = e^{\gamma}.$$

Thus, we have

$$\frac{\sigma(n)}{n} \ll \log\log n,$$

as $n \to \infty$ [11].

Proof. Let $n \in \mathbb{Z}^+$ and n be composite, by the Fundamental Theorem of Arithmetic we can write $n = p_1^{k_1} p_1^{k_2} \cdots p_1^{k_m}$. We get

$$\begin{aligned} \frac{\sigma(n)}{n} &= \frac{(1 + p_1 + \cdots + p_1^{k_1}) \cdots (1 + p_m + \cdots + p_m^{k_m})}{p_1^{k_1} \cdots p_k^{k_m}} \\ &= \sum_{0 \le j_1 \le k_1} \frac{1}{p_1^{j_1}} \cdots \sum_{0 \le j_m \le k_m} \frac{1}{p_1^{j_m}}. \end{aligned}$$

If we fix p such that $p \mid n$ we have that

$$1 + \frac{1}{p} + \cdots + \frac{1}{p^k} < \sum_{i \ge 0} \frac{1}{p^i} = \left(1 - \frac{1}{p}\right)^{-1}.$$

Thus, we get

$$\frac{\sigma(n)}{n} \le \prod_{p|n} \left(1 - \frac{1}{p}\right)^{-1}. \tag{2.1}$$

Let $m = \omega(n)$ from Definition 2.10. We have $2^m \le p_1 \cdots p_m \le n$. Applying logarithms, we get

$$m \le \frac{\log n}{\log 2} < 2 \log n.$$

Thus, n has at most $2\log n$ prime factors. The right hand side of (2.1)

$$\prod_{i=1}^{m}\left(1-\frac{1}{p_i}\right)^{-1} \tag{2.2}$$

is maximal when we replace $p_1, ..., p_m$ by the first m primes (but keep the value for m). By the Prime Number Theorem (Theorem 2.8),

$$\frac{p_m}{\log p_m} \ll \pi(p_m) = m.$$

We can now write

$$\log p_m + O(1) \leq \log m + \log\log p_m.$$

Since $\log\log p_m = o(\log p_m)$, it follows that $\log m \geq (1+o(1))\log p_m$. In particular, $\log m \geq (1/2)\log p_m$ so $p_m \leq m^2$ for sufficiently large enough m. We want to see what the most number of primes that can be counted in a somewhat accurate way.
We can obtain something slightly stronger than $p_m \leq m^2$ namely

$$p_m \ll m\log p_m \ll m\log m^2 \ll m\log m.$$

Since $m \ll \log n$, we have $p_m < (\log n)^2$ holds for sufficiently large n.
Applying Theorem 2.7 to (2.2) we get

$$\prod_{i=1}^{k}\left(1-\frac{1}{p_i}\right)^{-1} \leq \prod_{p\leq(\log n)^2}\left(1-\frac{1}{p}\right)^{-1} \ll \log((\log n)^2) \ll \log\log n.$$

Thus, we conclude the proof. □

2.5 Sieves

In this section, we make use of the inclusion-exclusion technique. Let us look at a basic sieve, the Sieve of Eratosthenes. Say we take the list of numbers from 1 to 100. Taking $p \leq \sqrt{100}$, we can then remove (sieve out) any $n <= 100$ s.t $p \mid n$ where $p \leq \sqrt{100}$. The numbers which were not removed will be the prime numbers $p > \sqrt{100}$.

Some quick results shall be mentioned to aid, with the mission of expressing the sieve, in a more desirable form.

We let P denote the product of primes less than x :

$$P = \prod_{p \leq x} p.$$

Definition 2.12. *The Möbius function is defined by*

$$\mu(n) = \begin{cases} 1, & \text{if } n = 1, \\ 0, & \text{if } p^2 \mid n \text{ for some prime } p, \\ (-1)^{\omega(n)}, & \text{otherwise.} \end{cases}$$

Theorem 2.11. *If*

$$E(n) = \begin{cases} 1, & \text{if } n = 1, \\ 0, & \text{otherwise,} \end{cases}$$

then

$$\sum_{d \mid n} \mu(d) = E(n).$$

Proof. If $n = 1$, then condition holds. Asumme that $n > 1$.
Let $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, and $p_i \neq p_j$, for any $i \neq j$, and where α_i, are positive integers. Let $m = p_1 \cdots p_r$. Then it is clear that

$$\sum_{d \mid n} \mu(d) = \sum_{d \mid m} \mu(d).$$

Since for each integer $j \in \{1, \ldots, r\}$, there are $\binom{r}{j}$ divisors of m with j prime factors, we have

$$\sum_{d \mid m} \mu(d) = 1 - \binom{r}{1} + \binom{r}{2} - \cdots + (-1)^r \binom{r}{r} = (1-1)^r = 0.$$

Thus we conclude the proof. □

If x is any positive real number, then an integer n, in the interval $(\sqrt{x}, x)$ is prime if and only if $\gcd(n, P) = 1$. Thus we can write

$$\begin{aligned}\pi(x) - \pi(\sqrt{x}) + 1 &= \sum_{n \le x} E((n, P)) = \sum_{n \le x} \sum_{\substack{d \mid n \\ d \mid P}} \mu(d) \\ &= \sum_{d \mid P} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor . \end{aligned} \tag{2.3}$$

We have

$$\left\lfloor \frac{x}{d} \right\rfloor = \frac{x}{d} + O(1),$$

and plugging the above into equation 2.3, we obtain

$$\begin{aligned}\pi(x) - \pi(\sqrt{x}) + 1 &= \sum_{d \mid P} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor \\ &= \sum_{d \mid P} \mu(d) \left(\frac{x}{d} + O(1) \right) \\ &= \sum_{d \mid P} \left(\mu(d) \frac{x}{d} + O(\mu(d)) \right) .\end{aligned}$$

We end up counting an $O(1)$, every time $d \mid P$, and since P has $\pi(\sqrt{x})$ distinct prime factors, we will have the following amount of divisors

$$\binom{\pi(\sqrt{x})}{0} + \binom{\pi(\sqrt{x})}{1} + \cdots + \binom{\pi(\sqrt{x})}{\pi(\sqrt{x})} = 2^{\pi(\sqrt{x})}.$$

Therefore we are left with,

$$\begin{aligned}\pi(x) - \pi(\sqrt{x}) + 1 &= \sum_{d \mid P} \left(\mu(d) \frac{x}{d} + O(1) \right) \\ &= x \prod_{p \le \sqrt{x}} \left(1 - \frac{1}{p} \right) + O(2^{\pi(\sqrt{x})}).\end{aligned}$$

Thus we will end up with an error term which grows exponentially and is larger than the main term, which is not the most pleasing sight.

Definition 2.13. *A function $f : \mathbb{N} \to \mathbb{C}$ is said to be multiplicative if $f(1) = 1$ and $f(nm) = f(n)f(m)$ for all postitive integers n and m such that* $\gcd(n, m) = 1$.

Definition 2.14. *Let $a, b \in \mathbb{Z}$. We say that $n \equiv b \pmod{a}$ if there exists $k \in \mathbb{Z}$ such that $n = ak + b$.*

Let $\mathcal{A} \subseteq \{1, \ldots, x\}$. For a prime p we write

$$\mathcal{A}_p = \{a \in \mathcal{A},\ a \equiv 0 \pmod{p}\}.$$

In the case of the Sieve of Eratosthenes we can clearly see that we had $\mathcal{A} = \{1, \ldots, x\}$ and we sieved out

$$\mathcal{A}_p = \{a \in \mathcal{A},\ a \equiv 0 \pmod{p} \mid p \leq \sqrt{x}\}.$$

We can generalize this notation, and work with numbers in the following set

$$\mathcal{A}_d = \{a \in \mathcal{A},\ a \equiv 0 \pmod{d}\}.$$

We use $\rho(p)$ to denote the number of distinct residue classes modulo p we wish to sieve out. In the case of the Sieve of Eratosthenes, we had 1 distinct residue class modulo p namely $0 \pmod{p}$, and thus $\rho(p) = 1$. So, we can now generalize the Sieve of Eratosthenes and write the following:

$$|\mathcal{A}_p| = x\frac{v(p)}{p} + O(v(p)).$$

Assume that $v(d)$ as a multiplicative function when d is square-free (We will show that $v(d)$ is multiplicative in the proof of Proposition 3). For d square-free $\mathcal{A}_d = \cap_{p|d}\mathcal{A}_p$. Then, putting $v(d) = \prod_{p|d} v(p)$ we have

$$|\mathcal{A}_d| = x\frac{v(d)}{d} + O(v(d)).$$

Furthermore, we denote $P(y) = \prod_{p \leq y} p$, and we will look at

$$S(\mathcal{A}, P, y) \quad = \quad \left| \mathcal{A} \setminus \bigcup_{p \in P(y)} \mathcal{A}_p \right|,$$

as the number of elements in $\mathcal{A}$ that are not in any of the distinguished residue classes modulo p.

We now want to expand

$$\begin{aligned} S(\mathcal{A}, P, y) &= \sum_{d|P(y)} \mu(d)|\mathcal{A}_d| \\ &= \sum_{d|P(y)} \mu(d) \left(x\frac{v(d)}{d} + O(v(d)) \right). \end{aligned}$$

The Brun sieve

The Brun pure sieve is a technique for estimating the cardinality of sets where we have applied a sieve on the positive integers, which satisfy a set of congruences. Here are some results.

Proposition 2. *Let $f, g : \mathbb{N} \to \mathbb{C}$ be two multiplicative functions. Define*

$$(f * g)(n) = \sum_{d|n} f(d)g(n/d).$$

Then $f * g$ is multiplicative.

Definition 2.15. *Let $f : \mathbb{N} \to \mathbb{N}$ then we denote the kernel of f to be*

$$\ker(f) = \{x \in \mathbb{N} | f(x) = 1_{\mathbb{N}}\},$$

where $1_{\mathbb{N}}$ is the identity element of $\mathbb{N}$. (Under multiplication the indentity element is 1).

The Sieve of Eratosthenes satisfies the identity

$$\mu * 1 = E,$$

where E is the characteristic function mentioned in Theorem 2.11. Brun's idea was to introduce two auxiliary functions μ_1 and μ_2 satisfying

$$\mu_1 * 1 \leq E \leq \mu_2 * 1$$

and vanishing often enough so that the number of non-zero terms in the resulting formula analogous with

$$\sum_{d|P(\sqrt{x})} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor,$$

is not overwhelming. Brun's initial choice led to what is now called Brun's pure sieve and is comprised of the following.

Theorem 2.12. *Denote by χ_t, the characteristic function of the set of integers n such that $\omega(n) \leq t$. Then for each integer $h \geq 0$, the functions defined by*

$$\mu_i(n) = \mu(n)\chi_{2h+2-i}(n) \quad (i = 1, 2)$$

satisfy the inequalities $\mu_1 * 1 \leq E \leq \mu_2 * 1$.

Proof. Since $\mu_i * 1(n)$ depends only on the kernel of n which is the product of the distinct primes dividing n, we may assume that $\mu(n) \neq 0$. If $\omega(n) = k$, then, for each r with $0 \leq r \leq k$, it is clear that n has exactly $\binom{k}{r}$ divisors d with $\omega(d) = r$. For any given $t \geq 0$, we can thus write

$$\chi_t * 1(n) = \sum_{\substack{d|n \\ \omega(d)\leq t}} \mu(d) = \sum_{0\leq r\leq t} (-1)^r \binom{k}{r}.$$

Finally, we have that

$$\sum_{0\leq r\leq t} (-1)^r \binom{k}{r} = (-1)^t \binom{k-1}{t}.$$

We shall prove this by induction on t. For $t = 0$ this is obviously true. Now we assume for $t \geq 0$, and we test for $t + 1$.

$$\begin{aligned}
\sum_{0\leq r\leq t+1} (-1)^r \binom{k}{r} &= \sum_{0\leq r\leq t} (-1)^r \binom{k}{r} + (-1)^{t+1} \binom{k}{t+1} \\
&= (-1)^t \binom{k-1}{t} + (-1)^{t+1} \binom{k}{t+1} \\
&= \frac{(-1)^t (k-1)!}{(k-1-t)!t!} + \frac{(-1)^{t+1} k!}{(k-t-1)!(t+1)!} \\
&= (-1)^t (k-1)! \frac{(t+1)-k}{(k-1-t)!(t+1)!} \\
&= (-1)^{t+1} (k-1)! \frac{(-t-1)+k}{(k-1-t)!(t+1)!} \\
&= \frac{(-1)^{t+1}(k-1)!}{(k-1-(t+1))!(t+1)!} \\
&= (-1)^{t+1} \binom{k-1}{t+1},
\end{aligned}$$

thus concluding the proof. □

Corollary 2.1. *Let $\mathcal{A}$ be a finite set of integers and let $\mathcal{P}$ be a set of prime numbers. Write*

$$\begin{aligned} \mathcal{A}_d &= \#\{a \in \mathcal{A} : a \equiv 0 \pmod{d}\}, \\ \mathcal{P}(y) &= \prod_{\substack{p \le y \\ p \in \mathcal{P}}} p, \\ \mathcal{S}(\mathcal{A}, \mathcal{P}, y) &= \#\{a \in \mathcal{A} : (a, P(y)) = 1\}. \end{aligned}$$

Then, for each integer $h \ge 0$,

$$\sum_{\substack{d \mid \mathcal{P}(y) \\ \omega(d) \le 2h+1}} \mu(d)\mathcal{A}_d \le \mathcal{S}(\mathcal{A}, \mathcal{P}, y) \le \sum_{\substack{d \mid \mathcal{P}(y) \\ \omega(d) \le 2h}} \mu(d)\mathcal{A}_d.$$

We now use the above result to obtain an upperbound for $\pi(x)$.

We shall begin by letting $\mathcal{A} = \{n : n \le x\}$, $\mathcal{P} = \{\text{p prime}\}$ and $P = P(y) = \prod_{p \le y} p$. Then $\mathcal{S}(\mathcal{A}, \mathcal{P}, y)$ is the number of positive integers $n \le x$ having no prime factor $p \le y$, so that

$$\begin{aligned} \mathcal{S}(\mathcal{A}, \mathcal{P}, y) &\le \sum_{\substack{d \mid P \\ \omega(d) \le 2h}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor = x \sum_{\substack{d \mid P \\ \omega(d) \le 2h}} \frac{\mu(d)}{d} + O\left(\sum_{\substack{d \mid P \\ \omega(d) \le 2h}} 1 \right) \\ &= x \prod_{p \le y} \left(1 - \frac{1}{p}\right) + O\left(\sum_{\substack{d \mid P \\ \omega(d) \le 2h}} 1 + x \sum_{\substack{d \mid P \\ \omega(d) > 2h}} \frac{1}{d} \right), \end{aligned} \tag{2.4}$$

and similarly

$$\begin{aligned} \mathcal{S}(\mathcal{A}, \mathcal{P}, y) &\ge \sum_{\substack{d \mid P \\ \omega(d) \le 2h+1}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor = x \sum_{\substack{d \mid P \\ \omega(d) \le 2h+1}} \frac{\mu(d)}{d} + O\left(\sum_{\substack{d \mid P \\ \omega(d) \le 2h+1}} 1 \right) \\ &= x \prod_{p \le y} \left(1 - \frac{1}{p}\right) + O\left(\sum_{\substack{d \mid P \\ \omega(d) \le 2h+1}} 1 + x \sum_{\substack{d \mid P \\ \omega(d) > 2h+1}} \frac{1}{d} \right). \end{aligned} \tag{2.5}$$

The first two error terms in (2.4) and (2.5) do not exceed y^{2h+1} since this is an upperbound for all integers d such that $d \mid P$ and $\omega(d) \leq 2h+1$. The sums over d that appear in the second error term are bounded by the following

$$\sum_{\substack{d|P \\ \omega(d)>2h+1}} \frac{1}{d} \leq \sum_{k>2h} \frac{1}{k!}(\log\log y + c_0)^k.$$

We make use of the weak form of Stirling's formula

$$n! > \left(\frac{n}{e}\right)^n.$$

We also make use of the fact that $y < x$, and get that

$$\sum_{\substack{d|P \\ \omega(d)>2h+1}} \frac{1}{d} \leq \sum_{k>2h} \frac{1}{k!}(\log\log x + c_0)^k \leq \sum_{k>2h} \left(\frac{e(\log\log x + c_0)}{k}\right)^k.$$

Choosing the smallest integer $h \geq e(\log\log x + c_0)$, we obtain

$$\sum_{\substack{d|P \\ \omega(d)>2h+1}} \frac{1}{d} \leq \left(\frac{1}{2}\right)^{2h} \sum_{j\geq 0} \left(\frac{1}{2}\right)^j \ll \frac{1}{(\log x)^{2e\log 2}} \ll \frac{1}{(\log x)^2},$$

because $2e\log 2 = e\log 4 > e > 2$. For this choice of h, we force $y^{2h+1} \leq x/(\log x)^2$, which for $h > 1$ is implied by

$$y \leq \frac{x^{1/(2h+1)}}{\log x} \leq \exp\left(\frac{\log x}{2e\log\log x + c_1} - \log\log x\right),$$

where we can take $c_1 = 2ec_0 + 1$. Since $1/2e > 1/10$, it follows that we may choose

$$y = \exp\left(\frac{\log x}{10\log\log x}\right),$$

in which case the inequality $y^{2h+1} \ll x/(\log x)^2$ holds for all x. With this choice of y, we have that

$$\prod_{p\leq y}\left(1 - \frac{1}{p}\right) \asymp \frac{1}{\log y} \asymp \frac{\log\log x}{\log x},$$

while the error terms in (2.4) and (2.5) are $O(x/(\log x)^2)$. Thus, we obtain the following

$$\mathcal{S}(\mathcal{A}, \mathcal{P}, y) = x\prod_{p\leq y}\left(1 - \frac{1}{p}\right)\left(1 + O\left(\frac{1}{\log y}\right)\right).$$

Since,

$$\mathcal{S}(\mathcal{A}, \mathcal{P}, y) \geq \pi(x) - \pi(y) \geq \pi(x) - y \geq \pi(x) + O(x^{1/2}),$$

we can deduce that

$$\pi(x) \ll x^{1/2} + x \prod_{p \leq y} \left(1 - \frac{1}{p}\right) \ll \frac{x \log\log x}{\log x}.$$

We notice this is quite an improvement compared to the upperbound of $\pi(x)$ obtained by the Eratosthenes sieve.

We will demonstrate the use of Brun's pure sieve again with the mention of a few powerful results which will aid us in the quest of proving the desired result whose proof makes use of Brun's pure sieve.

We want to count primes such that they are of the form $p' = p + h$ where $0 < h < d$, and h is a fixed even number. We look at the set

$$\mathcal{A} = \{n(n+h) : n \leq x\}.$$

For square-free d we have

$$v(d) = \#\{0 \leq n \leq d-1 \quad n(n+h) \equiv 0 \pmod{d}\}.$$

We see that for $q > 2$ we have 2 solutions to the congruence $n(n+h) \equiv 0 \pmod{p}$, namely $n = 0$ and $n = q - h$ except if $q \mid h$, in which case $v(q) = 1$. Furthermore, $v(2) = 1$. So if d is square-free and is odd $v(d) = 2^{\omega(d)}$, and if d is even $v(d) = 2^{\omega(d)-1}$. Now we have to sieve out the following residue classes $0 \pmod{p}$ and $-h \pmod{p}$.

Theorem 2.13 (Chinese Remainder Theorem)**.** *Let $m_1, \ldots, m_k$ be positive integers with $\gcd(m_i, m_j) = 1$ for all $1 \leq i < j \leq k$. Let $a_1, \ldots, a_k$ be arbitary integers. Then there is an integer a such that*

$$a \equiv a_i \pmod{m_i} \quad \textit{for all } i = 1, \ldots, k.$$

Theorem 2.14. *For each real number κ, there exists a positive constant c_k such that*

$$\prod_{p \leq x} \left(1 + \frac{\kappa}{p}\right) = c_k (\log x)^{\kappa} \left(1 + O\left(\frac{1}{\log x}\right)\right) \quad (x \to \infty).$$

Definition 2.16. *If p and q both be primes such that $q = p + 2$. Then we say p and q are a pair of twin primes.*

The Twin Prime Conjecture states that there are infinitely many twin primes. Hardy and Littlewood 'conjectured' the following:
Let $\pi_2(x)$ denote the number of primes $p \leq x$ such that $p + 2$ is also prime. Define the twin prime constant c_2 as

$$c_2 = \prod_{p \geq 3} \left(1 - \frac{1}{(p-1)^2}\right) \approx 0.660161\ldots$$

Then a special case of their Conjecture is

$$\pi_2(x) \sim 2c_2 \frac{x}{(\log x)^2} \sim 2c_2 \int_2^x \frac{dt}{(\log t)^2} \quad [4].$$

This is known as the "First Hardy and Littlewood Conjecture." We can see if this were proved, then we would already have a better lower bound to Theorem 1. Now finally the next result which makes use of Brun's sieve.

Proposition 3. *Let $\mathcal{J} = \{p : p \text{ and } p + 2 \text{ are both primes}\}$ and set $\mathcal{J} = \#\{p \leq : p \in \mathcal{J}\}$. Then*

$$\mathcal{J} \ll \frac{x(\log\log x)^2}{(\log x)^2}.$$

Proof. We set $\mathcal{A} = \{n(n+2) : n \leq x\}$. We also recall $P(y) = P = \prod_{p \leq y} p$, we shall choose our y at a later stage. Understand the following number of residues classes satisfying a given congruence:

$$\rho(d) = \#\{0 \leq n \leq d-1 : \ n(n+2) \equiv 0 \pmod{d}\}.$$

It will first be shown that $\rho(d)$ is multiplicative. If u and v are coprime and $c \pmod{uv}$ is some congruence class modulo uv such that $n(n+2) \equiv 0 \pmod{uv}$, then certainly $c \pmod{u}$ ($c \pmod{v}$, respectively) is a congruence class modulo u (modulo v, respectively). Conversely, if $a \pmod{u}$ and $b \pmod{v}$ are congruence classes for n modulo u and v which are solutions to $n(n+2) \equiv 0 \pmod{u}$ and $n(n+2) \equiv 0 \pmod{v}$, respectively, then by the Chinese Remainder Theorem (Theorem 2.13), there exists a class $c \pmod{uv}$ (which is unique) such that $c \equiv a \pmod{u}$ and $c \equiv b \pmod{v}$. Hence, $n(n+2) \equiv 0 \pmod{u}$ and $n(n+2) \equiv 0 \pmod{v}$, and since u and v are coprime, we get that $n(n+2) \equiv 0 \pmod{uv}$. This shows that

$\rho(uv) = \rho(u)\rho(v)$, therefore satisfying the properties of being multiplicative. As mentioned before $\rho(2) = 1$, but in addition to this $\rho(q) = 2$ for all odd $q > 2$ since $h = 2$ so it has no odd prime factors. Since there are precisely $\rho(d)$ solutions to n the congruence $n(n+2) \equiv 0 \pmod{d}$ in any interval of length d, and since the interval $[1, x]$ is made up of $\lfloor x/d \rfloor$ intervals of length d and (maybe) one shorter interval, we now get that

$$\begin{aligned} \mathcal{A}_d &= \#\{n \le x : d \mid n(n+2)\} = \rho(d)\left(\left\lfloor \frac{x}{d} \right\rfloor + O(1)\right) \\ &= \frac{x\rho(d)}{d} + O(\rho(d)) = \frac{x\rho(d)}{d} + O(2^{\omega(d)}). \end{aligned}$$

Note that if $p, p+2$ are twin primes, then either $p \le y$ or $p \in \mathcal{S}(\mathcal{A}, \mathcal{P}, y)$, we have by Corollary 2.1, that

$$\begin{aligned} \mathcal{J}(x) &\le \pi(y) + \sum_{\substack{d \mid P \\ \omega(d) \le 2h}} \mu(d)\mathcal{A}_d \\ &= \sum_{\substack{d \mid P \\ \omega(d) \le 2h}} \mu(d)\left(\frac{x\rho(d)}{d} + O(2^{\omega(d)})\right) + O(y) \\ &= x \sum_{\substack{d \mid P \\ \omega(d) \le 2h}} \frac{\mu(d)\rho(d)}{d} + O\left(y + \sum_{\substack{d \mid P \\ \omega(d) \le 2h}} 2^{\omega(d)}\right) \\ &= x \sum_{\substack{d \mid P \\ \omega(d) \le 2h}} \frac{\mu(d)\rho(d)}{d} + O\left(y + 2^{2h} \sum_{\substack{d \mid P \\ \omega(d) \le 2h}} 1 + x \sum_{\substack{d \mid P \\ \omega(d) > 2h}} \frac{2^{\omega(d)}}{d}\right) \\ &= x \prod_{p \le y}\left(1 - \frac{\rho(p)}{p}\right) + O\left(y + 2^{2h} \sum_{\substack{d \mid P \\ \omega(d) \le 2h}} 1 + x \sum_{\substack{d \mid P \\ \omega(d) > 2h}} \frac{2^{\omega(d)}}{d}\right) \\ &= \frac{x}{2} \prod_{3 \le p \le y}\left(1 - \frac{2}{p}\right) + O\left(y + (2y)^{2h} + x \sum_{\substack{d \mid P \\ \omega(d) > 2h}} \frac{2^{\omega(d)}}{d}\right). \end{aligned}$$

Using the combinatorial fact that

$$\sum_{\substack{d|P\\ \omega(d)>2h}} \frac{2^{\omega(d)}}{d} \leq \sum_{k>2h} \sum_{\substack{d|P\\ \omega(d)=k}} \frac{2^k}{d} \leq \sum_{k>2h} \frac{1}{k!} \left(\sum_{p\leq y} \frac{2}{p} \right)^k,$$

together with Theorem 2.5, we get

$$\sum_{\substack{d|P\\ \omega(d)>2h}} \frac{2^{\omega(d)}}{d} < \sum_{k>2h} \frac{(2\log\log x + 2c_0)^k}{k!} < \sum_{k>2h} \left(\frac{2e\log\log x + c_1}{k} \right)^k,$$

where $c_1 = 2ec_0$. Hence, we see that if we choose $h = 2h_0$ where h_0 is the minimal positive integer with $h_0 > 2e\log\log x + c_1$. We then get

$$\sum_{\substack{d|P\\ \omega(d)>2h}} \frac{2^{\omega(d)}}{d} < \frac{1}{2^{2h}} \left(\sum_{l\geq 0} \frac{1}{2^l} \right) = 2^{1-2h} \ll \frac{1}{(\log x)^{4e\log 2}} < \frac{1}{(\log x)^2}. \quad (2.6)$$

Choosing $y = \exp(\log x/(20\log\log x))$, we obtain that

$$(2y)^{2h} = 2^{2h} \exp\left(\frac{2h\log x}{20\log\log x} \right)$$

$$= \exp\left(\frac{(1+o(1))4e\log x}{20} \right) \ll \frac{x}{(\log x)^2}. \quad (2.7)$$

Inserting Equations 2.6 and 2.7 into our initial estimate, we get that

$$\mathcal{J}(x) \ll x \prod_{3\leq p\leq y} \left(1 - \frac{2}{p}\right) + \frac{x}{(\log x)^2}.$$

Finally we use Theorem 2.14 with $\kappa = -2$, we have that

$$\begin{aligned} \prod_{3\leq p\leq y} \left(1 - \frac{2}{p}\right) &= \frac{c_2}{(\log y)^2}(1+o(1)) = c_2 \left(\frac{20\log\log x}{\log x} \right)^2 (1+o(1)) \\ &= (1+o(1)) \frac{400c_2(\log\log x)^2}{(\log x)^2}, \end{aligned}$$

so that finally we get

$$\mathcal{J}(x) \ll \frac{x(\log\log x)^2}{(\log x)^2},$$

thus concluding the proof. □

This sieve was later refined by partitioning the interval $[1, y]$ into suitable subintervals $[y_j, y_{j+1}]$, where $1 = y_0 < y_1 < \ldots < y_k = y$ and selecting for $i = 1, 2$, $\mu_i(d) = \mu(d)\chi_i^*(d)$, where $\chi_i^*(d)$ is the characteristic function of the set of those positive integers n having exactly $2h_j + 2 - i$ prime factors in $[y_j, y_{j+1})$ for $j = 0, 1, \ldots, k-1$.

The following are basic results from the Brun combinatorial sieve.

Theorem 2.15. *Assume that there exists a non-negative multiplicative function w, some real number X (Where we say $\#\mathcal{A} = X$) and some positive constants κ and $\mathcal{A}$ such that*

$$(i) \quad \mathcal{A}_d = X\frac{v(d)}{d} + R_d \qquad (d \mid P(y)),$$

$$(ii) \quad \prod_{\eta \le p \le \zeta} \left(1 - \frac{v(p)}{p}\right)^{-1} < \left(\frac{\log \zeta}{\log \eta}\right)^{\kappa} \left(1 + \frac{\mathcal{A}}{\log \eta}\right) \quad (2 \le \eta \le \zeta).$$

Then uniformly for $\mathcal{A}, X, y, u \ge 1$, where $u = \dfrac{\log x}{\log y}$.

$$\mathcal{S}(\mathcal{A}, \mathcal{P}, y) \quad = \quad X \prod_{\substack{p \le y \\ p \in \mathcal{P}}} \left(1 - \frac{v(p)}{p}\right)^{-1} \{1 + O(u^{-u/2})\} + O\left(\sum_{\substack{d \le y^u \\ d \mid P(y)}} |R_d|\right).$$

Definition 2.17. *The Euler Phi function denoted as $\phi(n) = \#\{1 \le m \le n : (m, n) = 1\}$, counts all the positive integers $m \le n$ that are relatively prime (coprime) to n.*

One of the few known applications of the Euler Phi function and prime numbers can be explored in the field of Cryptography. The importance of having $\gcd(a, n) = 1$ implies that a has an inverse *mod* n. This will imply that if one applies $a \pmod n$ to encrypt a message then due to the fact that there exists an inverse, one can then decrypt (reverse the transformation) to obtain the original message. The Euler Phi function will tell you the number of possible keys that one can have to encrypt a message with such that the original message can be recovered. We also note that $\phi(p) = p - 1$.

Note 2.4. *Let $\pi(x; a, b) = \#\{p \le x : p \equiv b \pmod a\}$.*

Theorem 2.16. *Brun-Titchmarsh theorem: Let $1 \le a < b$ be integers with $\gcd(a, b) = 1$. Then the inequality*

$$\pi(x; a, b) - \pi(x - z; a, b) \ll \frac{z}{\phi(a) \log(z/a)},$$

holds uniformly for $1 \le a < z \le x$.

We make use of Mertens' formula for the following note: Let us denote $P(m)$, as the highest prime factor of m.

Note 2.5.

$$\sum_{\substack{m\leq x\\ P(m)\leq w}} \frac{1}{m} \ll \log w.$$

We add more terms to the sum, in such a way as to sum over all the numbers where the highest prime factor $P(m) \leq w$. We can sum all the numbers which have this property and by the Fundamental Theorem of Arithmetic, we get that the only numbers which are counted have to have factors of $p \leq w$. Now we see that

$$\sum_{P(m)\leq w} \frac{1}{m} \leq \left(1 + \frac{1}{p_1} + \frac{1}{p_1^2} + \cdots\right) \cdots \left(1 + \frac{1}{p_k} + \frac{1}{p_k^2} + \cdots\right) = \prod_{p\leq w} \left(1 - \frac{1}{p}\right)^{-1}.$$

Applying Mertens' formula (Theorem 2.7). We have that

$$\prod_{p\leq w} \left(1 - \frac{1}{p}\right)^{-1} \ll \log w.$$

2.6 The Bateman–Horn Conjecture

Definition 2.18. *We say that a system of non-constant polynomials $f_1(X), \dots, f_k(X) \in \mathbb{Z}[X]$ each having positive leading coefficient is acceptable if the following conditions hold:*

(i) $f_i(X)$ *is irreducible for all* $i = 1, \dots, k$.

(ii) there exists no prime number p *such that* $p \mid f_1(n)f_2(n)\cdots f_k(n)$ *for all* $n \geq 0$.

Given a system of acceptable polynomials, the Bateman–Horn Conjecture (see Section 2.10 in [7]) is a heuristic statement addressing the frequency of the positive integers n such that

$$f_1(n),\ f_2(n),\ \dots,\ f_k(n)$$

are all prime numbers. The actual statement is as follows.

Conjecture 2.1 (Bateman, Horn [3])**.** *Let $f_1(X), \dots, f_k(X)$ be an acceptable system of polynomials. For each prime number p let*

$$\omega(p) = \#\{0 \leq n \leq p-1 \ :\ f_1(n)f_2(n)\cdots f_k(n) \equiv 0 \pmod{p}\},$$

and set

$$\pi_{f_1,\dots,f_k}(x) := \#\{n \leq x \ :\ f_1(n),\ \dots,\ f_k(n) \text{ are all primes}\}.$$

Then

$$\pi_{f_1,\dots,f_k}(x) = (1+o(1))C(f_1,\dots,f_k)\,\frac{1}{d_1\cdots d_k}\,\frac{x}{(\log x)^k} \quad \text{as} \quad x \to \infty, \tag{2.8}$$

where $d_i := \deg(f_i)$, and where the constant $C(f_1, \dots, f_k)$ is given by

$$C(f_1,\dots,f_k) = \prod_{p \geq 2} \frac{1-\omega(p)/p}{(1-1/p)^k}. \tag{2.9}$$

2.7 Chudakov

In this section we refer to the following reference [2].

Let a and b be two-dimensional integer vectors and $\pi(x;a,b)$ be the number of positive integers n such that $an+b \leq x$. The probability that a_1n+b_1 and a_2n+b_2 are both prime is

$$\frac{1}{\log(a_1n+b_1)}\frac{1}{\log(a_2n+b_2)}.$$

So,we would now like to count through all of these numbers of the form $an+b \leq x$. Let

$$T(x;a,b) = \sum_{1<an+b\leq x} \frac{1}{\log(a_1n+b_1)(\log(a_2n+b_2)}.$$

We now let h, be an even positive integer and we set $(a_1,a_2)=(1,1)$ and $(b_1,b_2)=(0,h)$. Having done this, we shall get the following:

$$T(x;(1,1),(0,h)) = T_h(x) = \sum_{n+h\leq x} \frac{1}{\log n \log(n+h)}.$$

Chudakov proved that almost all even integers $h \leq x$ are the difference of two primes. The equation for $T_h(x)$ is the expected order of magnitude of $\pi(x,h)$, and

$$T_h(x) \sim c_h \frac{x}{(\log x)^2},$$

as $x \to \infty$. In fact we take note that $T_h(x) \asymp x/(\log x)^2$ uniformly for $h = o(x)$ as $x \to \infty$. We denote

$$c_h = 2\prod_{p\neq 2} \frac{p(p-2)}{(p-1)^2} \prod_{\substack{p|d \\ p\neq 2}} \frac{p-1}{p-2}.$$

Chudakov then proved that

$$\sum_{2|h} \left|\pi(x,h) - 2\prod_{p\neq 2} \frac{p(p-2)}{(p-1)^2} \prod_{\substack{p|d \\ p\neq 2}} \frac{p-1}{p-2} T_h(x)\right|^2 \ll \frac{x^3}{(\log x)^A}$$

for any $A>0$. We observe that the sum over h is finite, as for large $h>x$, we have $\pi(x,h)=T_h(x)=0$. For most even integers $h \leq x$, the number of prime pairs $1<p<x$, $1<p<p+h<x$ is equal to the expected number. However this does not imply that $\pi(x,h)$ tends to infinity for some fixed h.

2.8 Smooth numbers and the Dickman–de Brujin function

Definition 2.19. *Let $P(n)$ be the largest prime factor of n. Then the number n is said to be y-smooth if $P(n) \leq y$.*
We use the following function to denote the counting of y-smooth numbers,

$$\psi(x, y) = \#\{n \leq x : P(n) \leq y\}.$$

Definition 2.20. *The Dickman–de Brujin function $\rho(u)$ is a continuous function that satisfies the delay differential equation*

$$u\rho'(u) + \rho(u-1) = 0 \qquad (u > 1)$$

with initial conditions $\rho(u) = 1$ for $0 \leq u \leq 1$.

In number theory the Dickman–de Brujin function is used to estimate the proportion of smooth numbers up to a given bound.

Theorem 2.17.

$$\psi(x, y) = \psi(x, z) - \sum_{y < p \leq z} \psi\left(\frac{x}{p}, p\right) \qquad (x \geq 1, z \geq y > 0).$$

Proof. We use the fact that

$$\psi(x, z) = \sum_{p \leq z} \sum_{\substack{n = pm \leq x \\ P(m) \leq p}} 1 = \sum_{p \leq z} \psi\left(\frac{x}{p}, p\right).$$

Now we see that we have to take out the multiples where there is a prime factor bigger than z. So, we subtract

$$\sum_{y < p \leq z} \psi\left(\frac{x}{p}, p\right), \quad \text{where } z \geq y,$$

from $\psi(x, z)$ and thus we get the desired result, therefore we can conclude the proof. □

Theorem 2.18. *For each $y \geq 2$ and $x \geq 2$, set $u := \log x / \log y$. Then,*

$$\psi(x, y) = (1 + o(1))x\rho(u)$$

as x tends to infinity in such a way that u is fixed, where $\rho(u)$ is the Dickman–de Brujin function given by Definition 2.20.

Proof. Say we want to count all the smooth numbers when $y \geq x$, then it is easy to see that

$$\psi(x, y) = \lfloor x \rfloor + O(1).$$

We now take the case when $\sqrt{x} \leq y \leq x$, this will be $u = \log x / \log y$. Thus, $y = x^{1/u}$. So, we will be looking at the case when $1 \leq u \leq 2$. To count these smooth numbers we will have

$$\begin{aligned}
\psi(x, y) &= \lfloor x \rfloor - \sum_{x^{1/u} < p \leq x} \sum_{\substack{n \leq x \\ p \mid n}} 1 = \lfloor x \rfloor - \sum_{x^{1/u} < p \leq x} \left\lfloor \frac{x}{p} \right\rfloor \\
&= x + O(1) - x \sum_{x^{1/u} < p \leq x} \frac{1}{p} + O\left(\sum_{x^{1/u} < p \leq x} \frac{x}{p} \right) \\
&= x - x \left(\log\log x + O\left(\frac{1}{\log x} \right) + \log\log x^{1/u} + O\left(\frac{u}{\log x} \right) \right) \\
&\quad + O\left(\sum_{x^{1/u} < p \leq x} \frac{x}{p} \right) \\
&= x - x \left(\log\left(\frac{u \log x}{\log x} \right) + O\left(\frac{1}{\log x} \right) \right) + O\left(\log\left(\frac{u \log x}{\log x} \right) + \frac{x}{\log x} \right) \\
&= x\left(1 - \log u\right) + O\left(\frac{x}{\log x} \right).
\end{aligned}$$

We have thus established that, for $0 \leq u \leq 2$,

$$\psi(x, x^{1/u}) = (1 + o(1)) x \rho(u) \qquad (x \to \infty),$$

where

$$\rho(u) = \begin{cases} 1, & \text{if } 0 \leq u \leq 1, \\ 1 - \log u, & \text{if } 1 \leq u \leq 2. \end{cases}$$

We shall compute the Dickman–de Brujin function for $u \in (1, 2]$ using Definition 2.20. Due to $u \in (1, 2]$, $\rho(u - 1) = 1$.

$$\begin{aligned}
u\rho'(u) &= -\rho(u - 1) \\
\rho(u) &= -\int_1^u \frac{\rho(v - 1)}{v} dv + C
\end{aligned}$$

If $u = 1$ then integral above will equate to 0. Since $\rho(1) = 1$ this implies that $C = 1$.

Now we need to show

$$\psi(x, x^{1/u}) = (1 + o(1))x\rho(u) \qquad (x \to \infty),$$

for each $u \geq 0$. We will use an induction type of approach. So, we assume that $\psi(x, x^{1/u}) = (1 + o(1))x\rho(u)$ as $x \to \infty$, holds for all $u \in [0, N]$, we must now show that it holds for all $u \in [N, N + 1]$. We will now apply Theorem 2.17, and we see that

$$\psi(x, y) = 1 + \sum_{p \leq y} \psi\left(\frac{x}{p}, p\right).$$

We obtain

$$\psi(x, x^{1/u}) \quad = \quad \psi(x, x^{1/N}) - \sum_{x^{1/u} < p \leq x^{1/N}} \psi\left(\frac{x}{p}, p\right).$$

Since $p > x^{1/u}$ and $u \in [N, N + 1]$, we have

$$\frac{\log(x/p)}{\log p} = \frac{\log x}{\log p} - 1 < \frac{\log x}{\log(x^{1/u})} - 1 = u - 1 \leq N,$$

so this satisfies our assumption that $\psi(x, x^{1/u}) = (1+o(1))x\rho(u)$ as $x \to \infty$, holds for all $u \in [0, N]$, say with $u' = \log(x/p)/\log p$ instead of u. Now we can make the appropriate substitutions to arrive at the following:

$$\begin{aligned}
\psi(x, x^{1/u}) &= \psi(x, x^{1/N}) - \sum_{x^{1/u} < p \leq x^{1/N}} \psi\left(\frac{x}{p}, p\right) \\
&= (1 + o(1))x\rho(N) - \sum_{x^{1/u} < p \leq x^{1/N}} \psi\left(\frac{x}{p}, p\right) \\
&= (1 + o(1))x\rho(N) - (1 + o(1))x \sum_{x^{1/u} < p \leq x^{1/N}} \frac{1}{p}\rho\left(\frac{\log x}{\log p} - 1\right).
\end{aligned}$$

We shall now apply Proposition 1 to the last outer sum above. Choosing $a_n = 1$ when n is prime and $a_n = 0$ when n is composite. Thus we take $f(t) = 1/t$.

$$\sum_{x^{1/u} < p \leq x^{1/N}} \frac{1}{p} \quad = \quad \frac{\pi(x^{1/N}) - \pi(x^{1/u})}{x^{1/N}} + \int_{x^{1/u}}^{x^{1/N}} \frac{\pi(t)}{t^2} dt.$$

Just comparing $\pi(x^{1/N})$ and $\pi(x^{1/u})$ and $u \in [N, N+1]$ with $N \geq 2$. The difference between $\pi(x^{1/N})$ and $\pi(x^{1/u})$ can be at most $O(1)$ as we know $\pi(x)/x = O(1)$. Taking the limit as $x \to \infty$, we apply Theorem 2.9 to get

$$\begin{aligned}
\sum_{x^{1/u}<p\leq x^{1/N}} \frac{1}{p} &= O(1) + \int_{x^{1/u}}^{x^{1/N}} \frac{t}{t^2 \log t} + O\left(\frac{t}{t^2(\log t)^2}\right) dt \\
&= O(1) + \int_{x^{1/u}}^{x^{1/N}} \frac{1}{t \log t} + O\left(\frac{1}{t \log t}\right) dt \\
&= O(1) + \int_{x^{1/u}}^{x^{1/N}} \frac{1 + o(1)}{t \log t} dt.
\end{aligned}$$

Letting $t = v$ and combining it with above we get:

$$\begin{aligned}
\psi(x, x^{1/u}) &= (1+o(1))x\rho(N) - (1+o(1))x \sum_{x^{1/u}<p\leq x^{1/N}} \frac{1}{p} \rho\left(\frac{\log x}{\log p} - 1\right) \\
&= (1+o(1))x\rho(N) - (1+o(1))x \int_{x^{1/u}}^{x^{1/N}} \rho\left(\frac{\log x}{\log v} - 1\right) \frac{dv}{v \log v} \\
&= (1+o(1))x \left(\rho(N) - \int_N^u \frac{\rho(t-1)}{t} dt\right) \\
&= (1+o(1))x\rho(u).
\end{aligned}$$

Therefore we get a consequence of Definition 2.20, where $\rho(N)$ is a constant as we compute in intervals of length 1.

$$\rho(u) = \rho(N) - \int_N^u \frac{\rho(v-1)}{v} dv \qquad (N \leq u \leq N+1),$$

we have then shown that $\psi(x, x^{1/u}) = (1+o(1))x\rho(u)$ as $x \to \infty$, holds for $u \in [N, N+1]$. □

Definition 2.21. *Let $z \in \mathbb{C}$ with $\mathrm{Re}(z) > 0$. The Gamma function denoted $\Gamma(z)$ is defined as*

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx.$$

Lemma 2.

$$\Gamma(z+1) = z\Gamma(z).$$

For Lemma 2 integration by parts is used to prove the result.
We can now solve for $\Gamma(1) = 1$. We also see that that $\Gamma(2+1) = 2 \cdot 1$, we can prove by induction that

$$\Gamma(n) = (n-1)!,$$

for all positive integers n.

Theorem 2.19. *The Dickman–de Brujin function $\rho(u)$ is a continuous function that has the following four properties:*

$$\begin{aligned}
&(i) \quad \rho(u) = \rho(N) - \int_{u-1}^{u} \frac{\rho(v)dv}{v} \quad (u \in [N, N+1], \text{ and } N \geq 1);\\
&(ii) \quad \rho(u) > 0 \quad (u > 0);\\
&(iii) \quad \rho'(u) < 0 \quad (u > 1);\\
&(iv) \quad \rho(u) \leq \frac{1}{\Gamma(u+1)} \quad (u \geq 0).
\end{aligned}$$

Proof. (i) Let $u > 1$. Let $f(u) = \rho(u)$, and let $g(u) = \rho(N) - \int_{u-1}^{u} \frac{\rho(v)dv}{v}$.

$$\begin{aligned}
f'(u) &= \rho'(u) \quad \text{where by Definition 2.20,}\\
\rho'(u) &= -\frac{\rho(u-1)}{u}\\
g'(u) &= 0 - \frac{\rho(u-1)}{u} = f'(u).
\end{aligned}$$

So, $f(u) + c = g(u) + k$, but $f(1) = 1$, and since $u \in [N, N+1]$ with $u = 1$ $g(u) = 1$. So, $c = k = 0$.
(ii) It is clear from (i).
(iii) It is clear due to

$$\rho'(u) = \frac{-\rho(u-1)}{u},$$

and the fact from (ii) that $\rho(u) > 0$ $(u > 0)$.
(iv) We will use induction on $k = \lfloor u \rfloor$. The property is satisfied for $k = 0$, since $\rho(u) = 1$ for $0 \geq u \geq 1$. On the other hand, for all $u \geq 1$, we have

$$\rho(u) = \rho(N) - \int_{u-1}^{u} \frac{\rho(v)dv}{v} \leq \frac{\rho(u-1)}{u},$$

due to (i) and $\rho(u)$ decreases from (iii).
Therefore by the induction hypothesis, if $k = \lfloor u \rfloor \geq 1$, we have

$$\rho(u) \leq \frac{1}{u\Gamma(u)} = \frac{1}{\Gamma(u+1)}.$$

□

Theorem 2.20. *(Stirling's formula) As* $n \to \infty$,

$$n! \sim n^n e^{-n}\sqrt{2\pi n}.$$

Note that Stirling's formula can be applied to $\Gamma(z)$ as well. Using property (iv) of the Dickman–de Brujin function (Theorem 2.20) as well as Stirling's formula (Theorem 2.20) we can show the following:

Theorem 2.21.

$$\begin{aligned}
\rho(u) &\leq \frac{1}{\Gamma(\lfloor u \rfloor + 1)} \\
&= \frac{1}{\lfloor u \rfloor !} \\
&= \frac{e^u}{u^u\sqrt{2\pi u}}(1 + o(1)) \\
&= u^{-(1+o(1))u}.
\end{aligned}$$

2.9 Motivation for this Thesis

2.9.1 Composite numbers n such that the sum of its distinct divisors divide n.

Luca and De Koninck studied composite integers which are divisble by the sum of their prime factors. They applied sieve methods to count sets with different properties to ascertain which set these composite integers can be found, thus enabling one to count these integers [5].

Recall, that for a subset $\mathcal{A}$ of positive integers and a real number $x \geq 1$, we write

$$\mathcal{A}(x) = \mathcal{A} \cap [1, x].$$

Theorem 2.22. *Let $\mathcal{A} = \{n \leq x : \ n$ is composite and $\beta(n) \mid n\}$. For x sufficiently large,*

$$x(\exp\{-c_1(1+o(1))\ell(x)\}) < \#\mathcal{A}(x) < x(\exp\{-c_2(1+o(1))\ell(x)\}),$$

where $\ell(x) := \sqrt{\log x \log\log x}$ and c_1 and c_2 can be taken as $c_1 = 3/\sqrt{2}$ and $c_2 = 1/\sqrt{2}$.

Before the current paper which this thesis is based on, I wished to study another property which in the end led to the writing of this current thesis. Let us first define the set we initially aimed to work on:

$$\mathcal{C}(x) = \{n \leq x : \ n \text{ is composite and } \beta(n) \mid \phi(n)\}.$$

This along with another property inspired the question: How many positive composite integers $n \leq x$ such that $\beta(n) \mid \phi(n)$?

2.9.2 Product of twin primes

The formula for the Euler Phi function is as follows:

$$\phi(n) = \prod_{p^j \mid n} p^{j-1}(p-1).$$

One of the special sets which satisfy the conditions to be in the set $\mathcal{C} = \{n \leq x : n$ is composite and $\beta(n) \mid \phi(n)\}$ is the product of the twin primes.

Let's say we have a a pair twin primes, say $p, q = p + 2$. Put $n = pq$. Then

$$\begin{aligned} \beta(n) &= 2(p+1) \\ \text{and } \phi(n) &= (p+1)(p-1). \end{aligned}$$

It is clearly seen that $\beta(n) \mid \phi(n)$.

2.9.3 Vinogradov's theorem

Theorem 2.23. *Every odd integer $n \geq 7$ can be written as a sum of 3 odd primes.*

$$n = p_i + p_j + p_k,$$

where i, j, k do not necessarily have to be distinct [1].

In 1923, Hardy and Littlewood [10] showed the number of ways n can be written as a sum of three primes denoted as $D_3(n)$ grows asymptotically as follows:

$$D_3(n) \sim c_3 \frac{n^2}{(\log n)^3} \prod_{p|n} \left(1 - \frac{1}{p^2 - 3p + 3}\right),$$

where $c_3 = \prod_{p \geq 3} \left(1 + \frac{1}{(p-1)^3}\right)$.

2.9.4 $\beta(n)$ being prime

When we look at $\beta(n)$ we notice that if $\omega(n) > 1$ and $\beta(n)$ is prime then we clearly see that if p_i divides n then, $p_i \nmid \beta(n)$. We can now construct a new number N such that $N = n\beta(n)^2$. Clearly $\beta(N) \mid \phi(N)$. We need to see how often $\beta(n)$ is a prime when n is composite or not a power of a prime. This was one of the main reasons that led to the research of this thesis.

$$\beta(N) = 2\beta(n) \mid \beta(n)\phi(n\beta(n)) = \phi(N).$$

We will now begin with the proof of the main result.

3 Composite numbers n whose sum of prime factors is prime: The upper bound

The proof for the main result begins. Theorem 1.1

$$\frac{x}{(\log x)^3} \ll \#\mathcal{B}(x) \ll \frac{x}{\log x}.$$

We carve off a finite number of slices of $\mathcal{B}(x)$ denoted $\mathcal{B}_i(x)$ for $i = 1, 2$ each of cardinality $O(x/\log x)$. Then we will study the condition $\beta(n)$ is prime on the left–over subsets of $n \le x$, for $\mathcal{B}_3(x), \mathcal{B}_4(x), \mathcal{B}_5(x)$.

3.1 Eliminating smooth numbers from $\mathcal{B}(x)$

We let $P(n)$ be the largest prime factor of n. We put $y := \exp(\log x/\log\log x)$ and let

$$\mathcal{B}_1(x) = \{n \le x : P(n) \le y\}. \tag{3.1}$$

In classical notation, and in our range of y versus x, we have

$$\#\mathcal{B}_1(x) = \Psi(x, y) = \rho(u)(1 + o(1))x = xu^{-(1+o(1))u}, \quad \text{where} \quad u = \frac{\log x}{\log y},$$

where $\rho(u)$ is the Dickman–de Brujin function. Since

$$\begin{aligned}
u^u &= \left(\frac{\log x}{\log y}\right)^{\frac{\log x}{\log y}}, \quad \text{we have} \\
u\log u &= \frac{\log x}{\log y}\log\left(\frac{\log x}{\log y}\right) \\
u\log u &= \log\log x(\log\log x - \log\log y) \\
u\log u &= \log\log x\left(\log\log x - \log\left(\frac{\log x}{\log\log x}\right)\right) \\
u\log u &= \log\log x(\log\log x - \log\log x + \log\log\log x) \\
u\log u &= (\log\log\log x)\log\log x \\
u^u &= \log x^{\log\log\log x}
\end{aligned}$$

So, we have that $u^{-(1+o(1))u} = (\log x)^{-(1+o(1))\log\log\log x}$ as $x \to \infty$, it follows that

$$\#\mathcal{B}_1(x) = \frac{x}{(\log x)^{(1+o(1))\log\log\log x}} \ll \frac{x}{\log x}. \tag{3.2}$$

Here we showed that $n \in \mathcal{B}_1(x)$ (with the property $P(n) \leq y$) may or may not contain composite numbers n that fufill the property of $\beta(n)$ being prime. Despite this we have shown that the cardinality of this set cannot grow faster than $O\left(\frac{x}{\log x}\right)$.

3.2 Eliminating numbers divisible by the square of a large prime

We put $z := (\log x)^{10}$ and let

$$\mathcal{B}_2(x) = \{n \leq x : p^2 \mid n \text{ for some prime } p > z\}. \tag{3.3}$$

Given a prime p, the number of $n \leq x$ divisible by p^2 is $\lfloor n/p^2 \rfloor \leq x/p^2$. Thus,

$$\#\mathcal{B}_2(x) \leq \sum_{p>z} \frac{x}{p^2} < x \sum_{m>z} \frac{1}{m^2} \ll x \int_z^\infty \frac{dt}{t^2} \ll \frac{x}{z}.$$

In particular,

$$\#\mathcal{B}_2(x) \ll \frac{x}{\log x}. \tag{3.4}$$

Here we notice that for not too large values of x, $z > y$. When x is sufficiently large $y > z$.

3.3 Applying the sieve

Assume $n \in \mathcal{B}(x) \setminus (\mathcal{B}_1(x) \cup \mathcal{B}_2(x))$. Then $n = Pm$, where $P = P(n) > y$ (as we have eliminated the smooth numbers from $\mathcal{B}_1(x)$). Thus, $2 \leq m < x/y$. For large x, we have $P > y > z$, so $P \nmid m$. Fix m. Then

$$\beta(n) = P + \beta(m) = P_1,$$

where P_1 is also a prime. We apply Theorem 2.15, with $v(p) = 2$ if $p \nmid \beta(m)$ and $v(p) = 1$ otherwise. By this sieve the number of primes $P \leq x/m$ with

the property that $P+\beta(m)$ is a prime is

$$\begin{aligned}
&\ll \frac{x}{m}\prod_{3\le p\le x/m}\left(1-\frac{v(p)}{p}\right)\\
&\ll \frac{x}{m}\prod_{3\le p\le x/m}\left(1-\frac{2}{p}\right)\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\left(\left(1-\frac{1}{p}\right)\left(1-\frac{2}{p}\right)^{-1}\right)\\
&= \frac{x}{m}\prod_{3\le p\le x/m}\left(1-\frac{2}{p}\right)\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\left(\frac{p-1}{p-2}\right)\\
&= \frac{x}{m}\prod_{3\le p\le x/m}\left(1-\frac{2}{p}\right)\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\left(\frac{p-1}{p-2}\right)\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\frac{\left(1+\frac{1}{p}\right)}{\left(1+\frac{1}{p}\right)}\\
&= \frac{x}{m}\prod_{3\le p\le x/m}\left(1-\frac{2}{p}\right)\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\frac{\left(\frac{p-1}{p-2}\right)}{\left(1+\frac{1}{p}\right)}\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\left(1+\frac{1}{p}\right)\\
&= \frac{x}{m}\prod_{3\le p\le x/m}\left(1-\frac{2}{p}\right)\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\frac{p^2-p}{p^2-p-2}\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\left(1+\frac{1}{p}\right)\\
&= \frac{x}{m}\prod_{3\le p\le x/m}\left(1-\frac{2}{p}\right)\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\left(1+\frac{2}{p^2-p-2}\right)\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\left(1+\frac{1}{p}\right)\\
&\ll \frac{x}{m}\prod_{3\le p\le x/m}\left(1-\frac{2}{p}\right)\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\left(1+O\left(\frac{1}{p^2}\right)\right)\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\left(1+\frac{1}{p}\right)\\
&\ll \frac{x}{m}\prod_{3\le p\le x/m}\left(1-\frac{2}{p}\right)\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\left(1+\frac{1}{p}\right)\\
&\ll \frac{x}{m}\prod_{3\le p\le x/m}\left(1-\frac{2}{p}\right)\prod_{\substack{3\le p\le x/m\\ p|\beta(m)}}\left(1+\sum_{\substack{1\le k\le l\\ p^l||\beta(m)}}\frac{1}{p^k}\right)\\
&= \frac{x}{m}\prod_{3\le p\le x/m}\left(1-\frac{2}{p}\right)\left(\frac{\sigma(\beta(m))}{\beta(m)}\right)\\
&\ll \frac{x}{m(\log(x/m))^2}\left(\frac{\sigma(\beta(m))}{\beta(m)}\right).
\end{aligned}$$

Now we sum up over m. Note that since $n = Pm \le x$ and $P > P(m)$, it follows that the m's under scrutiny satisfy $mP(m) \le x$. Thus,

$$\#\left(\mathcal{B}(x)\setminus\left(\mathcal{B}_1(x)\cup\mathcal{B}_2(x)\right)\right) \ll \sum_{\substack{m\le x/y\\ mP(m)\le x}} \frac{x}{m(\log(x/m))^2}\left(\frac{\sigma(\beta(m))}{\beta(m)}\right).$$

If most of the sum over such m is concentrated on m's that are not too small, say $\log(x/m) \asymp \log x$ and if in addition $\sigma(\beta(m))/\beta(m) = O(1)$ "on the average" over such m, then we win since the sum of the reciprocals of m introduces at most another logarithm. Hence, we have to deal with $\log(x/m)$ and with $\sigma(\beta(m))/\beta(m)$. We first deal only with $\sigma(\beta(m))/\beta(m)$ when m is not too small as a warm-up example, and deal with the general case later.

3.4 Eliminating smooth m's

Let $w := \exp(\log x/(\log\log x)^3)$. Let

$$\mathcal{B}_3(x) = \{n \in \mathcal{B}(x)\setminus\left(\mathcal{B}_1(x)\cup\mathcal{B}_2(x)\right) \text{ and } P(m) \le w\}. \tag{3.5}$$

Then

$$\#\mathcal{B}_3(x) \le \sum_{\substack{m\le x/y\\ P(m)\le w}} \frac{x}{m(\log(x/m))^2}\left(\frac{\sigma(\beta(m))}{\beta(m)}\right).$$

Since $x/m \ge y$, we have $\log(x/m) \ge \log y = \log x/\log\log x$. Further, since $\beta(m) \le m \le x$, it follows, by the maximal order of the sum-of-divisors function, that $\sigma(\beta(m))/\beta(m) \ll \log\log x$. Hence,

$$\begin{aligned}
\#\mathcal{B}_3(x) &\le \sum_{\substack{m\le x/y\\ P(m)\le w}} \frac{x}{m(\log(x/m))^2}\left(\frac{\sigma(\beta(m))}{\beta(m)}\right)\\
&\le x\sum_{\substack{m\le x/y\\ P(m)\le w}} \frac{1}{m(\log y)^2}\,(\log\log x)\\
&\ll \frac{x(\log\log x)^3}{(\log x)^2}\sum_{\substack{m\le x/y\\ P(m)\le w}} \frac{1}{m}.
\end{aligned}$$

We have

$$\frac{x(\log\log x)^3}{(\log x)^2}\sum_{\substack{m\le x/y\\ P(m)\le w}}\frac{1}{m}\ll\frac{x(\log\log x)^3}{(\log x)^2}\prod_{p\le w}\left(1-\frac{1}{p}\right)^{-1}.$$

Next we use Theorem 2.7 to get

$$\prod_{p\le w}\left(1-\frac{1}{p}\right)^{-1}\ll\log w.$$

Hence, we get

$$\frac{x(\log\log x)^3}{(\log x)^2}\prod_{p\le w}\left(1-\frac{1}{p}\right)^{-1}\ll\frac{x(\log\log x)^3\log w}{(\log x)^2}=\frac{x}{\log x}.$$

Therefore we have:

$$\#\mathcal{B}_3(x)\ll\frac{x}{\log x}. \tag{3.6}$$

The next trick we use is to note that if d is a divisor of s, so is s/d.

$$\frac{\sigma(s)}{s}=\sum_{d|s}\frac{1}{d}\le\sum_{\substack{d|s\\ d\le\sqrt{s}}}\left(\frac{1}{d}+\frac{1}{s/d}\right)\le 2\sum_{\substack{d|s\\ d\le\sqrt{s}}}\frac{1}{d}.$$

We apply this with $s=\beta(m)$ and change the order of summation getting that

$$\#\left(\mathcal{B}(x)\setminus\left(\mathcal{B}_1(x)\cup\mathcal{B}_2(x)\cup\mathcal{B}_3(x)\right)\right)\ll x\sum_{d\le\sqrt{x/y}}\frac{1}{d}\sum_{\substack{m\in\mathcal{M}\\ \beta(m)\ge d^2\\ \beta(m)\equiv 0\pmod d}}\frac{1}{m(\log(x/m))^2},$$

where

$$\mathcal{M}:=\{m\le\min\{x/y,x/P(m)\};Q=P(m)>w,\ p^2\nmid m\text{ for }p>z\}.$$

3.5 The case $m\le x^{9/10}$

We write $\mathcal{M}_1=\mathcal{M}\cap[1,x^{9/10}]$ and

$$\mathcal{B}_4(x)=\left\{n\in\mathcal{B}(x)\setminus\left(\mathcal{B}_1(x)\cup\mathcal{B}_2(x)\cup\mathcal{B}_3(x)\right):m\in\mathcal{M}_1\right\}. \tag{3.7}$$

When $m \in \mathcal{M}_1$, we have $x/m \geq x^{1/10}$, so $\log(x/m) \geq \log(x^{1/10}) \gg \log x$. Thus,

$$\#\mathcal{B}_4(x) \ll \frac{x}{(\log x)^2} \sum_{d \leq \sqrt{x/y}} \frac{1}{d} \sum_{\substack{m \in \mathcal{M}_1 \\ \beta(m) \geq d^2 \\ \beta(m) \equiv 0 \pmod d}} \frac{1}{m} := \frac{x}{(\log x)^2} \sum_{d \leq \sqrt{x/y}} \frac{S_d}{d}. \tag{3.8}$$

We further separate variables in m. Namely we write $m = Q\ell$, where $Q = P(m)$. Since $Q > w$ and $w > z$ for large x, it follows that for large x, $P(\ell) < Q$. Now $\beta(m) = Q + \beta(\ell) \geq Q$. Further, since $\omega(m) \ll \log m / \log\log m \leq \log x / \log\log x$ holds for all our m, and $Q > z$, it follows that

$$\beta(m) \ll Q(\log x / \log\log x), \quad \text{therefore} \quad \beta(m) < Q \log x = Q z^{1/10} < Q^{1.1}$$

holds for all large x. Since $d^2 \leq \beta(m)$, it follows that $Q > d^{2/1.1} > d^{1.8}$. So, we fix d and ℓ. Then $\beta(m) = Q + \beta(\ell) \equiv 0 \pmod d$ puts Q into the arithmetic progression $d - \beta(\ell)$. which depends on ℓ.
Also, $Q \in [\max\{P(\ell), d^{1.8}\}, x]$. By Theorem 2.16, the counting function of primes $Q \leq t$ in this arithmetic progression is

$$\pi(t, d, -\beta(\ell)) \ll \frac{t}{\phi(d)\log(t/d))} \ll \frac{t}{\phi(d)\log t} \quad \text{for} \quad t \in [\max\{P(\ell), d^{1.8}\}, x]. \tag{3.9}$$

Fixing ℓ and summing up over all m with the fixed ℓ and the prime $P(m) = Q$ in the above progression modulo d, we get

$$S_{d,\ell} \ll \frac{1}{\ell} \sum_{\substack{\max\{P(\ell), d^{1/8}\} \leq Q \leq x \\ Q \equiv d - \beta(\ell)}} \frac{1}{Q}.$$

Here we use Abel's summation (Proposition 1) to evaluate the sum.

$$\sum_{\substack{\max\{P(\ell), d^{1/8}\} \leq Q \leq x \\ Q \equiv d - \beta(\ell)}} \frac{1}{Q}.$$

We choose $f(n) = 1/Q$ thus $f'(t) = -1/t^2$. And we choose our

$$a_n = \begin{cases} 1 & n = Q, \\ 0 & \text{otherwise.} \end{cases}$$

Hence,

$$\sum_{\substack{\max\{P(\ell),d^{1/8}\}\le Q\le x\\ Q\equiv d-\beta(\ell)}}\frac{1}{Q}\ll\frac{\pi(x,d,-\beta(\ell))}{x}+\int_{P(\ell)}^{x}\frac{\pi(t,d,-\beta(\ell))}{t^2}dt.$$

Applying (11) we get the next line and then simple integration follows.

$$\begin{aligned}
S_{d,\ell} &\ll \frac{1}{\ell}\sum_{\substack{\max\{P(\ell),d^{1/8}\}\le Q\le x\\ Q\equiv d-\beta(\ell)}}\frac{1}{Q}\\
&\ll \frac{1}{\ell}\left(\frac{\pi(x,d,-\beta(\ell))}{x}+\int_{P(\ell)}^{x}\frac{\pi(t,d,-\beta(\ell))}{t^2}dt\right)\\
&\ll \frac{1}{\phi(d)\ell}\left(\frac{1}{\log x}+\int_{P(\ell)}^{x}\frac{dt}{t\log t}\right)\\
&\ll \frac{1}{\phi(d)\ell}\left(\frac{1}{\log x}+(\log\log x-\log\log P(\ell))\right). \qquad (3.10)
\end{aligned}$$

Summing up the first terms $\ell\le m\le x^{0.9}$ and over $d\le x$, we get at most

$$O\left(\frac{x}{(\log x)^3}\sum_{\ell\le x^{0.9}}\frac{1}{\ell}\sum_{d\ge 1}\frac{1}{d\phi(d)}\right)$$

integers $n\in\mathcal{B}_4(x)$. We use the fact that the last series is convergent because $\phi(d)\gg d/\log\log d$, so $d\phi(d)\gg d^2/\log\log d\gg d^{3/2}$. The sum of the recipricol of $d^{3/2}$ is convergent another value could have been chosen. We now look at

$$\sum_{\ell\le x^{0.9}}\frac{1}{\ell}\le\sum_{\ell\le x}\frac{1}{\ell}\ll\log x.$$

So, summing up the first terms $\ell\le m\le x^{0.9}$ and over $d\le x$, we get a contribution of at most

$$\frac{x}{(\log x)^3}\sum_{\ell\le x^{0.9}}\frac{1}{\ell}\sum_{d\ge 1}\frac{1}{d\phi(d)}\ll\frac{x}{(\log x)^2}.$$

Summing up the second terms in (3.10) over ℓ then over $d \le x$, we get

$$\#\mathcal{B}_4(x) \ll \frac{x}{(\log x)^2} \sum_{3 \le \ell \le x^{0.9}} \left(\frac{\log\log x - \log\log P(\ell)}{\ell} \right) \sum_{d \ge 1} \frac{1}{d\phi(d)} + \frac{x}{(\log x)^2}$$

$$\ll \frac{x}{(\log x)^2} \sum_{3 \le \ell \le x} \frac{\log\log x - \log\log P(\ell)}{\ell}.$$

In the above, we discarded the cases $\ell = 1, 2$ since then $m = PQ,\ 2PQ$ with P and Q large primes, for these cases we have $\beta(m) = P + Q,\ P + Q + 2$ which results in $\beta(m)$ being even and large. We also absorbed the second term $x/(\log x)^2$ into the first term (say for $\ell = 3$). To continue, we may assume that $P(\ell) > y$. Indeed, the part of the above sum for $P(\ell) \le y$ gives a contribution of

$$\frac{x \log\log x}{(\log x)^2} \sum_{P(\ell) \le y} \frac{1}{\ell} \ll \frac{x(\log\log x)\log y}{(\log x)^2} \ll \frac{x}{\log x}.$$

As

$$\frac{x}{(\log x)^2} \sum_{3 \le \ell \le x} \frac{\log\log x - \log\log P(\ell)}{\ell} \ll \frac{x \log\log x}{(\log x)^2} \sum_{P(\ell) \le y} \frac{1}{\ell}.$$

And

$$\sum_{P(\ell) \le y} \frac{1}{\ell} \le \prod_{p \le y} \left(1 - \frac{1}{p}\right)^{-1} \ll \log y,$$

which is acceptable for us. For the rest, we put $P(\ell) \in [x^{1/(j+1)}, x^{1/j}]$ for some integer $j \in [1, \log\log x]$. Note $x > \log P(\ell) > \log y$, thus we have $\log x > \log P(\ell) > \log x/\log\log x$. Thus we can set j as we have done. Also note it is okay to count additional terms as long as we have covered our desired intervals.
For such ℓ, we have

$$\log\log P(\ell) = \log\log x + O(\log(j+1)),$$

so

$$\log\log x - \log\log P(\ell) = O(\log(j+1)).$$

For such an ℓ, we write $\ell = R\ell_1$, where $R = P(\ell)$. We then get

$$\begin{aligned}
\mathcal{B}_4(x) &\ll \frac{x}{(\log x)^2} \sum_{3\le \ell \le x^{0.9}} \frac{\log\log x - \log\log P(\ell)}{\ell} \\
&\ll \frac{x}{(\log x)^2} \sum_{1\le j\le \log\log x} \sum_{\substack{3\le \ell\le x^{0.9} \\ x^{1/(j+1)} < P(\ell) \le x^{1/j}}} \frac{\log(j+1)}{\ell} \\
&\ll \frac{x}{(\log x)^2} \\
&\times \sum_{1\le j\le \log\log x} \log(j+1) \sum_{x^{1/j+1} < R \le x^{1/j}} \frac{1}{R} \sum_{P(\ell_1)\le x^{1/j}} \frac{1}{\ell_1} \\
&\ll \frac{x}{(\log x)^2} \\
&\times \sum_{1\le j\le \log\log x} \log(j+1) \left(\log\log x^{1/j} - \log\log x^{1/j+1} + O\left(\frac{j}{\log x}\right)\right) \\
&\times \left(\log x^{1/j} + O(1)\right) \text{ (Apply Theorem 2.5)} \\
&\ll \frac{x}{(\log x)^2} \sum_{1\le j\le \log\log x} \log(j+1) \left(\log\left(1+\frac{1}{j}\right) + O\left(\frac{z^{0.1}}{\log x}\right)\right) \\
&\times \left(\frac{\log x}{j} + O(1)\right) \\
&\ll \frac{x(\log x)}{(\log x)^2} \sum_{j\ge 1} \frac{\log(j+1)}{j^2} \ll \frac{x}{\log x} \\
\mathcal{B}_4(x) &\ll \frac{x(\log x)}{(\log x)^2} \sum_{j\ge 1} \frac{\log(j+1)}{j^2} \ll \frac{x}{\log x}. \qquad (3.11)
\end{aligned}$$

We have $\mathcal{B}_4(x) \ll x/\log x$, which is what we wanted for $\mathcal{B}_4(x)$. In the above, we used the fact that the series $\sum_{j\ge 1}(\log(j+1))/j^2$ is convergent. Moving to the inequality from step 3 to step 4 for $\mathcal{B}_4(x)$ Mertens' Estimate Theorem 2.5 is applied to the following sum:

$$\sum_{a\le R\le b} \frac{1}{R} = \log\log b - \log\log a + O(1/\log a)$$

with $a = x^{1/(j+1)}$ and $b = x^{1/j}$.
With these choices, $\log\log b - \log\log a = \log(1+1/j)$, and we also have,

$$\lim_{j\to\infty} j\log(1+1/j) = 1,$$

therefore $\log(1+1/j) = O(1/j)$, and this is much larger than

$$O(1/\log a) = O(j/\log x) \text{ since } j \le \log\log x < \log x = z^{0.1}.$$

3.6 The case m large

Here, we put

$$\mathcal{B}_5(x) = \mathcal{B}(x) \setminus \left(\cup_{i=1}^{4} \mathcal{B}_i(x)\right). \tag{3.12}$$

We put $\mathcal{M}_2 = \mathcal{M}\setminus\mathcal{M}_1$ and then, by the argument leading to (3.8), we have

$$\#\mathcal{B}_5(x) \ll x \sum_{d\le\sqrt{x/y}} \frac{1}{d} \sum_{\substack{m\in\mathcal{M}_2\\ \beta(m)\ge d^2\\ \beta(m)\equiv 0 \pmod d}} \frac{1}{m(\log(x/m))^2} := x \sum_{d\le\sqrt{x/y}} \frac{T_d}{d}. \tag{3.13}$$

Every m participating in T_d satisfies $m \ge x^{9/10}$. We split the sum over m according to $x/m \in [x^{1/k+1}, x^{1/k}]$. For such m, we have $\log(x/m) \ge (\log x)/(k+1)$. Further, $m \in [x^{1-1/k}, x^{1-1/(k+1)}]$, and since $m \ge x^{9/10}$, it follows that $k \ge 9$. Further, $w < P(m) < P < x/m \le x^{1/k}$, therefore $k \le (\log\log x)^3 = \frac{3}{10}\log z \le z^{0.3}$. As here we have

$$\log w \le \frac{1}{k}\log x, \qquad \text{where } \log w = \frac{\log x}{(\log\log x)^3}.$$

Taking reciprocals and solve for k we get $k \le (\log\log x)^3$. Hence,

$$\begin{aligned} T_d &\le \frac{1}{(\log x)^2} \sum_{9\le k\le(\log\log x)^3} (k+1)^2 \sum_{\substack{x^{1-1/k}\le m\le x^{1-1/(k+1)}\\ P(m)<x^{1/k}\\ \beta(m)\ge d^2\\ \beta(m)\equiv 0 \pmod d}} \frac{1}{m} \\ &:= \frac{1}{(\log x)^2} \sum_{9\le k\le(\log\log x)^3} (k+1)^2 T_{d,k}. \end{aligned}$$

For each inner sum, we apply the argument used when bounding $\mathcal{B}_4(x)$. Namely, $\beta(m) = Q + \beta(\ell) \equiv 0 \pmod d$, which puts, for a fixed ℓ, the prime $Q = P(m)$ into the arithmetic progression $-\beta(\ell)$ modulo d. Further, $Q \ge \max\{P(\ell), d^{1.8}\}$ so the count (3.9) applies. Fixing ℓ and summing up over all Q, we get, by the argument used at (3.10), that

$$T_{d,k,\ell} \ll \frac{1}{\ell} \sum_{\substack{\max\{P(\ell),d^{1/8}\}\le Q\le x\\ Q\equiv -\beta(\ell) \pmod d}} \frac{1}{Q} \ll \frac{\log\log x - \log\log P(\ell)}{\phi(d)\ell}. \tag{3.14}$$

For the last estimate above, we used the fact that $P(\ell) \le x^{1/k} \le x^{1/10}$, so $\log\log x - \log\log P(\ell) \ge \log\log x - \log\log x^{1/10} \gg 1$. Thus

$$\frac{1}{\phi(d)\ell}\left(\frac{1}{\log x} + (\log\log x - \log\log P(\ell))\right) \ll \frac{(\log\log x - \log\log P(\ell))}{\phi(d)\ell}\left(\frac{1+\log x}{\log x}\right),$$

where $(1+\log x)/\log x = O(1)$. Hence,

$$T_d \ll \frac{1}{\phi(d)(\log x)^2} \sum_{9\le k<(\log\log x)^3} (k+1)^2 \sum_{\ell\in\mathcal{L}_k} \frac{\log\log x - \log\log P(\ell)}{\ell}.$$

Here $\mathcal{L}_k$ is the set of positive integers m such that $m = Q\ell$ for some prime $Q > P(\ell)$, and m satisfies the conditions that $m \in [x^{1-1/k}, x^{1-1/(k+1)}]$, $\beta(m) \equiv 0 \pmod{d^2}$. Therefore,

$$\#\mathcal{B}_5(x) \ll \frac{x}{(\log x)^2} \sum_{9\le k\le(\log\log x)^3} (k+1)^2 \sum_{\ell\in\mathcal{L}_k} \left(\frac{\log\log x - \log\log P(\ell)}{\ell}\right) \sum_{d\ge 1} \frac{1}{d\phi(d)}.$$

The last sum is $O(1)$ as noted on page 53 after (3.10). We may assume that

$$P(\ell) \ge w_1 := \exp(\log x/(\log\log x)^{10}),$$

since for $P(\ell) \le w_1$, we have

$$\begin{aligned}
&\frac{x}{(\log x)^2} \sum_{9\le k\le(\log\log x)^3} (k+1)^2 \sum_{P(\ell)\le w_1} \left(\frac{\log\log x - \log\log P(\ell)}{\ell}\right) \\
\ll\ &\frac{x(\log\log x)^6}{(\log x)^2} \sum_{3\le k\le(\log\log x)^3} \sum_{P(\ell)\le w_1} \frac{\log\log x}{\ell} \\
\ll\ &\frac{x(\log\log x)^{10}}{(\log x)^2} \sum_{P(\ell)\le w_1} \frac{1}{\ell} \\
\ll\ &\frac{x(\log\log x)^{10}\log w_1}{(\log x)^2} \ll \frac{x}{\log x}.
\end{aligned}$$

So, we have

$$\#\mathcal{B}_5(x) \ll \frac{x}{(\log x)^2} \sum_{9\le k\le(\log\log x)^3} (k+1)^2 \sum_{\substack{\ell\in\mathcal{L}_k \\ P(\ell)\ge w_1}} \left(\frac{\log\log x - \log\log P(\ell)}{\ell}\right),$$

and we may concentrate on the case $P(\ell) \geq w_1$. Note that since $Q \leq x^{1/k}$, it follows that $\ell = m/Q \geq x^{1-2/k}$. We now let $j \geq 0$ and assume that $P(\ell) \in [x^{1/(k+j+1)}, x^{1/(k+j)}]$. Since $P(\ell) \geq w_1$, it follows that

$$j < j + k \leq (\log\log x)^{10}.$$

Write $\ell = R\ell_1$, where $R = P(\ell)$. Then $\ell_1 = \ell/R \geq x^{1-2/k-1/(k+j)} \geq x^{1-3/k} \geq x^{1/3}$ is not too small. We easily verify this by showing

$$\frac{2}{k} + \frac{1}{k+j} = \frac{3k+j}{k(k+j)} \leq \frac{3(k+j)}{k(k+j)} \leq \frac{1}{3},$$

since $k \geq 9$ and $j \geq 0$. Since $P(\ell_1) \leq P(\ell)$, it follows that such numbers ℓ_1 are quite smooth, namely their $u = \log \ell_1/\log P(\ell_1)$ is at least

$$u = \frac{\log \ell_1}{\log P(\ell_1)} \geq \frac{\log(x^{1/3})}{\log x^{1/(k+j)}} = \frac{k+j}{3}.$$

For such numbers ℓ, we also have $\log\log x - \log\log P(\ell) = O(\log(k+j))$. We thus get that

$$\begin{aligned}\mathcal{B}_5(x) \ll \frac{x}{(\log x)^2} \sum_{9 \leq k < (\log\log x)^3} (k+1)^2 \sum_{j \geq 0} \log(k+j) \sum_{x^{1/(k+j+1)} \leq R \leq x^{1/(k+j)}} \frac{1}{R} \\ \times \sum_{\substack{\ell_1 \in [x^{1/3}, x] \\ u \geq (k+j)/3}} \frac{1}{\ell_1}.\end{aligned}$$

The last inner sum is

$$\ll \rho((k+j)/3)\log x.$$

As we get a $\log x$ from summing $(1/\ell_1)$ from 1 to x and we count all the smooth numbers using the Dickman–de Brujin function. The sum prior to it is

$$\begin{aligned}&\log\log(x^{1/(k+j)} - \log\log x^{1/(k+j+1)} + O\left(\frac{1}{\log x^{1/(k+j)}}\right) \\ = \ &\log\left(1 + \frac{1}{k+j}\right) + O\left(\frac{k+j}{\log x}\right) \ll \frac{1}{k+j} + O\left(\frac{(\log\log x)^{10}}{\log x}\right) \ll \frac{1}{k+j}.\end{aligned}$$

Thus,

$$\mathcal{B}_5(x) \ll \frac{x}{\log x} \sum_{3 \leq k \leq (\log\log x)^3} (k+1)^2 \sum_{0 \leq j \leq (\log\log x)^{10}} \frac{\rho((k+j)/3)\log(k+j)}{k+j}.$$

Since $\rho(u) = u^{-(1+o(1))u}$ as $u \to \infty$, it follows that $\rho((k+j)/3) \ll (k+j)^{-4}$. This is true as

$$\log(u^{-u}) = (-(k+j)/3)\log((k+j)/3) \leq -(k+j)\log(k+j) \leq -4\log(k+j),$$

as $(k+j) > 4$. Hence, the inner sum above is bounded by

$$\begin{aligned}\sum_{0\leq j\leq(\log\log x)^{10}} \frac{\rho((k+j)/3)\log(k+j)}{k+j} &\ll \sum_{j\geq 0}\frac{\log(k+j)}{(k+j)^5} \leq \sum_{K\geq k}\frac{\log K}{K^5}\\ &\ll \int_k^\infty \frac{\log t dt}{t^5} = -\frac{4\log x+1}{16x^4}\Big|_k^\infty\\ &\ll \frac{\log k}{k^4},\end{aligned}$$

and so, we have the following:

$$\begin{aligned}\mathcal{B}_5(x) &\ll \frac{x}{\log x}\sum_{9\leq k\leq(\log\log x)^3}\frac{(k+1)^2\log k}{k^4}\\ &\ll \frac{x}{\log x}\sum_{k\geq 9}\frac{\log k}{k^2} \ll \frac{x}{\log x}.\end{aligned}$$

Which is want we wanted to prove.

Recall the upperbound which we aimed for in Theorem 1.1.

$$\#\mathcal{B}(x) \ll \frac{x}{\log x}.$$

We have $\#\mathcal{B}(x) = \#(\mathcal{B}_1(x)\cup\mathcal{B}_2(x)\cup\mathcal{B}_3(x)\cup\mathcal{B}_4(x)\cup\mathcal{B}_5(x))$. Where

$$\begin{aligned}\mathcal{B}_1(x) &= \{n\leq x : P(n)\leq y\}\\ \mathcal{B}_2(x) &= \{n\leq x : p^2 \mid n \text{ for some prime } p > z\}\\ \mathcal{B}_3(x) &= \{n\in\mathcal{B}(x)\setminus(\mathcal{B}_1(x)\cup\mathcal{B}_2(x)) \text{ and } P(m)\leq w\}\\ \mathcal{B}_4(x) &= \{n\in\mathcal{B}(x)\setminus(\mathcal{B}_1(x)\cup\mathcal{B}_2(x)\cup\mathcal{B}_3(x)) : m\in\mathcal{M}_1\}\\ \mathcal{B}_5(x) &= \mathcal{B}(x)\setminus\left(\cup_{i=1}^4\mathcal{B}_i(x)\right).\end{aligned}$$

Namely equations (3.1), (3.3), (3.5), (3.7) and (3.12). Where all

$$\mathcal{B}_i(x) \ll x/\log x.$$

4 Composite numbers n whose sum of prime factors is prime: The lower bound

4.1 A conditional lower bound

We start with the conditional proof since it is easier. For all even positive integers h, let $\pi_2(x,h)$ be the number of primes p such that $p+h \le x$ is also prime. We refer to Section 2.6. For all positive integers h, let $\pi(x,h)$ be the number of primes p such that $p+h \le x$. The Bateman–Horn Conjecture for the acceptable system of polynomials $X, X+h$ implies that

$$\pi_2(x,h) \asymp c_h \frac{x}{(\log x)^2} \quad \text{where} \quad c_h = 2\left(\prod_{p\neq 2}\frac{p(p-2)}{(p-1)^2}\right)\left(\prod_{\substack{p\mid h\\ p>2}}\frac{p-1}{p-2}\right)$$

as $x\to\infty$. This follows from the Bateman–Horn Conjecture (Conjecture 2.1). Assume that $\pi_2(h) \gg c_h x/(\log x)^2$ holds for all even $h \le x^\delta$ for some small $\delta > 0$. We may assume that $\delta < 1/3$. Consider numbers $m \le x^{\delta/2}$ which are odd and for which $\omega(m)$ is even. Then $\beta(m)$ is even and

$$\beta(m) \le m \le x^{\delta/2} < (x/m)^\delta.$$

Thus, by the above assumption, with $h = \beta(m)$, we have

$$\pi_2(x/m,h) \gg c_h \frac{x}{m(\log(x/m))^2} \gg \frac{x}{m(\log x)^2}.$$

We may assume that the primes counted by $\pi_2(x/m,h)$ exceed the amount $x/(m(\log x)^2) > x^{(1-(2/3))} > x^{(1-(\delta/2)-(3\delta/2))} > x^{1-2\delta}$, as $(\log x)^2 \ll x^{1/4}$ (choosing $\delta < 1/3$), since the number of primes $p < x/(m(\log x)^2)$ is

$$\pi\left(\frac{x}{(m(\log x)^2)}\right) = \frac{x}{(m(\log x)^2)(\log \frac{x}{(m(\log x)^2)})} \ll \frac{x}{(m\log x)^3},$$

so we can ignore such primes from the count of $\pi_2(x/m,h)$ and still keep $\gg x/(m(\log x)^2)$ primes $p \le x/m$ with $p+\beta(m)$ prime. We now sum up over $m \le x^{\delta/2}$ getting a count of

$$\gg \frac{x}{(\log x)^2} \sum_{\substack{m\le x^{\delta/2}\\ m\equiv 1 \pmod 2\\ \omega(m)\equiv 0 \pmod 2}} \frac{1}{m}$$

pairs (m,p) with $m \le x^{\delta/2}$, $p \le x/m$ and $p+\beta(m)$ is prime. Thus, $n = pm \le x$. Further, $p > x/(m(\log x)^2) > m$ for large x, this will imply that $p \nmid m$, so given n, we have $p = P(n)$ is uniquely determined. Thus, this construction produces distinct integers $n \in \mathcal{B}(x)$. It is enough to show that the sum over m's is a positive proportion of $\log x$. Well, assume further $m \le x^{\delta/2}/3$ and that m is coprime to 6 (Note: we can write numbers which are congruent to 1 (mod 6) in the form of $6\ell+1$) and write

$$\sum_{\substack{m\le x^{\delta/2}/3\\ \gcd(m,6)=1}} \frac{1}{m} \ge \sum_{\ell\le \lfloor x^{\delta}/18\rfloor -1} \frac{1}{6\ell+1} \gg \sum_{\ell\le \lfloor x^{\delta}/18\rfloor -1} \frac{1}{\ell} \gg \log\left(\frac{x^{\delta/2}}{18}-1\right) \gg \log x.$$

Write the sum in the left as S_0+S_1, where S_i denotes the contribution to the sum of m with $\omega(m) \equiv i \pmod 2$ for $i=0,1$. If $S_1 \ge S_0$, then $S_1 \ge (S_0+S_1)/2 \gg \log x$, and we are through. Otherwise, $S_0 \gg \log x$ and

$$\sum_{\substack{m\le x^{\delta/2}\\ m\equiv 1 \pmod 2\\ \omega(m)\equiv 0 \pmod 2}} \frac{1}{m} \ge \sum_{\substack{m\le x^{\delta/2}/3\\ \gcd(m,6)=1\\ \omega(m)\equiv 1 \pmod 2}} \frac{1}{3m} \ge \frac{S_0}{3} \gg \log x,$$

so we are done with the conditional lowerbound.

4.2 An unconditional lower bound

We use an average version of the Bateman–Horn Conjecture due to Chudakov. Recall the following from Section 2.7 Chudakov, that

$$T_h(x) = \sum_{n+h\le x} \frac{1}{\log n \log(n+h)},$$

and that $T_h(x) \asymp x/(\log x)^2$ uniformly for $h = o(x)$ as $x \to \infty$. Now we use the following estimate:

$$\sum_{2|h} |\pi_2(x,h) - c_h T_h(x)|^2 \ll_A \frac{x^3}{(\log x)^A}.$$

To apply this result, we take $A = 10$ and let $\mathcal{H}$ be the set of even $h < x/20$ such that $\pi_2(x,h) < (c_h/2)T_h(x)$. Then

$$\sum_{h\in\mathcal{H}} (c_h T_h(x))^2 \ll \frac{x^3}{(\log x)^{10}}.$$

Note we could have taken any $A > 0$, but a smaller A results in a larger amount of numbers which we remove. Since $T_h(x) \gg x/(\log x)^2$ for $h \le x/20$ and $c_h \gg 1$, it follows that the left–hand side of the above sum is

$$\gg \#\mathcal{H}\left(\frac{x}{(\log x)^2}\right)^2 \gg \#\mathcal{H}\frac{x^2}{(\log x)^4}.$$

Thus,

$$\#\mathcal{H}\frac{x^2}{(\log x)^4} \ll \frac{x^3}{(\log x)^{10}},$$

implying that $\#\mathcal{H} \ll x/(\log x)^6$. Thus, the set of even integers $h \le x/20$ such that $\pi_2(x,h) \gg c_h T_h(x)$ is of cardinality $x/20 - O(x/(\log x)^6)$. In particular, this set contains most numbers of the form $h = Q+3$, where $Q \in (x/40, x/20]$ is a prime. Consider pairs of numbers of the form (Q,p), where Q is prime, $h = Q+3 \in (x/40, x/20]\backslash\mathcal{H}$ and p is a prime counted by $\pi_2(x,h)$. The number of such pairs is

$$\begin{aligned}
&\ge \sum_{\substack{x/40 \le Q+3 \le x/20 \\ Q+3 \notin \mathcal{H}}} \pi_2(x, Q+3) \\
&\gg \sum_{\substack{x/40 \le Q+3 \le x/20 \\ Q+3 \notin \mathcal{H}}} c_{Q+3} T_{Q+3}(x) \\
&\gg \sum_{\substack{x/40 \le Q+3 \le x/20 \\ Q+3 \notin \mathcal{H}}} \frac{x}{(\log x)^2} \\
&\gg \frac{x}{(\log x)^2}\left(\pi\left(\frac{x}{20}\right) - \pi\left(\frac{x}{40}\right) - O\left(\frac{x}{(\log x)^6}\right)\right) \gg \frac{x^2}{(\log x)^3}.
\end{aligned}$$

Here we used

$$\left(\pi(y) - \pi\left(\frac{y}{2}\right)\right) \gg \frac{y}{\log y}.$$

For each one of these pairs, $n = 3pQ \le x^2$ satisfies that $\beta(n) = p+Q+3 = p+h$ is a prime. Thus, $n \in \mathcal{B}(x^2)$. Further, since P and Q are large, it follows that each n can appear at most twice in the above count (once from $\pi_2(x, Q+3)$ and once from $\pi_2(x, p+3)$). Thus, the number of such distinct $n \le x^2$ is $\gg x^2/(\log x)^3 \gg x^2/(\log(x^2))^3$. Replacing x by $\sqrt{x}$ in the above argument we get the desired lower bound. This finishes the proof of the lower bound in Theorem 1.1.

4.3 A further conjecture on the count of $\mathcal{B}(x)$

In this section, we offer the following conjecture.

Conjecture 4.1. *We have*

$$\#\mathcal{B}(x) = e^{\gamma}(1+o(1))\frac{x}{\log x} \qquad as \qquad x \to \infty.$$

In the previous sections, we already proved that $\#\mathcal{B}(x) \ll x/\log x$ unconditionally, and $\#\mathcal{B}(x) \gg x/\log x$ conditionally under the Bateman–Horn Conjecture for all pairs of polynomials $X, X+h$ for even positive integers $h \le x^{\delta}$ for some $\delta > 0$. So, it is natural to conjecture that

$$\#\mathcal{B}(x) = (1+o(1))cx/\log x \qquad \text{holds as} \qquad x \to \infty \tag{4.1}$$

with some constant $c > 0$ and it remains to offer some guess for c. We will, assume that (4.1) holds. We shall estimate the sum

$$\sum_{n \in \mathcal{B}(x)} \frac{1}{n} \tag{4.2}$$

in two ways, as follows. In the first way, by estimate (4.1) and the Abel summation formula (this can been seen in Note 2.2), expression (4.2) should be asymptotically $c(1+o(1))\log\log x$ as $x \to \infty$. Secondly, let us assume that $\beta(n)$ is "randomly distributed" and as such the probability of it to be prime is $1/\log\beta(n)$. Then the sum w.r.t. primality (4.2) should be

$$\sum_{2 \le n \le x} \frac{1}{n\log\beta(n)}. \tag{4.3}$$

We have not seen the above sum evaluated in the literature, but we have seen the one for which $\beta(n)$ is replaced by $P(n)$. So, let compare them. We have $P(n) \le \beta(n) \le P(n)\omega(n)$. Let $s(n) := \log\beta(n) - \log P(n)$. If p_k denotes the kth prime, then $p_k > k$, so $P(n) \ge p_{\omega(n)} > \omega(n)$ holds for all $n \ge 2$. Thus, $s(n) \le \log\omega(n) \le \log P(n)$. Hence,

$$\begin{aligned}\frac{1}{\beta(n)} &= \frac{1}{\log P(n) + s(n)} = \frac{1}{\log P(n)}\left(\frac{1}{1+s(n)/\log P(n)}\right)\\ &= \frac{1}{\log P(n)} + O\left(\frac{s(n)}{(\log P(n))^2}\right) = \frac{1}{\log P(n)} + O\left(\frac{\log\omega(n)}{(\log P(n))^2}\right).\end{aligned}$$

In the above, we used the fact that

$$\frac{1}{1+z} = 1 + O(z) \qquad \text{for} \qquad |z| < 1,$$

with $z := s(n)/\log P(n)$. Thus,

$$\sum_{2\le n\le x}\frac{1}{n\log\beta(n)}=\sum_{2\le n\le x}\frac{1}{n\log P(n)}+O\left(\sum_{n\le x}\frac{\log\omega(n)}{n(\log P(n))^2}\right).$$

The first sum above was evaluated in [8]:

$$\sum_{2\le n\le x}\frac{1}{n\log P(n)}=e^{\gamma}\log\log x+O(1).$$

It remains to bound the sum inside the O term. For this, we break it into two parts, namely the part for which $\omega(n)$ is large and the part for which $\omega(n)$ is small. Let

$$\mathcal{A}_1=\{n:\omega(n)\ge 10\log\log x\}.$$

Recall that $\mathcal{A}(t)=\mathcal{A}\cap[1,t]$. Put $\tau(n)$ for the number of divisors of n. Since $\tau(n)\ge 2^{\omega(n)}\ge 2^{10\log\log x}>(\log x)^6$, we have that

$$\#\mathcal{A}(t)(\log x)^6\le\sum_{n\in\mathcal{A}(t)}\tau(n)\ll t\log t\ll t\log x.$$

The last inequality above we used Theorem 2.10. Thus,

$$\#\mathcal{A}(t)\ll\frac{t}{(\log x)^5}\qquad\text{holds for all}\qquad t\le x.$$

By the Abel summation formula, we have that

$$\sum_{\substack{n\le x\\ n\in\mathcal{A}}}\frac{1}{n}\le\frac{\#\mathcal{A}(t)}{t}\Big|_{t=2}^{t=x}+\int_2^x\frac{\#\mathcal{A}(t)dt}{t^2}\ll\frac{1}{(\log x)^5}\left(1+\int_2^x\frac{dt}{t}\right)\ll\frac{1}{(\log x)^4}.$$

Since $\omega(n)\ll\log n/\log\log n$, a estimate mentioned already during the proof of the upper bound for $\#\mathcal{B}_4(x)$, we have that $\log\omega(n)\le\log\log x$ holds for all $n\le x$ once x is sufficiently large. Thus,

$$\sum_{\substack{n\le x\\ n\in\mathcal{A}}}\frac{\log\omega(n)}{n(\log P(n))^2}\ll(\log\log x)\sum_{\substack{n\le x\\ n\in\mathcal{A}}}\frac{1}{n}\ll\frac{\log\log x}{(\log x)^4}=o(1)$$

as $x\to\infty$. Assume now that $n\le x$ is not in $\mathcal{A}$. We then have that $\omega(n)\le 10\log\log x$, so $\log\omega(n)\ll\log\log\log x$. Thus,

$$\sum_{\substack{2\le n\le x\\ n\notin\mathcal{A}}}\frac{\log\omega(n)}{n(\log P(n))^2}\ll(\log\log\log x)\sum_{2\le n\le x}\frac{1}{n(\log P(n))^2}.$$

The last sum on the right is $O(1)$. This follows easily by the Abel summation formula and the fact that

$$\sum_{2\le n\le x} \frac{1}{(\log P(n))^2} = (c_1 + o(1))\frac{x}{(\log x)^2} \qquad \text{as} \qquad x \to \infty,$$

which is a result of Wheeler from [15]. Further, $c_1 = \int_0^\infty \rho(t)(t+1)dt$, where $\rho(t)$ is the Dickman–de Brujin function mentioned in the upper bound for $\#\mathcal{B}_1(x)$ (for this, make $u := 1$, $\alpha := -2$ on the last display on page 516 in [15]). It thus follows that

$$\sum_{2\le n\le x} \frac{\log\omega(n)}{n(\log P(n))^2} = \sum_{\substack{n\le x\\ n\in\mathcal{A}}} \frac{\log\omega(n)}{n(\log P(n)^2} + \sum_{\substack{2\le n\le x\\ n\notin\mathcal{A}}} \frac{\log\omega(n)}{n(\log P(n))^2} \ll \log\log\log x,$$

so

$$\begin{aligned}\sum_{2\le n\le x} \frac{1}{n\log\beta(n)} &= \sum_{n\le x} \frac{1}{\log P(n)} + O\left(\sum_{n\le x} \frac{\log\omega(n)}{n(\log P(n))^2}\right)\\ &= e^\gamma \log\log x + O(\log\log\log x).\end{aligned}$$

Since the sums (4.2) and (4.3) should asymptotically be the same under the assumption that $\beta(n)$ is "randomly distributed", we conclude that if (4.1) holds, then we must indeed have $c = e^\gamma$.

We tested this computationally. We computed $r(x) := \#\mathcal{B}'(x)/\pi(x)$, where $\mathcal{B}'(x) = \{n \le x : \omega(n) \ge 2 \text{ and } \beta(n) \text{ is prime}\}$ for $x := 10^k$, and $k = 3, 4, 5, 6, 7, 8, 9$ getting the values from the table below.

k	$x = 10^k$	$\#\mathcal{B}'(x)$	$\pi(x)$	$r(x)$
3	10^3	130	168	0.77381
4	10^4	1196	1229	0.973149
5	10^5	11698	9592	1.21956
6	10^6	107315	78498	1.3671
7	10^7	961924	664579	1.44742
8	10^8	8641491	5761455	1.49988
9	10^9	78304633	50847534	1.53999
10	10^{10}	714962670	455052511	1.57117

This sequence $\{r(10^k)\}_{k\ge3}$ in the right–most column above seems to increase with k and it is anyone's guess whether by continuing the calculations one would get a limit of $e^\gamma = 1.78107\ldots$

Here is a list of some of the composite numbers which have this property of $\beta(n)$ with $\omega(n) > 1$ being prime: 6, 10, 12, 18, 20, 22, 24, 34, 36, 40, 44, 48, 50, 54, 58, 68, 72, 80, 82, 88, 96, 100, 108, 116, 118, 136, 142, 144, 160, 162, 164, 165, 176, 192, 200, 202, 210, 214, 216, 232, 236, 242, 250.

5 Remarks

It is found that 210 is the smallest n such that $\beta(34) = 19$, which takes 2 primes to write 2 and 17. If $n = 165$, then $\beta(165) = 19$ takes takes 3 primes to write namely $3, 5$ and 11. A further study can be done to see how many primes it takes to write a prime as a sum of primes such that their product is the smallest possible n, where n is composite. Another possible study is to count the number of integers n such that $\beta(n) \mid \phi(n)$ where $\phi(n)$ is the Euler Phi function of n.

There is room for other studies which this canidate plans to pursue in the near future.

References

[1] R. Alweiss, E. Chen, and S. Luo. Vinogradov's Three Prime Theorem. 2016.

[2] A. Balog. The prime k-tuplets conjecture on average. In *Berndt B.C., Diamond H.G., Halberstam H., Hildebrand A. (eds) Analytic Number Theory. Progress in Mathematics*, volume 85, pages 47–75. Springer, 1990.

[3] P. T. Bateman and R. A. Horn. A heuristic asymptotic formula concerning the distribution of prime numbers. *Mathematics of Computation*, 16(79):363–367, 1962.

[4] M. Bortolomasi and A. Ortiz-Tapia. Some remarks on the first Hardy-Littlewood Conjecture. *arXiv preprint arXiv:1904.03017*, 2019.

[5] J-M. De Koninck and F. Luca. Integers divisible by the sum of their prime factors. *Mathematika*, 52(1-2):69–77, 2005.

[6] J-M. De Koninck and F. Luca. Integers divisible by sums of powers of their prime factors. *Journal of Number Theory*, 128(3):557–563, 2008.

[7] J-M. De Koninck and F. Luca. *Analytic Number Theory: Exploring the Anatomy of Integers*, volume 134. American Mathematical Soc., 2012.

[8] J-M. De Koninck and R. Sitaramachandrarao. Sums involving the largest prime divisor of an integer. *Acta Arithmetica*, 48(1):1–8, 1987.

[9] P. Erdős and C. Pomerance. On the largest prime factors of n and n+1. *Aequationes Mathematicae*, 17(1):311–321, 1978.

[10] G. H. Hardy and J. E. Littlewood. Some problems of 'partitio numerorum'; iii: On the expression of a number as a sum of primes. *Acta Mathematica*, 44:1–70, 1923.

[11] G. H. Hardy and E. Wright. *"An Introduction to the Theory of Numbers"*. Oxford University Press, 1979.

[12] F. Luca and D. Moodley. Composite positive integers whose sum of prime factors is prime. *Archivum Mathematicum*, 56(1):49–64, 2020.

[13] C. Nelson, D.E. Penney, and C. Pomerance. 714 and 715. *Journal of Recreational Mathematics*, 7(2):87, 1974.

[14] C. Pomerance. Ruth–Aaron numbers revisited. *Paul Erdős and his Mathematics, Bolyai Soc. Math. Stud*, 11:567–579, 2002.

[15] F. S. Wheeler. Two differential-difference equations arising in number theory. *Transactions of the American Mathematical Society*, 318(2):491–523, 1990.