SIGN PATTERNS OF THE LIOUVILLE AND MÖBIUS FUNCTIONS

KAISA MATOMÄKI, MAKSYM RADZIWIŁŁ, AND TERENCE TAO

Abstract. Let λ and μ denote the Liouville and Möbius functions respectively. Hildebrand showed that all eight possible sign patterns for $(\lambda(n), \lambda(n+1), \lambda(n+2))$ occur infinitely often. By using the recent result of the first two authors on mean values of multiplicative functions in short intervals, we strengthen Hildebrand's result by proving that each of these eight sign patterns occur with positive lower natural density. We also obtain an analogous result for the nine possible sign patterns for $(\mu(n), \mu(n+1))$. A new feature in the latter argument is the need to demonstrate that a certain random graph is almost surely connected.

1. Introduction

In this paper we strengthen some results on the sign patterns of the Liouville function λ, as well as obtain new results on the Möbius function μ.

We begin with the Liouville function. It will be convenient (particularly in the combinatorial arguments used to prove our main theorems) to introduce the following notation.

Definition 1.1 (Liouville sign pattern)**.** Let k, l be non-negative integers, and let n be an integer. Let $\epsilon_{-k} \dots \epsilon_l$ be a string of $k+l+1$ symbols from the alphabet $\{+, -, *\}$. We write

$$n \mapsto \epsilon_{-k} \cdots \overset{\vee}{\epsilon_0} \dots \epsilon_l \qquad (1.1)$$

if $n > k$, $\lambda(n+i) = +1$ for all $-k \leq i \leq l$ with $\epsilon_i = +$, and $\lambda(n+i) = -1$ for all $-k \leq i \leq l$ with $\epsilon_i = -$. We write the negation of (1.1) as

$$n \not\mapsto \epsilon_{-k} \cdots \overset{\vee}{\epsilon_0} \dots \epsilon_l$$

The symbol $\vee$ is only present in the above notation as a positional marker (analogous to a decimal point in decimal notation) and has no further significance.

Example 1.2. The claim

$$n \mapsto -*\overset{\vee}{+}+$$

is equivalent to the assertion that $n > 2$, $\lambda(n-2) = -1$, $\lambda(n) = +1$, and $\lambda(n+1) = +1$, but makes no claim about the value of $\lambda(n-1)$.

In this notation, a well-known conjecture of Chowla [4] can now be phrased as follows:

Conjecture 1.3 (Chowla)**.** *For any $k \geq 1$ and signs $\epsilon_1, \dots, \epsilon_k \in \{-1, +1\}$, the set of natural numbers n for which*

$$n \mapsto \overset{\vee}{\epsilon_1} \dots \epsilon_k$$

has natural density $\frac{1}{2^k}$ (that is, the density of this set in $[1, x]$ converges to $1/2^k$ in the limit $x \to \infty$).

For $k = 1$, this claim is equivalent to the prime number theorem, but for $k > 1$ Chowla's conjecture remains open. For $k \leq 3$, we have the following partial result of Hildebrand [11]:

Theorem 1.4 (Hildebrand)**.** *For any $k = 1, 2, 3$ and signs $\epsilon_1, \dots, \epsilon_k \in \{-, +\}$, the claim*

$$n \longmapsto \overset{\vee}{\epsilon_1} \dots \epsilon_k$$

occurs for infinitely many n.

Hildebrand's method was elementary, relying on an *ad hoc* combinatorial analysis that relied primarily on the multiplicative properties of λ at the small primes $2, 3, 5$. In this paper, we combine the methods of Hildebrand with a recent result of the first two authors [12] to improve this result as follows.

Definition 1.5. A property $P(n)$ of an integer n is said to hold with *positive lower natural density* if

$$\liminf_{x \to \infty} \frac{1}{x} \sum_{n \leq x : P(n)} 1 > 0.$$

Theorem 1.6 (Liouville patterns of length three)**.** *For any $k = 1, 2, 3$ and signs $\epsilon_1, \dots, \epsilon_k \in \{-, +\}$, the claim*

$$n \longmapsto \overset{\vee}{\epsilon_1} \dots \epsilon_k$$

occurs with positive lower natural density.

As with Hildebrand's arguments, our arguments extend to other completely multiplicative functions f taking values in $-1, +1$ than λ, provided that f agrees with λ at the primes $2, 3, 5$ and obeys a prime number theorem in arithmetic progressions for any modulus dividing 60. We leave the details of this generalisation to the interested reader.

As it turns out, the most difficult sign patterns to handle for Theorem 1.6 are $+++$ and $---$. The problem is that the Liouville function $\lambda(n)$ could potentially behave like $f(n)\chi_3(n)$ for almost all n that are not multiples of 3, where χ_3 is the primitive Dirichlet character of conductor 3 (thus $\chi_3(3n+1) = 1$ and $\chi_3(3n+2) = -1$ for all n) and $f : \mathbb{N} \to \{-1, +1\}$ is a function that changes sign very rarely. In such a case, the sign patterns $+++$ and $---$ will almost never occur. Fortunately, the results in [12] preclude this scenario; the main difficulty is then to show that this is essentially the *only* scenario that could eliminate the $+++$ or $---$ patterns almost completely.

Remark 1.7. Strictly speaking, our arguments do not yield an explicit bound on the lower natural density, because we rely on Banach limits to simplify the presentation of the argument. However, we believe that one could extract an effective lower bound on the density if required by avoiding the use of Banach limits, and keeping track of all error terms without passing to an asymptotic limit.

Remark 1.8. In [3] it was shown that $(\lambda(n), \lambda(n+r), \lambda(n+2r), \lambda(n+3r))$ attains all sixteen sign patterns in $\{-1, +1\}^4$ infinitely often, if n ranges over the natural numbers and r ranges over a bounded set. It is plausible that using arguments similar to the ones here, one can show that these sixteen sign patterns also occur for n in a set of positive lower density. Note also from recent results on linear equations in the Liouville function (see [6, Proposition 9.1], together with the companion results in [7] and [8]) that for any fixed $k \geq 1$, the tuple $(\lambda(n), \dots, \lambda(n + (k-1)r))$ is asymptotically equidistributed in $\{-1,+1\}^k$ if n, r range uniformly over (say) $[x, 2x]$ for some large x going to infinity; see also [5] for a recent generalization of these sorts of results to other linear forms and to more general bounded multiplicative functions. It may be, in light of the results in [12], that one can obtain a similar equidistribution result with r restricted to a much smaller range that grows arbitrarily slowly with x, but we do not pursue this matter here.

We now turn to the Möbius function μ. This function takes values in $\{-1, 0, +1\}$ rather than $\{-1, 1\}$, and the presence of the additional 0 significantly complicates the analysis. Nevertheless, we can show an analogue for Theorem 1.6 for $k = 1, 2$ only:

Theorem 1.9 (Möbius patterns of length two)**.** *For any $k = 1, 2$ and $\epsilon_1, \dots, \epsilon_k \in \{-1, 0, +1\}$, the claim*

$$(\mu(n), \dots, \mu(n + k - 1)) = (\epsilon_1, \dots, \epsilon_k)$$

occurs with positive lower natural density.

The difficult case of this theorem ocurs when $k = 2$ and $\epsilon_1, \epsilon_2 \in \{+1, -1\}$. Suppose for instance that we wanted to show that $(\mu(n), \mu(n + 1)) = (+1, -1)$ occurred with positive lower natural density. In the case of the Liouville function, one could show (as was done in [12]) that if the analogous claim $(\lambda(n), \lambda(n + 1)) = (+1, -1)$ occurs with zero density, then there would often exist long chains $n, n + 1, \dots, n + h$ of consecutive natural numbers on which λ was constant, which is in contradiction with the results in [12]. This argument no longer works in the case of the Möbius function, due to the large number of zeroes of this function (for instance, the zeroes at the multiples of four). To get around this, one needs to replace the chains $n, n + 1, \dots, n + h$ by more complicated paths of nearby natural numbers, necessitating the analysis of the connectivity properties of a certain random graph, which will be done in Sections 6, 7. The further study of this random graph (or similar such graphs) may have some further applications; in particular, one may hope to use expansion properties of this graph to make progress towards the $k = 2$ case of the Chowla conjecture (Conjecture 1.3).

2. Asymptotic probability

When dealing with natural densities, there is the minor technical difficulty that the lower and upper natural densities for a given set of integers need not match, leading to a breakdown of additivity of density in these situations. To get around this problem we shall use the (artificial) device of Banach limits.

Let $P_0(n)$ be a property of the natural numbers that holds with zero lower natural density. In particular, we can find a sequence x_i of real numbers going to infinity such that

$$\frac{1}{x_i} \sum_{n \leq x_i : P_0(n)} 1 \leq 2^{-i}$$

for all i. This implies that

$$\sum_{\frac{1}{i} x_i \leq n \leq x_i : P_0(n)} \frac{1}{n} \leq i 2^{-i}.$$

In particular

$$\frac{1}{\log i} \sum_{\frac{1}{i} x_i \leq n \leq x_i : P_0(n)} \frac{1}{n} \to 0$$

as $i \to \infty$. Henceforth we fix the sequence x_i with this property.

Next, define a *Banach limit* to be a linear functional $\mathrm{LIM} : \ell^\infty(\mathbb{N}) \to \mathbb{R}$ from bounded sequences $(a_i)_{i=1}^\infty$ of real numbers to the real numbers with the property that

$$\liminf_{i\to\infty} a_i \le \mathrm{LIM}(a_i)_{i=1}^\infty \le \limsup_{i\to\infty} a_i$$

for all bounded sequences $(a_i)_{i=1}^\infty$; in particular, $\mathrm{LIM}(a_i)_{n=1}^\infty = \lim_{i\to\infty} a_i$ when a_i is convergent. As is well known, the existence of a Banach limit is guaranteed by the Hahn-Banach theorem, or by the existence of non-principal ultrafilters on the natural numbers. Henceforth we will fix[1] a single Banach limit LIM.

Given a property $P(n)$ of an integer n, we define the *asymptotic probability* that $P(n)$ occurs to be the quantity

$$\mathrm{LIM}\left(\frac{1}{\log i}\sum_{\frac{1}{i}x_i \le n \le x_i : P(n)} \frac{1}{n}\right)_{i=1}^\infty,$$

and say that $P(n)$ holds *asymptotically almost surely* (or *a.a.s.* for short) if its asymptotic probability is 1. Thus, for instance, with LIM and x_i as above, the property $P_0(n)$ is asymptotically almost surely false.

Note that as LIM is linear, asymptotic probability obeys all the axioms of finitely additive probability. For instance, if $P(n)$ and $Q(n)$ are properties with Q holding asymptotically almost surely, then the asymptotic probability of $P(n)$ is equal to the asymptotic probability of $P(n) \wedge Q(n)$.

We will call a quantity *fixed* if it does not depend on the parameter i. Since $x_i \to \infty$, it is clear that for any fixed constant $C > 0$, one has $n > C$ a.a.s.. (This does not contradict the finite nature of n, because our probability measure is only finitely additive, rather than countably additive.)

In view of the above construction, we see that Theorem 1.6 follows from

Theorem 2.1. *Fix a sequence $x_i \to \infty$ and a Banach limit* LIM*. For any $k = 1, 2, 3$ and signs $\epsilon_1, \dots, \epsilon_k \in \{-, +\}$, the claim*

$$n \longmapsto \overset{\vee}{\epsilon}_1 \dots \epsilon_k \tag{2.1}$$

occurs with positive asymptotic probability.

Indeed, if (2.1) occurs with zero natural density, one obtains a contradiction to Theorem 2.1 after selecting x_i as discussed in the start of the section, with P_0 taken to be the property (2.1). Similarly, Theorem 1.9 follows from

Theorem 2.2. *Fix a sequence $x_i \to \infty$ and a Banach limit* LIM*. For any $k = 1, 2$ and $\epsilon_1, \dots, \epsilon_k \in \{-1, 0, +1\}$, the claim*

$$(\mu(n), \dots, \mu(n+k-1)) = (\epsilon_1, \dots, \epsilon_k) \tag{2.2}$$

occurs with positive asymptotic probability.

[1] Strictly speaking, this means that we are assuming the axiom of choice (or at least the ultrafilter lemma) in our arguments. However, if desired, it is a routine matter to rewrite the arguments below in terms of limit superior and limit inferior instead of Banach limits and replacing additivity by subadditivity or superadditivity as appropriate, so that the axiom of choice is no longer required. We leave the details of this modification of the argument to the interested reader.

Henceforth the sequence $x_i \to \infty$ and Banach limit LIM will be fixed, and this fact will be omitted from the explicit formulation of all the propositions below. (Thus, for instance, with this convention the first sentence of Theorems 2.1 and 2.2 would be deleted.)

We now set out some basic rules for manipulating asymptotic probability, beyond the laws of finitely additive probability. The first rule allows one to make linear changes of variable.

Lemma 2.3 (Linear change of variable)**.** *Let $P(n)$ be a property of integers n. Then for any natural number q and integer r, the asymptotic probability of $P(qn+r)$ is equal to q times the asymptotic probability of "$P(n)$ and $n = r \mod q$".*

Note that the $q = 1$ case of Lemma 2.3 gives translation invariance: the asymptotic probability of $P(n)$ is equal to the asymptotic probability of $P(n+h)$ for any fixed h. Another useful corollary of Lemma 2.3 is that if $r_1, \dots, r_q$ is any fixed set of representatives of residues modulo q, then $P(n)$ holds a.a.s. if and only if $P(qn+r)$ holds a.a.s. for each $r = r_1, \dots, r_q$.

Proof. Let C be a large constant depending on q, r. By deleting an event of asymptotic probability zero, we may assume that $P(n)$ only holds for $n > C$. For sufficiently large C, this implies that if $P(qn+r)$ holds, then $\frac{1}{n} = (1+o(1))\frac{q}{qn+r}$ where $o(1)$ goes to zero as $C \to \infty$, and thus on making the change of variables $m = qn + r$,

$$\sum_{\substack{\frac{1}{i}x_i \leq n \leq x_i \\ P(qn+r)}} \frac{1}{n} = (1+o(1))q \sum_{\substack{\frac{q}{i}x_i + r \leq m \leq qx_i + r \\ P(m) \text{ and } m = r \bmod q}} \frac{1}{m}.$$

Also, we have

$$\sum_{\frac{q}{i}x_i + r \leq m \leq qx_i + r : P(m) \text{ and } m = r \bmod q} \frac{1}{m} = \sum_{\frac{1}{i}x_i \leq m \leq x_i : P(m) \text{ and } m = r \bmod q} \frac{1}{m} + O_{C,q}(1)$$

for a quantity $O_{C,q}(1)$ bounded in magnitude by a function of C and q. Dividing by $\log i$, we conclude that

$$\frac{1}{\log i} \sum_{\frac{1}{i}x_i \leq n \leq x_i : P(qn+r)} \frac{1}{n} = (1+o(1))q \frac{1}{\log i} \sum_{\frac{1}{i}x_i \leq n \leq x_i : P(n)} \frac{1}{n} + \frac{O_{C,q}(1)}{\log i},$$

and on taking Banach limits and then sending C to infinity, we obtain the claim. □

Now we encode some known facts about the Liouville and Möbius function in this language. From the prime number theorem in arithmetic progressions we have

$$\lim_{x\to\infty} \frac{1}{x} \sum_{n \leq x} \lambda(qn+r) = 0$$

for any $q \geq 1$ and $r \in \mathbb{Z}$, and thus by summation by parts

$$\lim_{x\to\infty} \frac{1}{\log x} \sum_{n \leq x} \frac{\lambda(qn+r)}{n} = 0.$$

Similarly for the Möbius function. We conclude:

Proposition 2.4. *Let $q \geq 1$ and $r \in \mathbb{Z}$. Then the assertions $\lambda(qn+r) = +1$ and $\lambda(qn+r) = -1$ (or in the notation of this paper, $qn+r \mapsto \overset{\vee}{+}$ and $qn+r \mapsto \overset{\vee}{-}$) each occur with asymptotic probability $1/2$. Also, the assertions $\mu(qn+r) = +1$ and $\mu(qn+r) = -1$ occur with equal asymptotic probability.*

This immediately gives the $k = 1$ case of Theorem 2.1. Since μ^2 has density

$$\frac{1}{\zeta(2)} = \frac{6}{\pi^2} = 0.6079\ldots,$$

we also know that $\mu^2(n) = 1$ with asymptotic probability $\frac{1}{\zeta(2)}$, and hence by the above proposition the three events $\mu(n) = +1, \mu(n) = 0, \mu(n) = -1$ occur with asymptotic probability $\frac{1}{2\zeta(2)}$, $1 - \frac{1}{\zeta(2)}$, $\frac{1}{2\zeta(2)}$ respectively. Thus we also obtain the $k = 1$ case of Theorem 2.2.

From Proposition 2.4 and the Chinese remainder theorem, we also see that for any fixed w, the asymptotic probability that $\mu(n+1) = +1$ and n is not divisible by p^2 for any $p \leq w$ is equal to the asymptotic probability that $\mu(n+1) = -1$ and n is not divisible by p^2 for any $p \leq w$. Taking limits as $w \to \infty$ (noting that the asymptotic probability that n is divisible by p^2 for some $p > w$ is $O(1/w)$), one concludes that the the pair $(\mu(n), \mu(n+1))$ takes the values $(0, +1)$ and $(0, -1)$ with equal asymptotic probability, and similarly takes the values $(+1, 0)$ and $(-1, 0)$ with equal asymptotic probability. Also, standard sieve theory arguments also show that the event $\mu^2(n) = \mu^2(n+1) = +1$ (which is excluding two residue classes modulo p^2 for each prime p) occurs with asymptotic probability

$$c := \prod_p \left(1 - \frac{2}{p^2}\right) = 0.3226\ldots \tag{2.3}$$

and hence by inclusion-exclusion, $(\mu(n), \mu(n+1))$ takes the value $(0, 0)$ with asymptotic probability

$$1 - \frac{2}{\zeta(2)} + c = 0.1067\ldots.$$

Further inclusion-exclusion then shows $(\mu(n), \mu(n+1))$ takes each of the four values $(+1, 0)$, $(-1, 0)$, $(0, +1)$, $(0, -1)$ with asymptotic probability

$$\frac{1}{4}\left(\frac{1}{\zeta(2)} - c\right) = 0.0713\ldots.$$

This gives all the cases of Theorem 2.2 except for those in which $k = 2$ and $(\epsilon_1, \epsilon_2) = (+1, +1), (+1, -1), (-1, +1), (-1, -1)$, the treatment of which we defer to Sections 5-7.

Recently, the first two authors [12] established (among other things) that

$$\limsup_{X \to \infty} \frac{1}{X} \sum_{X \leq x \leq 2X} \left| \frac{1}{h} \sum_{x \leq n \leq x+h} \lambda(n) \right| \leq c(h)$$

for any $h \geq 1$ some quantity $c(h)$ that goes to zero as $h \to \infty$, and similarly with λ replaced by μ. From summation by parts this implies that

$$\limsup_{x \to \infty} \frac{1}{\log x} \sum_{n \leq x} \frac{1}{n} \left| \frac{1}{h} \sum_{j=0}^{h} \lambda(n+j) \right| \leq c(h)$$

and thus

Theorem 2.5 (Liouville or Möbius in short intervals)**.** *For any $\varepsilon > 0$ and any h that is sufficiently large depending on ε, one has*

$$\left|\sum_{j=0}^{h} \lambda(n+j)\right| \leq \varepsilon h$$

with asymptotic probability at least $1-\varepsilon$. Similarly with λ replaced by μ.

The above results also hold when the Liouville function is twisted by a fixed real Dirichlet character:

Theorem 2.6 (Twisted Liouville or Möbius in short intervals)**.** *Let χ be a fixed real Dirichlet character. For any $\varepsilon > 0$ and any sufficiently large h, one has*

$$\left|\sum_{j=0}^{h} \lambda(n+j)\chi(n+j)\right| \leq \varepsilon h$$

with asymptotic probability at least $1-\varepsilon$. Similarly with λ replaced by μ.

This result was recently extended to the case of complex Dirichlet characters in [13, Appendix A], but we will not need that extension here. Indeed, we will only need Theorem 2.6 with χ the non-trivial character χ_3 of period 3.

As already essentially observed in [12], Theorem 2.5 gives the $k=2$ case of Theorem 2.1:

Theorem 2.7. *The assertions $n \mapsto \overset{\vee}{+}-$, $n \mapsto \overset{\vee}{-}+$ hold with positive asymptotic probability, and the assertions $n \mapsto \overset{\vee}{+}+$, $n \mapsto \overset{\vee}{-}-$ hold with asymptotic probability at least $1/6$ each.*

We remark that the second part of this theorem is essentially due to Harman, Pintz, and Wolke [10].

Proof. We begin with the sign pattern $\overset{\vee}{+}-$. Assume for contradiction that $n \mapsto \overset{\vee}{+}-$ holds with zero asymptotic probability. By Proposition 2.4 and Lemma 2.3, the claims $n \mapsto \overset{\vee}{+}*$ and $n \mapsto \overset{\vee}{*}-$ hold with asymptotic probability 1/2, so that $\overset{\vee}{+}+$ and $\overset{\vee}{-}-$ hold with asymptotic probability 1/2. Hence $n \mapsto \overset{\vee}{-}+$ holds with zero asymptotic probability. In other words, we have $\lambda(n+1) = \lambda(n)$ a.a.s., hence by Lemma 2.3 and finite additivity we have for any fixed h that $\lambda(n) = \lambda(n+1) = \cdots = \lambda(n+h)$ a.a.s.. But this contradicts Theorem 2.5 if h is large enough.

The same argument also works for the sign pattern $\overset{\vee}{-}+$. Now we consider the sign pattern $\overset{\vee}{+}+$. From Proposition 2.4 and inclusion-exclusion, the sign pattern $\overset{\vee}{-}-$ occurs with the same asymptotic probability as $\overset{\vee}{+}+$. Thus it suffices to show that $\lambda(n) = \lambda(n+1)$ holds with asymptotic probability at least 1/3. But from the pigeonhole principle, at least one of $\lambda(2n+1) = \lambda(2n)$, $\lambda(2n+2) = \lambda(2n+1)$, and $\lambda(2n) = \lambda(2n+2)$ must hold for any n, which implies that the asymptotic probabilities of $\lambda(2n+1) = \lambda(2n)$, $\lambda(2n+2) = \lambda(2n+1)$, and $\lambda(n+1) = \lambda(n)$ add up to at least 1. On the other hand, from Lemma 2.3 we see that the asymptotic probability of $\lambda(n+1) = \lambda(n)$ is the average of the asymptotic probabilities of $\lambda(2n+1) = \lambda(2n)$ and $\lambda(2n+2) = \lambda(2n+1)$, and the claim follows. □

As essentially already observed by Hildebrand [11], a simple translation argument then gives four of the eight subcases of the $k = 3$ case of Theorem 2.1:

Corollary 2.8. *The assertions $n \mapsto \overset{\vee}{+}+-$, $n \mapsto \overset{\vee}{-}++$, $n \mapsto \overset{\vee}{+}--$, $n \mapsto \overset{\vee}{-}-+$ each hold with positive asymptotic probability.*

Proof. We show this for the sign pattern $\overset{\vee}{+}+-$ only, as the other three cases are similar. Suppose for contradiction that $n \not\mapsto \overset{\vee}{+}+-$ a.a.s.. In particular, $n \mapsto \overset{\vee}{+}+$ implies $n+1 \mapsto \overset{\vee}{+}+$ a.a.s.. Iterating this (using the translation invariance from Lemma 2.3), we see that for any fixed $h \geq 1$, $n \mapsto \overset{\vee}{+}+$ implies that $\lambda(n) = \lambda(n+1) = \cdots = \lambda(n+h) = +1$ a.a.s., and hence by Theorem 2.7, one has $\lambda(n) = \lambda(n+1) = \cdots = \lambda(n+h) = +1$ with asymptotic probability at least c for some $c > 0$ independent of h. But this contradicts Theorem 2.5 if h is chosen sufficiently large. □

Remark 2.9. The same argument, in combination with the remaining $k = 3$ cases of Theorem 2.1 that we will handle shortly, show that the $k = 4$ case of Theorem 2.1 holds for the sign patterns $\overset{\vee}{+}++-$, $\overset{\vee}{-}+++$, $\overset{\vee}{-}--+$, and $\overset{\vee}{+}---$. However, there does not seem to be a similarly simple argument to handle the remaining twelve sign patterns of length 4.

It remains to verify the $k = 3$ cases of Theorem 2.1 for the sign patterns $n \mapsto \overset{\vee}{+}++$, $n \mapsto \overset{\vee}{-}--$, $n \mapsto \overset{\vee}{+}-+$, $n \mapsto \overset{\vee}{-}+-$. We now address these cases in the next two sections.

3. The $\overset{\vee}{+}++$ case

We now verify Theorem 2.1 for the sign pattern $n \mapsto \overset{\vee}{+}++$. The same argument (flipping all the signs) applies to the sign pattern $n \mapsto \overset{\vee}{-}--$, and is left to the reader[2].

Throughout this section we assume for sake of contradiction that Theorem 2.1 fails for $n \mapsto \overset{\vee}{+}++$. Thus we have

Hypothesis 3.1. *We have $n \not\mapsto \overset{\vee}{+}++$ a.a.s..*

As remarked in the introduction, Hypothesis 3.1 is morally compatible with λ "pretending" to be like the Dirichlet character χ_3 of conductor 3, defined by setting $\chi_3(3n+1) = +1$, $\chi_3(3n+2) = -1$, $\chi_3(3n) = 0$ for any n.

The strategy will be to leverage Hypothesis 3.1, together with Lemma 2.3 and the multiplicative nature of λ, to force λ to behave increasingly like χ_3 in various senses, until we can use Theorem 2.6 to obtain a contradiction.

We first give some simple consequences of Hypothesis 3.1 and Lemma 2.3.

Corollary 3.2. *Asymptotically almost surely in n, the following claims hold.*

(a) *If $n \mapsto \overset{\vee}{+}*+$, then $n \mapsto \overset{\vee}{+}-+$.*

[2] Alternatively, one can observe that the only properties of λ used in these arguments are those given by Proposition 2.4, Theorem 2.5, Theorem 2.6, that λ takes values in $\{-1, +1\}$, and that $\lambda(pn) = -\lambda(n)$ for all n and $p = 2, 3, 5$. All of these properties hold with λ replaced by $-\lambda$, and so the arguments in this section applied to $-\lambda$ handle the sign pattern $n \mapsto \overset{\vee}{-}--$. Similarly for the arguments in the next section.

(a.2) *If* $2n \mapsto \overset{\vee}{-}***-$, *then* $2n \mapsto \overset{\vee}{-}*+*-$.

(b) *If* $n \mapsto \overset{\vee}{+}+$, *then* $n \mapsto -\overset{\vee}{+}+-$.

(b.3) *If* $3n \mapsto \overset{\vee}{-}**-$, *then* $3n \mapsto +**\overset{\vee}{-}**-**+$.

(b.5) *If* $5n \mapsto \overset{\vee}{-}****-$, *then* $5n \mapsto +****\overset{\vee}{-}****-****+$.

(c) *If* $3n \mapsto \overset{\vee}{*}++$, *then* $3n \mapsto +**\overset{\vee}{-}++-**+$.

Proof. The claim (a) is immediate from Hypothesis 3.1. As $\lambda(2m) = -\lambda(m)$ for all m, that the claim (a.2) is equivalent to (a).

For (b), observe from Hypothesis 3.1 and Lemma 2.3 that $n \not\mapsto \overset{\vee}{+}++$ and $n \not\mapsto ++\overset{\vee}{+}$ a.a.s., giving the claim. The multiplicativity properties $\lambda(3m) = -\lambda(m)$ and $\lambda(5m) = -\lambda(m)$ then give the claims (b.3) and (b.5) respectively.

Now we prove (c). Suppose $3n \mapsto \overset{\vee}{*}++$. From (b) and Lemma 2.3, we then have $3n \mapsto \overset{\vee}{-}++-$ a.a.s.. The claim then follows from (b.3). □

The next implication is essentially due to Hildebrand [11].

Proposition 3.3. *One has* $15n \not\mapsto +\overset{\vee}{*}+$ *a.a.s..*

Proof. We will restrict to the event $15n \mapsto +\overset{\vee}{*}+$ and show that this leads a.a.s. to a contradiction, giving the claim.

By hypothesis and Corollary 3.2(a) we have $15n \mapsto +\overset{\vee}{-}+$ a.a.s.. Since $\lambda(4m) = \lambda(m)$, we thus have

$$60n \mapsto +***\overset{\vee}{-}***+$$

a.a.s.. Clearly this implies

$$60n \not\mapsto \overset{\vee}{+}**-++-**+$$

so by Corollary 3.2(c) in the contrapositive (and Lemma 2.3) we have

$$60n \not\mapsto \overset{\vee}{*}****++$$

and thus

$$60n \mapsto +***\overset{\vee}{-}***+-$$

a.a.s.. A similar argument gives

$$60n \not\mapsto ++***\overset{\vee}{*}$$

and thus

$$60n \mapsto -+***\overset{\vee}{-}***+-$$

a.a.s.. But from Corollary 3.2(b.5) and Lemma 2.3, we must then have

$$60n \mapsto +****\overset{\vee}{-}****-$$

a.a.s., giving the desired contradiction since $\lambda(60n-5)$ cannot simultaneously be $+1$ and -1. □

This gives us our first step towards demonstrating that λ "pretends to be like" χ_3:

Corollary 3.4. *We have* $\lambda(15n+1) = -\lambda(15n-1)$ *a.a.s..*

Proof. By Proposition 2.4, we have $\lambda(15n+1) = +1$ and $\lambda(15n-1) = +1$ with an asymptotic probability of $1/2$, while from Proposition 3.3 we have $\lambda(15n+1) = \lambda(15n-1) = +1$ with asymptotic probability zero. The claim then follows from the inclusion-exclusion principle. □

A variant of the above arguments gives

Proposition 3.5. *One has* $6n \not\mapsto -\overset{\vee}{+}-$ *a.a.s..*

Proof. As before, we restrict to the event $6n \mapsto -\overset{\vee}{+}-$ and show that this leads to a contradiction a.a.s..

The multiplicativity property $\lambda(5m) = -\lambda(m)$ for all m gives

$$30n \mapsto +****\overset{\vee}{-}****+.$$

In particular,

$$30n \not\mapsto \overset{\vee}{+}**-++-**+$$

and hence from Corollary 3.2(c) in the contrapositive (and Lemma 2.3) we have

$$30n \not\mapsto \overset{\vee}{*}***++$$

and hence

$$30n \mapsto +****\overset{\vee}{-}***-+$$

a.a.s.. A similar argument gives

$$30n \not\mapsto ++***\overset{\vee}{*}$$

and thus

$$30n \mapsto +-***\overset{\vee}{-}***-+$$

a.a.s. By two applications of Corollary 3.2(a.2) and Lemma 2.3, we then have

$$30n \mapsto +-*+*\overset{\vee}{-}*+*-+$$

a.a.s., and hence since $\lambda(2m) = -\lambda(m)$

$$\lambda(15n-1) = \lambda(15n+1) = -1,$$

but this contradicts Corollary 3.4 a.a.s.. □

Corollary 3.6. *One has* $3n \not\mapsto -\overset{\vee}{-}-$ *a.a.s..*

Proof. We restrict to the event that $3n \mapsto -\overset{\vee}{-}-$. The multiplicativity property $\lambda(2m) = -\lambda(m)$ then gives

$$6n \mapsto +*\overset{\vee}{+}*+$$

and hence by two applications of Corollary 3.2(a) and Lemma 2.3 we have

$$6n \mapsto +-\overset{\vee}{+}-+$$

a.a.s.. But this contradicts Proposition 3.5 a.a.s.. □

Corollary 3.7. *Asymptotically almost surely, we have* $\lambda(6n-1) = -\lambda(6n+1)$.

Proof. By Proposition 2.4, we have $\lambda(6n-1) = -1$ and $\lambda(6n+1) = -1$ with an asymptotic probability of $1/2$ each. From Proposition 3.5, Corollary 3.6, and Lemma 2.3 we see that $\lambda(6n-1) = \lambda(6n+1) = -1$ holds with asymptotic probability zero. The claim then follows from the inclusion-exclusion principle. □

Next, we work on improving Corollary 3.2(c).

Proposition 3.8. *One has* $9n+3 \not\mapsto \overset{\vee}{*}{+}{+}$ *a.a.s..*

Proof. We restrict to the event that $9n+3 \mapsto \overset{\vee}{*}{+}{+}$. By Corollary 3.2(c) and Lemma 2.3 we then have

$$9n+3 \mapsto {+}{*}{*}\overset{\vee}{-}{+}{+}{-}{*}{*}{+}$$

a.a.s.. Since $\lambda(2m) = -\lambda(m)$, we then have

$$18n+9 \mapsto {+}{*}{-}\overset{\vee}{*}{-}{*}{+}.$$

By Corollary 3.6 and Lemma 2.3 we thus have

$$18n+9 \mapsto {+}{*}{-}\overset{\vee}{+}{-}{*}{+}$$

a.a.s.; since $\lambda(3m) = -\lambda(m)$, we then have

$$6n+3 \mapsto {-}\overset{\vee}{-}{-}$$

which by Corollary 3.6 and Lemma 2.3 leads to a contradiction a.a.s.. □

Proposition 3.9. *Asymptotically almost surely, the claim* $9n \mapsto \overset{\vee}{*}{+}{+}$ *implies* $6n-3 \mapsto \overset{\vee}{*}{+}{+}$*, and similarly* $9n-3 \mapsto \overset{\vee}{*}{+}{+}$ *implies* $6n \mapsto \overset{\vee}{*}{+}{+}$*.*

Proof. Let us first restrict to the event that $9n \mapsto \overset{\vee}{*}{+}{+}$. By Corollary 3.2(c) and Lemma 2.3, we then have

$$9n \mapsto {+}{*}{*}\overset{\vee}{-}{+}{+}{-}{*}{*}{+}$$

a.a.s.. Since $\lambda(2m) = -\lambda(m)$, this implies

$$18n \mapsto {-}{*}{*}{*}{*}{*}\overset{\vee}{+}{*}{-}{*}{-}.$$

From Corollary 3.6 and Lemma 2.3 we a.a.s. have

$$18n \not\mapsto \overset{\vee}{*}{*}{-}{-}{-}$$

and hence

$$18n \mapsto {-}{*}{*}{*}{*}{*}\overset{\vee}{+}{*}{-}{+}{-};$$

since $\lambda(3m) = -\lambda(m)$, this implies that

$$6n-3 \mapsto \overset{\vee}{*}{+}{*}{-}{-}.$$

But from Corollary 3.6 and Lemma 2.3 we have

$$6n-3 \not\mapsto \overset{\vee}{*}{+}{-}{-}{-}$$

a.a.s., and hence $6n-3 \mapsto \overset{\vee}{*}{+}{+}$ a.a.s. as required.

Similarly, if we instead restrict to the event $9n-3 \mapsto \overset{\vee}{*}{+}{+}$, then Corollary 3.2(c) and Lemma 2.3 give

$$9n \mapsto {+}{*}{*}{-}{+}{+}\overset{\vee}{-}{*}{*}{+}$$

and then

$$18n \mapsto -*-*\overset{\vee}{+}*****-$$

and then by Corollary 3.6 and Lemma 2.3 as before

$$18n \mapsto -+-*\overset{\vee}{+}*****-$$

and thus

$$6n \mapsto --\overset{\vee}{*}+$$

and then by Corollary 3.6 and Lemma 2.3 we have $6n \mapsto \overset{\vee}{*}++$ as required. □

By combining the two preceding propositions with an iteration argument, we obtain

Corollary 3.10. *We have* $3n \not\mapsto \overset{\vee}{*}++$ *a.a.s..*

Proof. Let $f : \mathbb{N} \to \mathbb{N} \cup \{\perp\}$ be the partially defined function given by the formulae $f(3n) := 2n - 1$, and $f(3n-1) := 2n$ for $n \geq 2$, with all other values of f equal to the undefined symbol $\perp$. We can then combine Propositions 3.8, 3.9 to give the following assertion (by dividing into cases based on the residue of n modulo 3): if $3n \mapsto \overset{\vee}{*}++$, then a.a.s. $f(n) \neq \perp$ and $3f(n) \mapsto \overset{\vee}{*}++$. Iterating this, we conclude in particular that for any fixed k, we have a.a.s. that if $3n \mapsto \overset{\vee}{*}++$, then the sequence $n, f(n), f^2(n), \dots, f^k(n)$ avoids $\perp$. But a routine count shows that the event that $n, f(n), f^2(n), \dots, f^k(n)$ avoids $\perp$ occurs with asymptotic probability $(2/3)^{k+1}$. As k can be arbitrarily large, we obtain the claim. □

Corollary 3.11. *Asymptotically almost surely, we have* $\lambda(3n+1) = -\lambda(3n+2)$.

Proof. By Proposition 2.4, we have $\lambda(3n+1) = +1$ and $\lambda(3n+2) = +1$ with an asymptotic probability of $1/2$ each. From Corollary 3.10 we see that $\lambda(3n+1) = \lambda(3n+2) = +1$ holds with asymptotic probability 0. The claim then follows from the inclusion-exclusion principle. □

Corollary 3.12. *Asymptotically almost surely, we have* $\lambda(3n-1) = -\lambda(3n+1)$.

Proof. For n even, this follows from Corollary 3.7, so suppose that $n = 2N+1$. But then since $\lambda(2m) = -\lambda(m)$, the claim $\lambda(3n-1) = -\lambda(3n+1)$ in this case is equivalent to $\lambda(3N+1) = -\lambda(3N+2)$, and the claim follows from Corollary 3.11. □

We can combine Corollary 3.11, 3.12 using the Dirichlet character χ_3 to conclude that asymptotically almost surely, $\lambda\chi_3$ is constant on the set $\{3n-1, 3n+1, 3n+2\}$. Shifting n repeatedly by 1 using Lemma 2.3, we conclude that for any fixed $k \geq 1$, $\lambda\chi_3$ is a.a.s. constant on the set $\{3n-1, 3n+1, 3n+2, 3n+4, 3n+5, \dots, 3n+3k-1, 3n+3k+1, 3n+3k+2\}$, and in particular

$$|\sum_{j=0}^{3k} \lambda(3n+j)\chi_3(3n+j)| = 2k$$

a.a.s.. By Lemma 2.3 this implies that

$$|\sum_{j=0}^{3k} \lambda(n+j)\chi_3(n+j)| \geq 2k - 6$$

(say) a.a.s.. But this contradicts Theorem 2.6 if k is large enough. This (finally!) completes the proof of Theorem 2.1 for the sign pattern $\overset{\vee}{+}++$. The case $\overset{\vee}{-}--$ is proven similarly by reversing all the signs in the above argument.

4. The $\overset{\vee}{+}-+$ case

We now prove Theorem 2.1 for the sign pattern $\overset{\vee}{+}-+$; the argument for $\overset{\vee}{-}+-$ is analogous and follows by reversing all the signs below. Our arguments follow that of Hildebrand [11], adapted to the notation of this paper.

For sake of contradiction, we assume that

Hypothesis 4.1. *We have $n \not\mapsto \overset{\vee}{+}-+$ a.a.s..*

This leads to the following implications:

Proposition 4.2. *Asymptotically almost surely, the following two claims hold:*

(a) *If $2n \mapsto \overset{\vee}{+}+$, then $3n \mapsto \overset{\vee}{+}+$.*
(b) *If $2n \mapsto +\overset{\vee}{+}$, then $3n \mapsto +\overset{\vee}{+}$.*

Proof. We just prove (a), as (b) is analogous (reflecting all sign patterns around the positional marker $\vee$). Suppose that $2n \mapsto \overset{\vee}{+}+$, then since $\lambda(3m) = -\lambda(m)$, we have

$$6n \mapsto \overset{\vee}{-}{*}{*}-.$$

Since $\lambda(2m) = -\lambda(m)$, our objective is to show that $\lambda(6n+2) = -1$ a.a.s.. Suppose instead that $\lambda(6n+2) = +1$, thus

$$6n \mapsto \overset{\vee}{-}{*}+-.$$

By Hypothesis 4.1 and Lemma 2.3, we have

$$6n \not\mapsto \overset{\vee}{*}{*}+-+$$

and thus

$$6n \mapsto \overset{\vee}{-}{*}+--$$

a.a.s.. But as $\lambda(2m) = -\lambda(m)$, this implies

$$3n \mapsto \overset{\vee}{+}-+$$

which contradicts Hypothesis 4.1 and Lemma 2.3 a.a.s.. □

Corollary 4.3. *Let a, b be integers with $a < b$. Asymptotically almost surely, if $\lambda(m) = +1$ for all $n + a \leq m \leq n + b$, then $\lambda(m) = +1$ for all $\frac{3}{2}(n + a) \leq m \leq \frac{3}{2}(n + b)$.*

Proof. By the union bound and Lemma 2.3, it suffices to verify this in the case $a = 0, b = 1$. Proposition 4.2(a) then handles the case when n is even, and Proposition 4.2(b) handles the case when n is odd, and the claim then follows from Lemma 2.3. □

Corollary 4.4. *For any natural number k, we have $\lambda(n) = \ldots = \lambda(n + k - 1) = +1$ with positive asymptotic probability.*

Proof. We first establish the case $k = 4$, which of course implies the $k = 1, 2, 3$ cases as well. Suppose for contradiction that the $k = 4$ claim failed, thus

$$n \not\mapsto \overset{\vee}{+}+++$$

a.a.s.. In particular, from Lemma 2.3 one a.a.s. has

$$3n \not\mapsto \overset{\vee}{+}+++.$$

By Proposition 4.2 in the contrapositive we then have

$$2n \not\mapsto \overset{\vee}{+}++$$

a.a.s.; but by Hypothesis 4.1 and Lemma 2.3 we have

$$2n \not\mapsto \overset{\vee}{+}-+$$

and thus (since $\lambda(2m) = -\lambda(m)$)

$$n \not\mapsto \overset{\vee}{-}-$$

a.a.s.. But this contradicts Theorem 2.7.

Now assume inductively that the claim holds for some $k \geq 4$. By partitioning according to the parity of n and using Lemma 2.3, we thus see that with positive asymptotic probability, either

$$\lambda(2n) = \lambda(2n+1) = \cdots = \lambda(2n+k-1) = +1$$

or

$$\lambda(2n-1) = \lambda(2n) = \cdots = \lambda(2n+k-2) = +1.$$

In the former case, we see from Corollary 4.3 (and the hypothesis $k \geq 4$, which implies that $k \leq 3\frac{k-1}{2}$) that

$$\lambda(3n) = \cdots = \lambda(3n+k) = +1,$$

and in the latter case we similarly have (since $k - 1 \leq 3\frac{k-2}{2}$)

$$\lambda(3n-1) = \lambda(3n) = \cdots = \lambda(3n+k-1) = +1.$$

In either case we obtain the $k+1$ case of the claim from Lemma 2.3. □

Next, for any fixed real number $a > 0$, let $A_a \subset (a, +\infty)$ be the set

$$A_a := \{t \in (a, +\infty) : \lambda(n) = +1 \text{ for all } t - a < n < t + a\}$$

and consider the quantity

$$p_a := \operatorname{LIM} \left(\frac{1}{\log i} \int_{x_i/i}^{x_i} 1_{A_a}(t) \frac{dt}{t} \right)_{i=1}^{\infty}.$$

From Corollary 4.4 applied with, say, $k = \lfloor a + 10 \rfloor$, and rounding to the nearest integer, we see that $p_a > 0$ for any fixed $a > 0$. On the other hand, from Theorem 2.5 (and another rounding argument) we see that

$$\lim_{a \to \infty} p_a = 0. \tag{4.1}$$

We now claim that

$$p_{\frac{3}{2}a+10} \geq p_{a+10} \tag{4.2}$$

for any $a \geq 0$; iterating this starting from (say) $a = 1$, we see that $\limsup_{a \to \infty} p_a \geq p_{11} > 0$, contradicting (4.1).

We now show (4.2). Suppose that $t \in A_{a+10}$, thus

$$\lambda(n) = +1 \text{ for all } t - a - 10 < n < t + a + 10.$$

Applying Corollary 4.3 (and rounding to the nearest integer), we then conclude that

$$\lambda(n) = +1 \text{ for all } \frac{3}{2}t - \frac{3}{2}a - 10 < n < \frac{3}{2}t + \frac{3}{2}a + 10$$

for $t \in A_{a+10}$ outside of an exceptional set $E_a \subset (0, +\infty)$ which has zero asymptotic density in the sense that

$$\mathrm{LIM} \left(\frac{1}{\log i} \int_{x_i/i}^{x_i} 1_{E_a}(t) \frac{dt}{t} \right)_{i=1}^\infty = 0.$$

We conclude that

$$\mathrm{LIM} \left(\frac{1}{\log i} \int_{x_i/i}^{x_i} 1_{A_{\frac{3}{2}a+10}} \left(\frac{3}{2} t \right) \frac{dt}{t} \right)_{i=1}^\infty \geq p_a,$$

and hence by the change of variables $t' := \frac{3}{2}t$

$$\mathrm{LIM} \left(\frac{1}{\log i} \int_{3x_i/2i}^{3x_i/2} 1_{A_{\frac{3}{2}a+10}}(t) \frac{dt}{t} \right)_{i=1}^\infty \geq p_a.$$

We may replace the limits of integration from $\int_{3x_i/2i}^{3x_i/2}$ to $\int_{x_i/i}^{x_i}$ incurring an error of $O\left(\frac{1}{\log i}\right)$ which vanishes in the limit, and (4.2) follows.

5. A random graph theory question

We now return to the proof of Theorem 2.2. In this section we will reduce this theorem to the task of establishing that a certain random graph is almost surely connected; this connectedness will be verified in the next section.

Recall that the only remaining tasks are to show that $(\mu(n), \mu(n+1))$ attains each of the four values $(+1, +1), (+1, -1), (-1, +1), (-1, -1)$ with positive asymptotic probability. From the asymptotic probabilities already computed in Section 2, one can check that the four events

$$\begin{gathered} \mu(n) = +1, \mu^2(n+1) = 1 \\ \mu(n) = -1, \mu^2(n+1) = 1 \\ \mu^2(n) = 1, \mu(n+1) = +1 \\ \mu^2(n) = 1, \mu(n+1) = -1 \end{gathered}$$

each occur with asymptotic probability $c/2$, where c was defined in (2.3). In particular we see that $(\mu(n), \mu(n+1))$ takes the values $(+1, +1)$ and $(-1, -1)$ with equal asymptotic probability, and also takes the values $(+1, -1)$ and $(-1, +1)$ with equal asymptotic probability. Thus, we only need to show that the values $(+1, +1)$ and $(+1, -1)$ are attained with positive asymptotic probability.

Suppose for sake of contradiction that there was a sign $\epsilon \in \{-1, +1\}$ such that $(\mu(n), \mu(n+1))$ avoided $(+1, \epsilon)$ a.a.s.. Then it also avoids $(-1, -\epsilon)$ a.a.s., and so we conclude that we have the implication

$$\mu^2(n) = \mu^2(n+1) = 1 \implies \mu(n) = -\epsilon\mu(n+1)$$

a.a.s..

We can eliminate the sign $-\epsilon$ as follows. Define χ to be the completely multiplicative function such that $\chi(p) := +1$ for all $p > 2$, and $\chi(2) := -\epsilon$. Then $\chi(n) = -\epsilon\chi(n+1)$ for any n for which $n, n+1$ are not divisible by 4, and thus we have

$$\mu^2(n) = \mu^2(n+1) = 1 \implies \mu\chi(n) = \mu\chi(n+1)$$

a.a.s.. As $\mu\chi$ is a multiplicative function, we can use Lemma 2.3 and conclude more generally that

$$(\mu^2(n) = \mu^2(n+d) = 1 \wedge d|n) \implies \mu\chi(n) = \mu\chi(n+d) \tag{5.1}$$

a.a.s. for any fixed natural number n.

The strategy is now to use (5.1) create large "chains" $n + a_1, n + a_2, \ldots$ on which $\mu\chi$ is constant, and demonstrate that this is incompatible with Theorem 2.6. It will be convenient to pass from the finitely additive world of asymptotic probability to the countably additive world of genuine probability. Recall that the *profinite integers* $\hat{\mathbb{Z}}$ are defined as the inverse limit of the cyclic groups $\mathbb{Z}/N\mathbb{Z}$; we embed the ordinary integers $\mathbb{Z}$ as a subgroup of $\hat{\mathbb{Z}}$ in the usual fashion. The group $\hat{\mathbb{Z}}$ is compact (in the profinite topology) and thus has a well-defined Haar probability measure, so we can meaningfully talk about a random profinite integer $\mathbf{n} \in \hat{\mathbb{Z}}$, whose reductions $\mathbf{n}\ (N)$ to any cyclic group $\mathbb{Z}/N\mathbb{Z}$ are uniformly distributed in that group. (We will use boldface notation to indicate random variables that are generated from a random profinite integer.) We say that a profinite integer n is divisible by a natural number N if the reduction of n to N vanishes. The Möbius function μ does not obviously extend to the profinite integers, but we can (by abuse of notation) define the quantity $\mu^2(n)$ for profinite n to equal 1 if n is *squarefree* in the sense that it is not divisible by p^2 for any prime p, and equal to 0 otherwise.

We now construct a random graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ as follows. Let $\mathbf{n} \in \hat{\mathbb{Z}}$ be a random profinite integer, and define the random vertex set $\mathbf{V} \subset \mathbb{Z}$ to be the random set of integers defined by

$$\mathbf{V} := \{a \in \mathbb{Z} : \mu^2(\mathbf{n} + a) = 1\}.$$

We then define $\mathbf{E}$ to be the set of pairs $\{a, b\}$ of distinct vertices a, b in $\mathbf{V}$, such that $|b - a|$ is an odd prime number dividing $\mathbf{n} + a$ (or equivalently $\mathbf{n} + b$), and define the random graph $\mathbf{G}$ by setting $\mathbf{V}$ as the set of vertices and $\mathbf{E}$ as the set of edges. (The restriction of $|b - a|$ to be an odd prime will be needed in order to keep the analysis of $\mathbf{G}$ tractable.) Thus, for instance, the integers $2, 5$ will be connected by an edge in $\mathbf{G}$ if $\mu^2(\mathbf{n} + 2) = \mu^2(\mathbf{n} + 5) = 1$ and 3 divides $\mathbf{n} + 2$.

In the rest of the paper we will show

Theorem 5.1. *The random graph* $\mathbf{G}$ *is almost surely connected.*

Let us assume this theorem for now and finish establishing the contradiction required to conclude Theorem 2.2 and hence Theorem 1.9. Let a, b be integers. By Theorem 5.1, we see almost surely that if $\mu^2(\mathbf{n} + a) = \mu^2(\mathbf{n} + b) = 1$, then there exists a finite path $a = c_1, c_2, \ldots, c_k = b$ of distinct integers with $k \geq 1$ with the property that $\mu^2(\mathbf{n} + c_1) = \cdots = \mu^2(\mathbf{n} + c_k) = 1$ and such that $|c_{i+1} - c_i|$ divides $\mathbf{n} + c_i$ for each $i = 1, \ldots, k-1$. (The $|c_{i+1} - c_i|$ are also odd primes, but we will discard this information as it will not be needed here.) There are only countably many choices for $c_1, \ldots, c_k$, so from countable additivity we see that for any $\varepsilon > 0$ and $a, b \in \mathbb{Z}$ one can find a natural number M (depending on ε, a, b) with the property that with probability at least $1 - \varepsilon$, if $\mu^2(\mathbf{n}+a) = \mu^2(\mathbf{n}+b) = 1$, then there exist distinct integers $a = c_1, c_2, \dots, c_k = b$ with $k \leq M$ and $|c_1|, \dots, |c_k| \leq M$, such that $\mu^2(\mathbf{n}+c_1) = \dots = \mu^2(\mathbf{n}+c_k) = 1$ and $|c_{i+1} - c_i|$ divides $\mathbf{n}+c_i$ for each $i = 1, \dots, k-1$.

Note from the Chinese remainder theorem that for any fixed number of congruence conditions $n = a_i \ (q_i)$, the asymptotic probability that n obeys these congruence conditions is equal to the probability that the random profinite integer $\mathbf{n}$ obeys the same congruence conditions. Because of this, we can transfer the previous claim from $\mathbf{n}$ to n. That is to say, for any $\varepsilon > 0$ and $a, b \in \mathbb{Z}$, there exists M with the property that with asymptotic probability at least $1-\varepsilon$, if $\mu^2(n+a) = \mu^2(n+b) = 1$, then there exist distinct integers $a = c_1, c_2, \dots, c_k = b$ with $k \leq M$ and $|c_1|, \dots, |c_k| \leq M$, such that $\mu^2(n+c_1) = \dots = \mu^2(n+c_k) = 1$ and $|c_{i+1} - c_i|$ divides $n+c_i$ for each $i = 1, \dots, k-1$. But from (5.1), this conclusion implies a.a.s. that $\mu\chi(n+c_{i+1}) = \mu\chi(n+c_i)$ for each $i = 1, \dots, k-1$. Chaining these equalities together, we conclude that with asymptotic probability at least $1-\varepsilon$, we have

$$\mu^2(n+a) = \mu^2(n+b) = 1 \implies \mu\chi(n+a) = \mu\chi(n+b).$$

Sending ε to zero, we conclude that this claim holds a.a.s. for any fixed choice of a, b. Applying this for all $a, b \in \{1, \dots, h\}$, we conclude that for any fixed h, we have

$$|\sum_{j=1}^{h} \mu\chi(n+j)| = \sum_{j=1}^{h} \mu^2(n+h)$$

a.a.s.. On the other hand, by using Theorem 2.6 (treating the contribution of odd and even $n+j$ separately) we see that

$$|\sum_{j=1}^{h} \mu\chi(n+j)| \leq \varepsilon h$$

with asymptotic probability at least $1-\varepsilon$, if $\varepsilon > 0$ and h is sufficiently large depending on ε. Thus we see that the quantity $\sum_{j=1}^{h} \mu^2(n+h)$ has an asymptotic expectation of at most $2\varepsilon h$. On the other hand, from linearity of expectation this asymptotic expectation is $\frac{h}{\zeta(2)}$, and one obtains a contradiction if ε is small enough and h is large enough.

It remains to establish Theorem 5.1. In next section we will make several reductions, some of which are easy and some of which are more involved. In the following paragraph we give a sketch of all the reductions, forgetting about some additional technical conditions on the lengths of paths between elements, etc..

First, we will show that instead of showing that almost surely any vertices a and b are connected, it is enough to show that almost surely 0 and X are connected for any large enough odd X. Then, we will show that it is enough to consider the graph with a slight extension of the vertex set, where instead of requiring $\mu^2(\mathbf{n}+a) = 1$, one only requires that $\mathbf{n}+a$ is not divisible by p^2 for any prime $p \leq (\log X)^5$. Next, we will show that instead of connecting 0 to all large enough vertices X, it is enough to connect it to at least $X^{9/10}$ even elements up to $X/10$, which means that it is enough to show that, for all $X' \asymp X$, 0 is almost surely connected to at least one element in $[X' - X^{1/20}, X']$. Then, we will show that this follows if we can show that 0 is connected to at least $\log^{10} X$ odd elements in $[0, X^{1/100}]$. Thus we will dramatically reduce the number of distinct vertices we need to connect 0 to, and this last claim is shown through a very careful study of paths of length somewhat shorter than $\log\log x$ in Section 7.

6. Initial reductions

For the remainder of this paper, we use the usual asymptotic notation of writing $X \ll Y$, $Y \gg X$, or $X = O(Y)$ to denote the estimate $|X| \leq CY$ for some absolute constant C, and write $X \asymp Y$ for $X \ll Y \ll X$.

We now turn to the formal proof of Theorem 5.1. We begin by deducing this theorem from the following claim.

Proposition 6.1. *Let X be a sufficiently large odd natural number, which we view as an asymptotic parameter going to infinity. Then, with probability $1 - o(1)$, if 0 and X both lie in $\mathbf{V}$, then there is a path in $\mathbf{G}$ from 0 to X.*

Let us see how Proposition 6.1 implies Theorem 5.1. Observe that the random graph $\mathbf{G}$ is stationary, in the sense that any translate of $\mathbf{G}$ has the same distribution as $\mathbf{G}$. Thus it suffices to show that for any natural number h, one almost surely has 0 and h connected in $\mathbf{G}$ whenever $0, h$ both lie in $\mathbf{G}$. Stationarity also shows that Proposition 6.1 automatically extends to even X as well as odd X, since one can write a large even number as the sum of two large odd ones.

Let w be a natural number (larger than h), and let $W := \prod_{p \leq w} p$. Note that if 0 and h both lie in $\mathbf{V}$, then $\mathbf{n}$ and $\mathbf{n}+h$ are not divisible by p^2 for any $p \leq w$. Conditioning on the event that 0 and h both lie in $\mathbf{V}$, we see that $\mathbf{n} + W^2$ is not divisible by p^2 for any $p \leq w$, and an application of the union bound shows that $\mathbf{n} + W^2$ is square-free with probability $1 - O(1/w)$; that is to say $W^2 \in \mathbf{V}$ with probability $1 - O(1/w)$. Applying Proposition 6.1 (and stationarity) we conclude that conditionally on 0 and h both lying in $\mathbf{V}$, we have a path in $\mathbf{G}$ from 0 to h through W^2 with probability $1 - O(1/w) - o(1)$ as $W^2 \to \infty$. Sending $w \to \infty$, we obtain Theorem 5.1.

It remains to prove Proposition 6.1. Let X be a sufficiently large odd natural number, viewed as an asymptotic parameter going to infinity. It is convenient to work with a slight enlargement $\mathbf{G}_X$ of $\mathbf{G}$ with slightly more vertices. Set

$$w := \log^5 X$$

and define $\mu_X^2(n)$ for a profinite integer n to equal one when n is not divisible by p^2 for any $p \leq w$, and zero otherwise, and set $\mathbf{V}_X := \{a \in \mathbb{Z} : \mu_X^2(\mathbf{n} + a) = 1\}$; then $\mathbf{V}_X$ contains $\mathbf{V}$, and a vertex a of $\mathbf{V}_X$ lies in $\mathbf{V}$ unless $p^2 | \mathbf{n} + a$ for some $p > w$.

Define $\mathbf{G}_X$ to be the random graph with vertex set $\mathbf{V}_X$, and two distinct elements a, b of $\mathbf{V}_X$ connected by an edge if $|a - b|$ is an odd prime dividing $\mathbf{n} + a$. Thus $\mathbf{G}$ is the restriction of $\mathbf{G}_X$ to $\mathbf{V}$. We will first show that it suffices to prove

Proposition 6.2. *Let X be a sufficiently large odd natural number, which we view as an asymptotic parameter going to infinity. Then, with probability $1 - o(1)$, if 0 and X both lie in $\mathbf{V}_X$, then there is a path in $\mathbf{G}_X$ from 0 to X of length at most $10 \log^2 X$ and contained in $[-10X, 10X]$.*

Let us see how Proposition 6.2 implies Proposition 6.1. The key point is that the restriction of $\mathbf{G}_X$ to $[-10X, 10X]$ does not require knowledge of the entirety of the profinite random integer $\mathbf{n}$; only knowledge of the reductions $\mathbf{n} \bmod p$ for $p \leq 20X$ and $\mathbf{n} \bmod p^2$ for $p \leq w$ is required. The remaining components of $\mathbf{n}$ can then be used to restrict $\mathbf{G}_X$ to $\mathbf{G}$.

Let us first show that, almost surely every number in $[\mathbf{n} - 10X, \mathbf{n} + 10X]$ has at most $\log^2 X$ distinct prime factors in the interval $[w, 20X]$. Notice that this property depends only on reductions modulo primes $p \leq 20X$. Write $W = \prod_{p \leq 20X} p = e^{20X(1+o(1))}$, and then it is enough to show that the number of integers in $[W, 2W)$ that have more than $\log^2 X$ distinct prime factors in $[w, 20X]$ is $o(W/X)$. Such numbers have $20X$-smooth part at least $w^{\log^2 X}$, and we get that the number of them is at most

$$\sum_{\substack{W \leq mn \leq 2W \\ p|m \implies p > 20X \\ p|n \implies p \leq 20X \\ n \geq w^{\log^2 X}}} 1 \leq \sum_{\substack{m \leq 2W \\ p|m \implies p > 20X}} \sum_{\substack{w^{\log^2 X} \leq n \leq 2W/m \\ p|n \implies p \leq 20X}} 1 \ll \sum_{m \leq 2W} \frac{2W}{m} (\log X)^{-c \log X \log\log X} \ll \frac{W}{X^{100}}$$

for an absolute constant $c > 0$, by the standard estimate (see e.g. [2]) that the number of y-smooth numbers up to x is $u^{-(1+o(1))u}$ where $u = \log x / \log y \leq y^{1-\varepsilon}$ (in our case, to get an upper bound, we can take, for every m, $u = \log(w^{\log^2 X})/\log(20X) \asymp \log X \log\log X$).

Now suppose we condition the reductions $(\mathbf{n} \bmod p)_{p \leq 20X}$ and $(\mathbf{n} \bmod p^2)_{p \leq w}$ to be a value for which there is a path in $\mathbf{G}_X$ from 0 to X of length at most $10 \log^2 X$ contained in $[-10X, 10X]$, and for which every number in $[\mathbf{n} - 10X, \mathbf{n} + 10X]$ has at most $\log^2 X$ prime factors in $[w, 20X]$. After this conditioning, the residue classes $\mathbf{n} \bmod p^2$ for $w < p \leq 20X$ are restricted to a single coset of $\mathbb{Z}/p\mathbb{Z}$ in $\mathbb{Z}/p^2\mathbb{Z}$, but are uniformly distributed in that coset, whereas the residue classes $\mathbf{n} \bmod p^2$ for $p > 20X$ are uniformly distributed on all of $\mathbb{Z}/p^2\mathbb{Z}$. Also, the $\mathbf{n} \bmod p^2$ are independent in p across all primes p. Let $0 = a_1, a_2, \ldots, a_k = X$ be a path in $\mathbf{G}_X$ from 0 to X of length $k \leq 10 \log^2 X$ contained in $[-10X, 10X]$. Then, for each $j = 1, \ldots, k$, the number of primes $p \in [w, 20X]$ such that p divides $\mathbf{n} + a_k$ is at most $\log^2 X$. Hence the probability that all of the $\mathbf{n} + a_i$ are not divisible by p^2 for any $p > w$ (which implies that this path lies in $\mathbf{G}$ and not just in $\mathbf{G}_X$) is at least

$$\left(1 - \frac{k}{w}\right)^{\log^2 X} \times \prod_{p > 20X} \left(1 - \frac{k}{p^2}\right).$$

From the bounds on w, k we see that this expression is $1 - o(1)$. This then gives Proposition 6.1 from Proposition 6.2.

We will want to substantially reduce the number of vertices to which we need to connect 0. To do this, we will use the following lemma several times.

Lemma 6.3. *Let X be large and fix the residue classes $\mathbf{n}$ mod p for $p \leq X$ and $\mathbf{n}$ mod p^2 for $p \leq w$. Let A and B be respectively subsets of odd and even integers in $[0, X] \cap \mathbf{V}_X$ such that $|A||B| \gg X \log^{10} X$. Then, with conditional probability $1 - O(\log^{-2} X)$, there is a path of length 3 in $\mathbf{V}_X$ contained in $[-8X, 8X]$ connecting an element $a \in A$ with an element $b \in B$.*

Proof. Let S denote the set of quadruples (a, b, p_1, p_2, p_3) obeying the following constraints:

- We have $a \in A, b \in B$, and p_1, p_2, p_3 are primes in $(X, 3X]$, $(5X, 7X]$, $(3X, 5X]$ respectively obeying the equation
$$b = a - p_1 + p_2 - p_3.$$
- $a - p_1$ and $a - p_1 + p_2$ lie in $\mathbf{V}_X$, that is to say $\mathbf{n} + a - p_1, \mathbf{n} + a - p_1 + p_2$ are not divisible by p^2 for any $p \leq w$.

Note that the set S is deterministic due to our conditioning. A routine application of the circle method (see Proposition 8.1 below) shows that each pair $(a,b) \in A \times B$ contributes $\asymp \frac{X^2}{\log^3 X}$ quintuples to S, and so

$$|S| \asymp |A||B|X^2/\log^3 X. \tag{6.1}$$

Each quintuple (a,b,p_1,p_2,p_3) will yield a path $a, a-p_1, a-p_1+p_2, a-p_1+p_2-p_3 = b$ of the desired form provided that one has the divisibility conditions

$$p_1|\mathbf{n}+a; \quad p_2|\mathbf{n}+a-p_1; \quad p_3|\mathbf{n}+a-p_1+p_2.$$

Let $E_{(a,b,p_1,p_2,p_3)}$ be the event that these three divisibility conditions occur; thus it suffices to show that

$$\mathbb{P}\left(\bigvee_{s\in S} E_s\right) = 1 - O((\log X)^{-2}).$$

We use the second moment method. By the Cauchy-Schwarz inequality,

$$\mathbb{P}(\bigvee_{s\in S} E_s) \geq \frac{(\mathbb{E}\sum_{s\in S} 1_{E_s})^2}{\mathbb{E}(\sum_{s\in S} 1_{E_s})^2} = \frac{\sum_{s,s'\in S}\mathbb{P}(E_s)\mathbb{P}(E_{s'})}{\sum_{s,s'\in S}\mathbb{P}(E_s\cap E_{s'})}.$$

Hence it suffices to show that

$$\sum_{s,s'\in S:\mathbb{P}(E_s\cap E_{s'})>\mathbb{P}(E_s)\mathbb{P}(E_{s'})} \mathbb{P}(E_s\cap E_{s'}) \ll \frac{1}{\log^2 X}\sum_{s,s'\in S}\mathbb{P}(E_s)\mathbb{P}(E_{s'}). \tag{6.2}$$

On the one hand, from the Chinese remainder theorem we have

$$\mathbb{P}(E_s) = \frac{1}{p_1p_2p_3} \asymp \frac{1}{X^3}$$

and so by (6.1)

$$\sum_{s,s'\in S}\mathbb{P}(E_s)\mathbb{P}(E_{s'}) \asymp \frac{|A|^2|B|^2}{X^2\log^6 X}. \tag{6.3}$$

On the other hand, if $s = (a,b,p_1,p_2,p_3)$ and $s' = (a',b',p_1',p_2',p_3')$ lie in S, then $\mathbb{P}(E_s\cap E_{s'})$ is equal to either $\mathbb{P}(E_s)\mathbb{P}(E_{s'})$ or 0 unless at least one of the following three situations occur:

(i) $p_1 = p_1'$ and $a = a'$.
(ii) $p_2 = p_2'$ and $a - p_1 = a' - p_1'$.
(iii) $p_3 = p_3'$ and $b = b'$.

(Here we are using the fact that p_1, p_2, p_3 lie in disjoint intervals, and p_1 is larger than the diameter of the interval in which a ranges, p_2 is larger than the diameter of the interval in which $a - p_1$ ranges and p_3 is larger than the diameter of the interval in which b ranges.) Furthermore, in these exceptional cases, $\mathbb{P}(E_s\cap E_{s'})$ may be bounded by $O(\frac{1}{X^{6-j}})$, where $j = 1,2,3$ is the number of situations (i), (ii), (iii) that are occurring simultaneously.

Meanwhile, when j of the situations occur simultaneously, a simple degree of freedom counting gives $O(|A||B|(X/\log X)^2)$ choices for s and then $O((X/\log X)^{3-j})$ choices for s' (one can choose the primes p_j' not fixed by the first conditions in (i)–(iii) freely, and then everything else is fixed by the conditions and definition of S). Thus the left-hand side of (6.2) is bounded by

$$O\left(\sum_{j=1}^{3}|A||B|\left(\frac{X}{\log X}\right)^{5-j}\cdot\frac{1}{X^{6-j}}\right)=O\left(\frac{|A||B|}{X(\log X)^2}\right)$$

and (6.2) follows from (6.3) since $|A||B|\geq X(\log X)^{10}$. □

Now we use the previous lemma to make a reduction, in which we content ourselves with connecting 0 to many vertices in an interval $[0,X/10]$, rather than trying to reach a specific vertex X. Namely, we reduce Proposition 6.2 to

Proposition 6.4. *Let X be a sufficiently large odd natural number, which we view as an asymptotic parameter going to infinity. Then, with probability $1-o(1)$, if 0 lies in $\mathbf{V}_X$, then there are at least $X^{9/10}$ even elements of $[0,X/10]$ which are connected to 0 in $\mathbf{G}_X$ by a path of length at most $2\log^2 X$ contained in $[0,X/10]$.*

We now explain why Proposition 6.4 implies Proposition 6.2. For any integer a, let E_a be the event that a lies in $\mathbf{V}_X$, and there are at least $X^{9/10}$ elements of $[a,a+X/10]$ of the same parity as a which are connected to a in $\mathbf{G}_X$ by a path of length at most $2\log^2 X$ contained in $[a,a+X/10]$. By stationarity and Proposition 6.4, we see that for all a, one has with probability $1-o(1)$ that E_a holds whenever $a\in\mathbf{V}_X$. In particular, with probability $1-o(1)$, one has $E_0\cap E_X$ holding whenever $0,X\in\mathbf{V}_X$.

Now suppose that $E_0\cap E_X$ holds. This event only depends on the residue classes $\mathbf{n}$ mod p for $p\leq X/10$ and $\mathbf{n}$ mod p^2 for $p\leq w$, so we now condition these residue classes to be deterministic. We now have deterministic subsets A,B of $[0,X/10]$ and $[X,11X/10]$ respectively of cardinality at least $X^{9/10}$ each, such that any element of A (resp. B) is connected to 0 (resp. X) by a path in $\mathbf{G}_X$ of length at most $2\log^2 X$ contained in $[-10X,10X]$. Furthermore, A consists entirely of even numbers, and B consists entirely of odd numbers. The claim now follows from Lemma 6.3.

Next, we reduce Proposition 6.4 to the following variant, in which we connect 0 to one element in a short interval, as opposed to many elements in a long interval:

Proposition 6.5. *Let X be a sufficiently large odd natural number, which we view as an asymptotic parameter going to infinity. Let X' be an element of the interval $[X/40,X/20]$, chosen uniformly at random. Then, with probability $1-o(1)$, if 0 lies in $\mathbf{V}_X$, then there is an element of $[X'-X^{1/20},X']$ that is connected to 0 in $\mathbf{G}_X$ by a path of length at most $2\log^2 X$ contained in $[0,X/10]$.*

Indeed, assuming Proposition 6.5, observe that if 0 lies in $\mathbf{V}_X$ and X' is chosen uniformly at random from $[X/40,X/20]$, then the expected number of elements in $[X'-X^{1/20},X']$ that are connected to 0 in the indicated fashion is at least $1-o(1)$. From linearity of expectation, this implies that the number of elements of $[X/40-X^{1/20},X/20]$ that are connected to 0 in the indicated fashion is $\gg X^{1-1/20}$, and Proposition 6.4 follows.

Finally, we reduce to a weaker version of Proposition 6.4, in which 0 connects to far fewer elements in (a somewhat narrower) interval:

Proposition 6.6. *Let X be a sufficiently large odd natural number, which we view as an asymptotic parameter going to infinity. Then, with probability $1-o(1)$, if 0 lies in*

$\mathbf{V}_X$, then there are at least $\log^{10} X$ odd elements of $[0, X^{1/100}]$ which are connected to 0 in $\mathbf{G}_X$ by a path of length at most $\log X$ contained in $[0, X^{1/100}]$.

Let us now see how Proposition 6.6 implies Proposition 6.5. Call an integer a *good* if $a \in \mathbf{V}_X$ is even, and a is connected to at least $\log^{10} X$ odd elements of $[a, a + X^{1/100}]$ in $\mathbf{G}_X$ by a path of length at most $\log X$ contained in $[a, a + X^{1/100}]$. From Proposition 6.6 and stationarity, we see that each even element a of $\mathbf{V}_X$ is good with probability $1 - o(1)$. From linearity of expectation, we thus see with probability $1 - o(1)$ that all but $o(X)$ of the even elements of $\mathbf{V}_X \cap [-X, X]$ are good.

Note that the property of an integer being good is only dependent on the values of $\mathbf{n} \bmod p^2$ for $p \leq w$ and $\mathbf{n} \bmod p$ for $w < p \leq X^{1/100}$. We now condition on these values to be fixed, in such a fashion that all but $o(X)$ of the even elements of $\mathbf{V}_X \cap [-X, X]$ are good; now the property of being good is deterministic, as is the vertex set $\mathbf{V}_X$. We now use the following weak version of the Hardy-Littlewood inequality.

Proposition 6.7. *Let a_n be a sequence of non-negative real numbers supported in $[-X, X]$. Then,*

$$\frac{1}{X} \sum_{|n| \leq X} \Big(\sup_{r \geq 1} \frac{1}{r} \sum_{|n-m| \leq r} a_m \Big) \leq C \Big(\frac{1}{X} \sum_{|n| \leq X} a_n^2 \Big)^{1/2} \tag{6.4}$$

for some absolute constant $C > 0$.

Proof. The sequence $\Big(\sup_{r \geq 1} \frac{1}{r} \sum_{|n-m| \leq r} a_m \Big)_{n \in \mathbb{Z}}$ is the (discrete) Hardy-Littlewood maximal operator applied to this sequence $(a_n)_{n \in \mathbb{Z}}$. As the Hardy-Littlewood maximal operator is bounded in $\ell^2(\mathbb{Z})$ (see e.g. [14]), we have

$$\left(\sum_n \Big(\sup_{r \geq 1} \frac{1}{r} \sum_{|n-m| \leq r} a_m \Big)^2 \right)^{1/2} \leq C \Big(\sum_{|n| \leq X} a_n^2 \Big)^{1/2}$$

for some $C > 0$, and the claim then follows from the Cauchy-Schwarz inequality. □

Choosing a_m so that $a_m = 1$ if $m \in V_{\mathbf{X}} \cap [-X, X]$ is not good, and $a_m = 0$ otherwise, the proposition implies that if X' is chosen uniformly at random from $[X/40, X/20]$, then with probability $1 - o(1)$, the quantity X' is "excellent" in the sense that for any $1 \leq r \leq X$, all but at most $o(r)$ of the even elements of $\mathbf{V}_X \cap [X' - r, X' + r]$ are good. Note that because of our conditioning, the property of being excellent is deterministic.

Now let $X' \in [X/40, X/20]$ be an excellent number. We will show that with probability $1 - o(1)$, if $0 \in \mathbf{G}_X$, then there is an element of $[X' - X^{1/20}, X']$ that is connected to 0 in $\mathbf{G}_X$ by a path of length at most $2 \log^2 X$ contained in $[0, X/10]$; this clearly suffices to give Proposition 6.5.

We introduce the scales $R_j := 100^{1-j} X'$ for $j = 1, \ldots, J$, where J is the first number for which $R_J \leq X^{1/40}$, thus $J \leq \log X$. Introduce the intervals $I_j := [X' - R_j, X' - 0.99 R_j]$, thus the I_j are disjoint, with I_1 containing 0 and I_J contained in $[X' - X^{1/20}, X']$. We will "hop" from 0 to I_J by a path passing through each of the I_j in turn. More precisely, let A_j denote the even elements of $\mathbf{V}_X \cap I_j$ that are good; with all of our conditioning, this is a deterministic set. A routine sieve shows that there are $\gg R_j$ even elements of $\mathbf{V}_X \cap I_j$, and as X' is excellent, all but at most $o(R_j)$ of these elements are good. We conclude that

$$|A_j| \gg R_j.$$

We shall shortly establish the following lemma allowing one to hop from A_j to A_{j+1}:

Lemma 6.8. *Let $1 \leq j < J$, and let $a_j \in A_j$, which is allowed to be a random variable depending on the values of $\mathbf{n}$ mod p for $p \leq R_j/10$ but not on the reductions for higher p. Then, with probability $1 - O(\frac{1}{\log^2 X})$, there is a path of length at most $\log X + 3$ in $\mathbf{G}_X$ in $[0, X/10]$ connecting a_j to an element a_{j+1} of A_{j+1}. Furthermore, a_{j+1} depends only on the values of $\mathbf{n}$ mod p for $p \leq R_{j+1}/10$.*

Iterating this lemma with the union bound, starting from $a_1 = 0$, we conclude with probability $1 - O(\frac{J}{\log^2 X}) = 1 - o(1)$ that there is a path of length at most $J(\log X + 3) \leq 2\log^2 X$ in $\mathbf{G}_X$ in $[0, X/10]$ connecting 0 to an element of $A_J \subset [X' - X^{1/20}, X']$, giving Proposition 6.5. Thus it suffices to prove Lemma 6.8.

We do this by an argument similar to that used to obtain Proposition 6.4 from Proposition 6.2. Condition on the values of $\mathbf{n}$ mod p for $p \leq R_j/10$, so that a_j is now deterministic. By definition of A_j, a_j is connected to a (deterministic) set A by paths of length at most $\log X$ in $\mathbf{G}_X$ in $[0, X/10]$, where $A \subset [a_j, a_j + X^{1/100}]$ is a collection of odd numbers of cardinality

$$|A| \geq \log^{10} X.$$

Since $|A||A_{j+1}| \gg R_j \log^{10} X$, Lemma 6.3 implies that there is a path of length 3 connecting some $a \in A$ to some $b \in A_{j+1}$, and the claim follows.

It remains to prove Proposition 6.6. This will be done in the next section.

7. Conclusion of the argument

We now prove Proposition 6.6. Condition the residue classes $\mathbf{n}$ mod p^2 for $p \leq w$ to be fixed, which makes the vertex set $\mathbf{V}_X$ deterministic, while keeping the $\mathbf{n}$ mod p for $p > w$ uniformly and independently distributed on $\mathbb{Z}/p\mathbb{Z}$.

Set k to be the odd number

$$k := 100 \left\lfloor \frac{\log\log X}{\sqrt{\log\log\log X}} \right\rfloor + 1; \tag{7.1}$$

the reason for this somewhat strange choice is that $(\log\log X)^k$ will be significantly larger than $\log^{10} X$, while k remains significantly smaller than $\log\log X$. Let Γ be the set of paths γ in $\mathbf{V}_X$ of the form

$$0, p_1, p_1 + p_2, \dots, p_1 + \dots + p_k \tag{7.2}$$

where $p_1, \dots, p_k$ are distinct primes in the interval $I := [\exp(\sqrt{\log X}), X^{1/200}]$. We write $\gamma(k) := p_1 + \dots + p_k$ for the endpoint of such a path, which is automatically odd since k and the $p_1, \dots, p_k$ are odd, and by abuse of notation write $\gamma \subset \mathbf{G}_X$ if the path γ lies in $\mathbf{G}_X$, or equivalently that

$$p_i | \mathbf{n} + p_1 + \dots + p_{i-1}$$

for all $i = 1, \dots, k$.

It will be technically convenient to weight the paths γ in Γ. For each path γ of the form (7.2), define the weight $w_\gamma > 0$ by the formula

$$w_\gamma := \prod_{i=1}^k \frac{1}{\sum_{p \in I : p_1 + \dots + p_{i-1} + p \in \mathbf{V}_X} \frac{1}{p}}.$$

Lemma 7.1. *For every $\gamma \in \Gamma$, one has*

$$\sum_{p \in I : p_1 + \dots + p_{i-1} + p \in \mathbf{V}_X} \frac{1}{p} \asymp \log \log X$$

for all $i = 1, \dots, k$, so that

$$w_\gamma = \exp(O(k))/(\log \log X)^k \tag{7.3}$$

Proof. Since $p \leq X^{1/200}$ the upper bound

$$\sum_{p \in I : p_1 + p_2 + \ldots + p_{i-1} + p \in \mathbf{V}_X} \frac{1}{p} \leq \sum_{p \in I} \frac{1}{p} = \frac{1}{2} \log \log X + O(1)$$

is clear. For the lower bound note that since the reduction $\mathbf{n} \pmod{p^2}$ is fixed for $p \leq w$, there is a positive integer $A \leq X + \prod_{p \leq w} p^2$ such that the condition $p_1 + \ldots + p_{i-1} + p \in \mathbf{V}_X$ is equivalent to the condition $q^2 \nmid p + A$ for all $q \leq w$. Pick a large constant C, let $\mathcal{P} = \prod_{p \leq C} p^2$. Note that by the Chinese Remainder Theorem there are $\ell = \prod_{p \leq C} (\varphi(p^2) - 1)$ residues classes $a_1, \ldots, a_\ell \pmod{\mathcal{P}}$ such that if p belongs to one of them then $q^2 \nmid p + A$ for all $q \leq C$. Therefore,

$$\mathbf{1}_{p_1 + p_2 + \ldots + p_{i-1} + p \in \mathbf{V}_X} \geq \sum_{i=1}^{\ell} \mathbf{1}_{p \equiv a_i \pmod{\mathcal{P}}} - \sum_{C < q \leq w} \mathbf{1}_{p \equiv -A \pmod{q^2}}$$

Summing this with a weight of $1/p$ and using the prime number theorem in arithmetic progressions we conclude that

$$\sum_{p \in I : p_1 + \ldots + p_{i-1} + p \in \mathbf{V}_X} \frac{1}{p} \geq \Big(\prod_{p \leq C} \frac{\varphi(p^2) - 1}{\varphi(p^2)} - \frac{B}{C} \Big) \sum_{p \in I} \frac{1}{p}$$

for some absolute constant $B > 0$. Therefore if C is choosen large enough then we obtain the desired lower bound. □

Let us first show that Proposition 6.6 follows once we have shown the three estimates

$$\sum_{\gamma \in \Gamma} w_\gamma \mathbb{P}(\gamma \subset \mathbf{G}_X) = 1 + o(1), \tag{7.4}$$

$$\sum_{\gamma, \gamma' \in \Gamma} w_\gamma w_{\gamma'} \mathbb{P}(\gamma, \gamma' \subset \mathbf{G}_X) \leq 1 + o(1), \tag{7.5}$$

and

$$\sum_{\gamma, \gamma' \in \Gamma : \gamma(k) = \gamma'(k)} w_\gamma w_{\gamma'} \mathbb{P}(\gamma, \gamma' \subset \mathbf{G}_X) \ll \log^{-100} X. \tag{7.6}$$

Indeed, from (7.4), (7.5), and Chebyshev's inequality we have

$$\sum_{\gamma \in \Gamma} w_\gamma 1_{\gamma \subset \mathbf{G}_X} = 1 + o(1)$$

with probability $1 - o(1)$, while (7.6) and Markov's inequality gives

$$\sum_{\gamma, \gamma' \in \Gamma : \gamma(k) = \gamma'(k)} w_\gamma w_{\gamma'} 1_{\gamma, \gamma' \subset \mathbf{G}_X} \ll \log^{-99} X$$

with probability $1-o(1)$. From Cauchy-Schwarz we have

$$\begin{aligned}\left(\sum_{\gamma\in\Gamma} w_\gamma 1_{\gamma\subset\mathbf{G}_X}\right)^2 &= \left(\sum_m \sum_{\gamma\in\Gamma\colon \gamma(k)=m} w_\gamma 1_{\gamma\subset\mathbf{G}_X}\right)^2\\ &\le |\{\gamma(k)\colon \gamma\in\Gamma, \gamma\subset\mathbf{G}_X\}|\cdot \sum_{\substack{\gamma,\gamma'\in\Gamma\\ \gamma(k)=\gamma'(k)}} w_\gamma w_{\gamma'} 1_{\gamma,\gamma'\subset\mathbf{G}_X}.\end{aligned}$$

Hence once we have shown the bounds (7.4), (7.5), (7.6), we get

$$|\{\gamma(k)\colon \gamma\in\Gamma, \gamma\subset\mathbf{G}_X\}| \gg (\log X)^{99}$$

with probability $1-o(1)$ and Proposition 6.6 follows.

We begin with (7.4). From the Chinese remainder theorem we have

$$\mathbb{P}(\gamma\subset\mathbf{G}_X) = \frac{1}{p_1\dots p_k}$$

for a path γ of the form (7.2), since the $p_1,\dots,p_k$ were assumed to be distinct; thus the left-hand side of (7.4) becomes

$$\sum_{\gamma\in\Gamma}\frac{w_\gamma}{p_1\dots p_k}. \tag{7.7}$$

We can interpret this expression probabilistically as follows. Consider a random path $0, p_1, p_1+p_2, \dots, p_1+\dots+p_k$, constructed iteratively by requiring that whenever $1\le i\le k$ and $p_1,\dots,p_{i-1}$ have already been chosen, then $p_i\in I$ is chosen with probability

$$\frac{1/p_i}{\sum_{p\in I: p_1+\dots+p_{i-1}+p\in\mathbf{V}_X} 1/p}$$

if $p_1+\dots+p_i\in\mathbf{V}_X$, and chosen with probability zero otherwise. Then the quantity (7.7) is nothing more than the probability that this random path actually lies in Γ. This gives the upper bound for (7.4) automatically. For the lower bound, observe that the only way the path $0, p_1, \dots, p_1+\dots+p_k$ could fail to lie in Γ is if there is a collision $p_i=p_j$ for some $1\le i<j\le k$. But if $1\le i<j$, then after fixing i,j and $p_1,\dots,p_{j-1}$ we see from Lemma 7.1 that the probability of the event $p_i=p_j$ occurring is $\ll \frac{1}{p_i\log\log X} \ll \exp(-\sqrt{\log X})$, for a total failure probability of $\ll k^2\exp(-\sqrt{\log X}) = o(1)$. This proves (7.4).

Now we prove (7.5). By (7.4), it suffices to show that

$$\sum_{\gamma,\gamma'\in\Gamma:\mathbb{P}(\gamma,\gamma'\subset\mathbf{G}_X)>\mathbb{P}(\gamma\subset\mathbf{G}_X)\mathbb{P}(\gamma'\subset\mathbf{G}_X)} w_\gamma w_{\gamma'}\mathbb{P}(\gamma,\gamma'\subset\mathbf{G}_X)\le o(1).$$

Let $\gamma = 0, p_1, \dots, p_1+\dots+p_k$ and $\gamma' = 0, p_1', \dots, p_1'+\dots+p_k'$ be two paths in Γ. The quantity $\mathbb{P}(\gamma,\gamma'\subset\mathbf{G}_X)$ is usually equal to either $\mathbb{P}(\gamma\subset\mathbf{G}_X)\mathbb{P}(\gamma'\subset\mathbf{G}_X)$ or zero. The only exceptions occur if we have at least one collision of the form $p_i = p_j'$ for some $1\le i,j\le k$. Furthermore, if such a collision occurs, the quantity $p_1+\dots+p_{i-1}-p_1'-\dots-p_{j-1}'$ must be divisible by p_i. If there are exactly r collisions $p_{i_l}=p_{j_l}'$ for some pairs $(i_1,j_1),\dots,(i_r,j_r)\in\{1,\dots,k\}^2$, then we have

$$\mathbb{P}(\gamma,\gamma'\subset\mathbf{G}_X)\le\frac{p_{i_1}\dots p_{i_r}}{p_1\dots p_k p_1'\dots p_k'}.$$

It thus suffices to show that

$$\sum_{r=1}^{k}\sum_{(i_1,j_1),\dots,(i_r,j_r)}\sum_{\gamma,\gamma'\in\Gamma} w_\gamma w_{\gamma'}\prod_{l=1}^{r} 1_{p_{i_l}=p'_{j_l}} 1_{p_{i_l}|p_1+\dots+p_{i_l-1}-p'_1-\dots-p'_{j_l-1}} \\ \frac{p_{i_1}\dots p_{i_r}}{p_1\dots p_k p'_1\dots p'_k} = o(1), \tag{7.8}$$

where the second sum is over distinct pairs $(i_1,j_1),\dots,(i_r,j_r)$ in $\{1,\dots,k\}^2$, with the $i_1,\dots,i_r$ and the $j_1,\dots,j_r$ distinct, with the ordering $i_1<\dots<i_r$ (to avoid duplicates).

Consider first the contribution of the case where $i_l=l$ for $l=1,\dots,r$. In this case we will omit the condition $1_{p_{i_l}|p_1+\dots+p_{i_l-1}-p'_1-\dots-p'_{j_l-1}}$, which in principle gives a large reduction to the size of the expression in (7.8), but is difficult to analyse. We also drop the condition that $\gamma,\gamma'\in\Gamma$ (thus allowing duplicates among p_i and among p'_i). We can thus bound the left-hand side of this contribution to (7.8) by

$$\sum_{r=1}^{k}\sum_{(1,j_1),\dots,(r,j_r)}\mathbb{E}\prod_{l=1}^{r}(p_l 1_{p_l=p'_{j_l}}) \tag{7.9}$$

the paths $\gamma=(0,p_1,\dots,p_1+\dots+p_k)$ and $\gamma'=(0,p'_1,\dots,p'_1+\dots+p'_k)$ are selected randomly as in the proof of (7.4). Observe that if one conditions $p'_1,\dots,p'_k$ and $p_1,\dots,p_{l-1}$ to be fixed, then $p_l 1_{p_l=p'_{j_l}}$ has conditional expectation

$$\frac{1}{\sum_{p\in I:p_1+\dots+p_{i-1}+p\in\mathbf{V}_X}\frac{1}{p}}\ll\frac{1}{\log\log X}$$

by Lemma 7.1. Thus by multiplying together the conditional expectations, we can bound (7.9) by

$$\sum_{r=1}^{k}\sum_{(1,j_1),\dots,(r,j_r)}O\left(\frac{1}{\log\log X}\right)^r.$$

Because we have constrained $i_l=l$ for $l=1,\dots,r$, there are only k^r choices for the $(i_1,j_1),\dots,(i_r,j_r)$, so we can bound the total contribution to (7.9) or (7.8) by

$$\sum_{r=1}^{k}O\left(\frac{k}{\log\log X}\right)^r$$

which is (barely) of the form $o(1)$ thanks to (7.1).

Now we consider the contribution of the case where $i_l\neq l$ for at least one $1\le l\le r$; in particular, there exists $1\le l_0\le r$ such that $i_{l_0}>i_{l_0-1}+1$ (with the convention that $i_0=0$). We temporarily fix r, $(i_1,j_1),\dots,(i_r,j_r)$ and l_0 and consider the corresponding component of

$$\sum_{\gamma,\gamma'\in\Gamma}w_\gamma w_{\gamma'}\prod_{l=1}^{r}1_{p_{i_l}=p'_{j_l}}1_{p_{i_l}|p_1+\dots+p_{i_l-1}-p'_1-\dots-p'_{j_l-1}}\frac{p_{i_1}\dots p_{i_r}}{p_1\dots p_k p'_1\dots p'_k} \tag{7.10}$$

to (7.8). Here we will keep only one of the conditions $1_{p_{i_l}|p_1+\dots+p_{i_l-1}-p'_1-\dots-p'_{j_l-1}}$, and also estimate the w_γ by (7.3), arriving at an upper bound of

$$\ll\frac{\exp(O(k))}{(\log\log X)^{2k}}\sum_{p_1,\dots,p_k,p'_1,\dots,p'_k\in I}\frac{\prod_{l=1}^{r}(p_{i_l}1_{p_{i_l}=p'_{j_l}})}{p_1\dots p_k p'_1\dots p'_k}1_{p_{i_{l_0}}|p_1+\dots+p_{i_{l_0}-1}-p'_1-\dots-p'_{j_{l_0}-1}}. \tag{7.11}$$

We sum first over $p_{i_{l_0}-1}$, keeping all the other variables in $p_1, \dots, p_k, p'_1, \dots, p'_k$ fixed. Then $p_{i_{l_0}-1}$ is constrained to a single residue class a mod $p_{i_{l_0}}$. We can crudely bound

$$\sum_{\substack{p_{i_{l_0}-1} \in I \\ p_{i_{l_0}-1} = a \bmod p_{i_{l_0}}}} \frac{1}{p} \leq \sum_{\substack{n \in I \\ n = a \bmod p_{i_{l_0}}}} \frac{1}{n} \ll \frac{\log X}{\exp(\sqrt{\log X})} \ll \exp(-(1+o(1))\sqrt{\log X})$$

since $n, p_{i_{l_0}} \geq \exp(\sqrt{\log X})$ (we could have done slightly better using the Brun-Titchmarsh inequality, but this is not necessary here).

The factor $\frac{\exp(O(k))}{(\log\log X)^{2k}}$ in (7.11) is $\exp(o(\sqrt{\log X}))$ and so can be absorbed into the $o(1)$ error in the preceding estimate, arriving at an upper bound of

$$\ll \exp(-(1+o(1))\sqrt{\log X}) \sum_{p_1, \dots, p_{i_{l_0}-2}, p_{i_{l_0}}, \dots, p_k, p'_1, \dots, p'_k \in I} \frac{\prod_{l=1}^{r} (p_{i_l} 1_{p_{i_l} = p'_{j_l}})}{p_1 \dots p_{i_{l_0}-2} p_{i_{l_0}} \dots p_k p'_1 \dots p'_k}.$$

Each of the variables p_j or p'_j that is not of the form p_{i_l} or p_{j_l} for some l can be summed using

$$\sum_{p \in I} \frac{1}{p} \ll \log\log X, \tag{7.12}$$

and then noting that $O(\log\log X)^{2k} \ll \exp(o(\sqrt{\log X}))$, we arrive (after using the constraints $p_{i_l} = p'_{j_l}$ to collapse the sum) at

$$\ll \exp(-(1+o(1))\sqrt{\log X}) \sum_{p_{i_1}, \dots, p_{i_r} \in I} \frac{1}{p_{i_1} \dots p_{i_r}},$$

and a further application of (7.12) then gives a total contribution of $\exp(-(1+o(1))\sqrt{\log X})$ for (7.11). Finally, the total number of choices for r and $(i_1, j_1), \dots, (i_r, j_r)$ may be crudely bounded by $k \times k^{2k}$, so we have a net bound of

$$k \times k^{2k} \times \exp(-(1+o(1))\sqrt{\log X}) = o(1).$$

This concludes the proof of (7.5).

Finally, we prove (7.6). We consider first the contribution of the case where there are no collisions, so that $\{p_1, \dots, p_k\}$ is disjoint from $\{p'_1, \dots, p'_k\}$. In this case, we have

$$\mathbb{P}(\gamma, \gamma' \subset \mathbf{G}_X) = \frac{1}{p_1 \dots p_k p'_1 \dots p'_k}$$

and so this contribution to (7.6) is bounded by

$$\sum_{\gamma, \gamma' \in \Gamma : \gamma(k) = \gamma'(k)} \frac{w_\gamma w_{\gamma'}}{p_1 \dots p_k p'_1 \dots p'_k}.$$

One can bound this by the probability that $\gamma(k) = \gamma'(k)$, where γ, γ' are selected as in the proof of (7.4). But if we fix the variables $p'_1, \dots, p'_k, p_1, \dots, p_{k-1}$, then the constraint $\gamma(k) = \gamma'(k)$ is only satisfiable for a single value p_k^0 of p_k in I, and so the probability here can be bounded using Lemma 7.1 by

$$\ll \frac{1}{p_k^0 \log\log X} \ll \frac{1}{\exp(\sqrt{\log X})}$$

which is acceptable.

Now we consider the contribution of the case where there is at least one collision. By the computations used to prove (7.5), this contribution is bounded by

$$\sum_{r=1}^{k} \sum_{(i_1,j_1),\dots,(i_r,j_r)} \sum_{\gamma,\gamma'\in\Gamma:\gamma(k)=\gamma'(k)} w_\gamma w_{\gamma'} \prod_{l=1}^{r} 1_{p_{i_l}=p'_{j_l}} 1_{p_{i_l}|p_1+\dots+p_{i_l-1}-p'_1-\dots-p'_{j_l-1}} \\ \frac{p_{i_1}\dots p_{i_r}}{p_1\dots p_k p'_1\dots p'_k}.$$

If one does not have $i_l = l$ for all $l = 1,\dots,r$, then we can discard the $\gamma(k) = \gamma'(k)$ constraint and use the computations used to prove (7.5) to obtain a bound of $\exp(-(1+o(1))\sqrt{\log X})$, which is acceptable. Thus we may assume that $i_l = l$ for all $l = 1,\dots,r$.

Now suppose that $r < k$, so that p_k is not one of the p'_{i_l}. If we fix $p'_1,\dots,p'_k,p_1,\dots,p_{k-1}$, then as before the constraint $\gamma(k) = \gamma'(k)$ is satisfiable for only a single value of p_k. Repeating the argument used to control (7.11), we see that the contribution of this case is also $\exp(-(1+o(1))\sqrt{\log X})$, which is acceptable. Thus we may assume that $r = k$. The constraint $\gamma(k) = \gamma'(k)$ is now automatic, and we simplify this contribution to

$$\sum_{(1,j_1),\dots,(k,j_k)} \sum_{\gamma,\gamma'\in\Gamma} w_\gamma w_{\gamma'} \prod_{l=1}^{k} \frac{1_{p_l=p'_{j_l}} 1_{p_l|p_1+\dots+p_{l-1}-p'_1-\dots-p'_{j_l-1}}}{p_l} \tag{7.13}$$

where the outer sum is over permutations $(j_1,\dots,j_k)$ of $(1,\dots,k)$.

Suppose that $j_l \neq l$ for some $1 \le l \le k$. Let l_0 be the least such l, so that $j_l = l$ for $l < l_0$ and $j_{l_0} > l_0$. Then the constraint

$$p_{l_0}|p_1+\dots+p_{l_0-1}-p'_1-\dots-p'_{j_{l_0}-1}$$

simplifies (using $p_l = p'_{j_l}$) to

$$p_{l_0}|p'_{l_0}+\dots+p'_{j_{l_0}-1}$$

which is a non-trivial constraint since $j_{l_0} > l_0$. In particular, if we fix $p_1,\dots,p_{l_0-1},p_{l_0+1},\dots,p_k$, and hence all of the p'_j except $p'_{j_{l_0}}$, we see that p_{l_0} is constrained to be a prime factor of a number of size at most $kX^{1/200}$; since p_{l_0} lies in I, we see that there are at most $O(\sqrt{\log X})$ choices for p_{l_0}, and the total sum of $\frac{1}{p_{l_0}}$ across these choices is thus $O(\sqrt{\log X}\exp(-\sqrt{\log X}))$. Using this and Lemma 7.1, and discarding all the other constraints $1_{p_l|p_1+\dots+p_{l-1}-p'_1-\dots-p'_{j_l-1}}$, we bound the contribution of (7.13) of a single such $(1,j_1),\dots,(k,j_k)$ (which determines l_0) as

$$\exp(-(1+o(1))\sqrt{\log X}) \sum_{p_1,\dots,p_{l_0-1},p_{l_0+1},\dots,p_k\in I} \frac{1}{p_1\dots p_{l_0-1}p_{l_0+1}\dots p_k}$$

which by (7.12) is bounded by $O(\log\log X)^k \exp(-(1+o(1))\sqrt{\log X}) \le \exp(-(1+o(1))\sqrt{\log X})$ which is acceptable.

The only remaining case occurs when $j_l = l$ for all $1 \le l \le k$, at which point γ' is equal to γ, the constraints $p_l|p_1+\dots+p_{l-1}-p'_1-\dots-p'_{j_l-1}$ can be discarded, and the contribution to (7.13) collapses to

$$\sum_{\gamma\in\Gamma} w_\gamma^2 \frac{1}{p_1\dots p_k}.$$

This can be bounded by the expectation of w_γ, where γ is chosen as in the proof of (7.4). But from Lemma 7.1, this expectation is at most $\exp(O(k))/(\log\log X)^k$, which from the choice (7.1) of k is (barely) $O(\log^{-100} X)$, which is acceptable. This proves (7.6), and tracing back all the preceding reductions we finally arrive at Theorem 1.9.

8. A Vinogradov type result

In this section we prove the following result of Vinogradov type which was needed in the proof of Lemma 6.3.

Proposition 8.1. *Let X be large, let I_1, I_2, I_3 denote the intervals*

$$I_1 := (X, 3X]; \quad I_2 := (5X, 7X]; \quad I_3 := (3X, 5X]$$

and let $m \in [-X, X]$ be odd. Let A be an integer. Then there are $\asymp X^2/(\log X)^3$ triples (p_1, p_2, p_3) of primes with $p_1 \in I_1, p_2 \in I_2, p_3 \in I_3$ such that

$$m = -p_1 + p_2 - p_3$$

and such that $A - p_1, A - p_1 + p_2$ are not divisible by p^2 for any $p \leq w$.

The proof of Proposition 8.1 relies crucially on the following slight modification of Vinogradov's result on representing odd integers as sums of three primes.

Lemma 8.2. *We have, for m odd and fixed square-free k, and any a_1, a_2,*

$$\sum_{\substack{m=-p_1+p_2-p_3 \\ p_j \in I_j \\ p_1 \equiv a_1 \pmod{k^2} \\ p_2 \equiv a_2 \pmod{k^2}}} 1 = \frac{\mathcal{G}(m)\mathfrak{S}(m)}{(\log X)^3} \cdot \mathbf{1}_{(k,-a_1+a_2-m)=(k,a_1)=(k,a_2)=1} \cdot \frac{f((k,m))g(k)}{k\varphi(k)^3} + o\Big(\frac{X^2}{(\log X)^3}\Big),$$

where $\mathcal{G}(m) = \#\{(n_1, n_2, n_3) \in I_1 \times I_2 \times I_3 : m = -n_1 + n_2 - n_3\}$ and f, g are multiplicative functions such that

$$f(p^\alpha) = f(p) = \Big(1 + \frac{1}{(p-1)^3}\Big) \cdot \Big(1 - \frac{1}{(p-1)^2}\Big)^{-1}$$
$$g(p^\alpha) = g(p) = \Big(1 + \frac{1}{(p-1)^3}\Big)^{-1}.$$

Finally

$$\mathfrak{S}(m) := \prod_{p|M}(1 - \frac{1}{(p-1)^2}) \times \prod_{p\nmid M}(1 + \frac{1}{(p-1)^3})$$

is the usual singular series appearing in Vinogradov's three-primes theorem.

Proof. This follows from a very minor modification of a generalization of Vinogradov's result due to Ayoub [1] (we only need to handle the additional condition $p_j \in I_j, \forall j \leq 3$). □

One can easily compute that $\mathcal{G}(m) \asymp X^2$ and $\mathfrak{S}(m), f((k,m)), g(k) \asymp 1$. Thus we have a cruder version

$$\sum_{\substack{m=-p_1+p_2-p_3 \\ p_j \in I_j \\ p_1 \equiv a_1 \pmod{k^2} \\ p_2 \equiv a_2 \pmod{k^2}}} 1 \asymp \frac{X^2}{\log^3 X}\Big(\frac{1}{k\varphi(k)^3} + o(1)\Big) \tag{8.1}$$

of the above lemma, when $a_1, a_2, -a_1 + a_2 - m$ are all coprime to k. This is the only consequence of the above lemma that we will need.

We are now ready to prove Proposition 8.1.

Proof of Proposition 8.1. Note that the only properties of the integer A which are relevant are its reductions modulo p^2 for $p \leq w$. Thus, by the Chinese remainder theorem, we may assume that $0 \leq A < \prod_{p \leq w} p^2$. We may rewrite the desired claim as

$$\sum_{\substack{m=-p_1+p_2-p_3\\ p_j \in I_j}} \mu_w^2(A-p_1)\mu_w^2(A-p_1+p_2) \asymp \frac{X^2}{\log^3 X} \tag{8.2}$$

where $\mu_w^2(n)$ is the indicator function of integers not divisible by a $p^2 \leq w$, and p_1, p_2, p_3 are understood to be prime.

Notice that for any fixed C, we have

$$\mu_w^2(A-p_1)\mu_w^2(A-p_1+p_2) \geq \mathbf{1}_{\substack{p^2 \nmid A-p_1\\ p^2 \nmid A-p_1+p_2\\ \forall p \leq C}} - \sum_{\substack{C \leq p \leq w\\ p^2 \mid A-p_1}} 1 - \sum_{\substack{C \leq p \leq w\\ p^2 \mid A-p_1+p_2}} 1. \tag{8.3}$$

We view the last two terms on the right-hand side as contributing error terms to be upper bounded using sieves. The first error term contributes to the left hand side of (8.2)

$$\begin{aligned} &\sum_{C \leq p \leq w} \sum_{\substack{m=-p_1+p_2-p_3\\ p_j \in I_j, \forall j \leq 3\\ p^2 \mid A-p_1}} 1 \\ \leq &\sum_{C \leq p \leq w} \sum_{\substack{p_1 \equiv A \pmod{p^2}\\ p_1 \in I_1}} |\{n \in I_2 \cap (m+p_1+I_3) : n, n-m-p_1 \text{ prime}\}|. \end{aligned} \tag{8.4}$$

The requirement that $n, n-m-p_1$ are prime removes two residue classes mod p for $p \leq X$ not dividing $m+p_1$, and one residue class mod p for $p \leq X$ dividing $m+p_1$. A standard upper bound sieve (see e.g. [9, Theorem 3.12]) then gives

$$\begin{aligned} |\{n \in I_2 \cap (m+p_1+I_3) : n, n-m-p_1 \text{ prime}\}| &\ll \frac{X}{\log^2 X} \prod_{p \leq X : p \mid m+p_1} (1+\frac{1}{p}) \\ &\ll \frac{X}{\log^2 X} \sum_{d \mid m+p_1} \frac{1}{d} \\ &\ll \frac{X}{\log^2 X} \sum_{d \ll \sqrt{X} : d \mid m+p_1} \frac{1}{d} \end{aligned}$$

since $m+p_1 \ll X$ and we may pair d with $\frac{m+p_1}{d}$. Thus we may upper bound (8.4) by

$$\ll \frac{X}{\log^2 X} \sum_{C \leq p \leq w} \sum_{d \ll \sqrt{X}} \frac{1}{d} \sum_{\substack{p_1 \equiv A \pmod{p^2}\\ p_1 \in I_1 : d \mid m+p_1}} 1.$$

Applying the Brun-Titchmarsh inequality, we may bound this by

$$\ll \frac{X^2}{\log^3 X} \sum_{C \leq p \leq w} \sum_{d} \frac{1}{d\varphi([p^2, d])}$$

where $[p^2, d]$ denotes the least common multiple of p^2 and d. By Euler products the innermost sum is $O(1/p^2)$, and so this expression is $O\left(\frac{1}{C}\frac{X^2}{\log^3 X}\right)$.

Similarly, the second error term in (8.3) contributes to the left hand side of (8.2)

$$\begin{aligned}&\sum_{C\leq p\leq w}\sum_{\substack{m=-p_1+p_2-p_3\\ p_j\in I_j,\forall j\leq 3\\ p^2|A-p_1+p_2}} 1\\ &= \sum_{C\leq p\leq w}\sum_{\substack{p_3\in I_3\\ p^2|A+m+p_3}} |\{n\in I_1\cap(I_2-m-p_3) : n, m+n+p_3 \text{ prime}\}|.\end{aligned} \tag{8.5}$$

As before, standard upper bound sieves give

$$|\{n\in I_1\cap(I_2-m-p_3) : n, m+n+p_3 \text{ prime}\}| \ll \frac{X}{\log^2 X}\sum_{d\ll\sqrt{X}:d|m+p_3}\frac{1}{d}$$

and an application of Brun-Titchmarsh as before shows that the second error term is also $O\left(\frac{1}{C}\frac{X^2}{\log^3 X}\right)$.

It remains to understand the contribution of the main term. Observe that for each prime p, there are at least $(p-1)(p-2) > p^2-3p$ pairs of residue classes $a_1, a_2 \pmod p$ such that $a_1, a_2, -a_1+a_2-m$ are all coprime to p. Thus, there are at least p^4-3p^3 pairs of residue classes $a_1, a_2 \pmod{p^2}$ such that $a_1, a_2, -a_1+a_2-m$ are coprime to p. Of these pairs, there are at most $2p^2$ pairs such that one of $A-a_1$ or $A-a_1+a_2$ is divisible by p^2. By the Chinese remainder theorem, setting $\mathcal{P} := \prod_{p\leq C} p$, we conclude that there are at least $\prod_{p\leq C}(p^4-3p^3-2p^2)$ residue classes $a_1, a_2 \pmod{\mathcal{P}^2}$ such that $a_1, a_2, -a_1+a_2-m$ are coprime to all primes $p\leq C$, and $A-a_1, A-a_1+a_2$ are not divisible by p^2 for any $p\leq C$. For any such fixed tuple (a_1, a_2) we apply (8.1) to conclude (for X sufficiently large depending on C) that

$$\sum_{\substack{m=-p_1+p_2-p_3\\ p_j\in I_j,\forall j\leq 3\\ p_1\equiv a_1 \pmod{\mathcal{P}^2}\\ p_2\equiv a_2 \pmod{\mathcal{P}^2}}} 1 \gg \frac{1}{\mathcal{P}\varphi(\mathcal{P})^3}\cdot\frac{X^2}{\log^3 X}.$$

Summing, we may thus lower bound the contribution of the main term (for C large) by

$$\begin{aligned}\gg \frac{\prod_{p\leq C}(p^4-3p^3-2p^2)}{\mathcal{P}\varphi(\mathcal{P})^3}\frac{X^2}{\log^3 X} &= \prod_{p\leq C}\left(\frac{p^4-3p^3-2p^2}{p(p-1)^3}\right)\frac{X^2}{\log^3 X}\\ &= \prod_{p\leq C}\left(\frac{(p-1)^3-5p+1}{(p-1)^3}\right)\frac{X^2}{\log^3 X} \gg \frac{X^2}{\log^3 X}\end{aligned}$$

with the implied constant uniform in C. For C large enough (and X sufficiently large depending on C), this lower bound dominates the two error terms, and we obtain the claim. □

References

[1] R. Ayoub, *On Rademacher's extension of the Goldbach-Vinogradoff theorem*, Trans. Amer. Math. Soc. **74**, (1953). 482491.

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

[3] Y. Buttkewitz, C. Elsholtz, *Patterns and complexity of multiplicative functions*, J. Lond. Math. Soc. (2) **84** (2011), no. 3, 578–594.

[4] S. Chowla, The Riemann hypothesis and Hilbert's tenth problem, Gordon and Breach, New York, 1965.

[5] N. Frantzikinakis, B. Host, *Asymptotics for multilinear averages of multiplicative functions*, preprint. `arXiv:1502.02646`.

[6] B. Green, T. Tao, *Linear equations in primes*, Ann. of Math. (2) **171** (2010), no. 3, 1753–1850.

[7] B. Green, T. Tao, *The Möbius function is strongly orthogonal to nilsequences*, Ann. of Math. (2) **175** (2012), no. 2, 541–566.

[8] B. Green, T. Tao, T. Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Ann. of Math. (2) **176** (2012), no. 2, 1231–1372.

[9] H. Halberstam, H.-E. Richert, Sieve Methods. London Academic Press, 1974.

[10] G. Harman, J. Pintz, D. Wolke, *A note on the Möbius and Liouville functions*, Studia Sci. Math. Hungar. **20** (1985), no. 1-4, 295–299.

[11] A. Hildebrand, *On consecutive values of the Liouville function*, Enseign. Math. (2) **32** (1986), no. 3-4, 219–226.

[12] K. Matomäki, M. Radziwiłł, *Multiplicative functions in short intervals*, preprint. `arXiv:1501.04585`.

[13] K. Matomäki, M. Radziwiłł, T. Tao, *An averaged form of Chowla's conjecture*, preprint. `arXiv:1503.05121`.

[14] E. M. Stein, Harmonic Analysis: real-variable methods, orthogonality, and oscillatory integrals. With the assistance of Timothy S. Murphy. Princeton Mathematical Series, 43. Monographs in Harmonic Analysis, III. Princeton University Press, Princeton, NJ, 1993.

Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland
E-mail address: `ksmato@utu.fi`

Department of Mathematics, Rutgers University, Hill Center for the Mathematical Sciences, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019
E-mail address: `maksym.radziwill@gmail.com`

Department of Mathematics, UCLA, 405 Hilgard Ave, Los Angeles CA 90095, USA
E-mail address: `tao@math.ucla.edu`