

## On the Enumeration of Finite Groups\*

P. ERDÖS

*Math. Institute, Hungarian Academy of Sciences, Budapest, Hungary*

M. RAM MURTY

*Department of Mathematics,  
McGill University, Montreal, Quebec, Canada*

AND

V. KUMAR MURTY

*School of Mathematics, Tata Institute of Fundamental Research,  
Homi Bhabha Road, Bombay, India*

*Communicated by H. Zassenhaus*

Received November 13, 1985

Let  $G(n)$  denote the number of finite groups of order  $n$  (up to isomorphism). We prove that for  $n$  squarefree,  $G(n) = \Omega(n^{1-\varepsilon})$  for any  $\varepsilon > 0$ , and that for almost all squarefree integers  $n$ ,  $\log G(n) = (1 + o(1))(\log \log n) \sum_{p|n} (\log p)/(p-1)$ . If we let  $F_k(x)$  be the number of  $n \leq x$  such that  $G(n) = k$ , then we prove  $F_k(x) = (c(a) + o(1))x/(\log \log \log x)^{a+1}$  for  $k = 2^a$ , and  $c(a)$  is an appropriate constant, as  $x \rightarrow \infty$ . If  $k \neq 2^a$ , then we show that  $F_k(x) = O(x/(\log \log x)^{1-\varepsilon})$ . © 1987 Academic Press, Inc.

### 1. INTRODUCTION

Let  $G(n)$  denote the number of groups (up to isomorphism) of order  $n$ . With the recent classification of finite simple groups, we know that

$$G(n) \leq n^{c(\log n)^2}$$

for some  $c > 0$ . (See Neumann [9].) This upper bound can be significantly improved if we confine our attention to certain classes of groups. For example, if  $n$  is squarefree, it was shown in [7] that

$$G(n) \leq \varphi(n),$$

\* Research partially supported by NSERC Grant U0237.

where  $\varphi$  is the Euler  $\varphi$ -function. Moreover, it was proved in [8] that

$$\sum_{n \leq x} \mu^2(n) \log G(n) = (1 + o(1)) cx \log \log x$$

as  $x \rightarrow \infty$ , for a certain constant  $c$ .

This was established by utilising the following beautiful formula due to Hölder [6]. Let  $V_p(n)$  denote the number of prime divisors of  $n$  which are  $\equiv 1 \pmod p$ . Then for  $n$  squarefree.

$$G(n) = \sum_{d|n} \prod_{p|d} \left( \frac{p^{V_p(n/d)} - 1}{p - 1} \right), \tag{1}$$

where the inner product runs over prime divisors  $p$  of  $d$ .

Formula (1) has other applications. It will be the essential ingredient in the proofs of the main theorems of this paper.

**THEOREM 1.** *For  $n$  squarefree,*

$$G(n) = \Omega(n^{1-\varepsilon})$$

for every  $\varepsilon > 0$ .

*Remark.* This theorem shows that the estimate  $G(n) \leq \varphi(n)$  for  $n$  squarefree, is nearly best possible.

**THEOREM 2.** *For almost all squarefree  $n$ ,*

$$\log G(n) = (1 + o(1))(\log \log n) \sum_{p|n} \frac{\log p}{p-1},$$

or in other words,

$$\frac{\log G(n)}{\log \log n}$$

has a distribution function.

*Remark.* A weaker version of this result was proved in [7].

The main interest in formula (1) is that it can be utilised in obtaining information concerning the distribution of the values of  $G(n)$ . To this end, let us define

$$F_k(x) = \text{card}(n \leq x: G(n) = k).$$

The following theorem shows that  $G(n)$  is a power of 2 more often than any other value.

THEOREM 3. Let  $a$  be a nonnegative integer.

(i) if  $k \neq 2^a$ , then  $F_k(x) \ll \frac{x}{(\log \log x)^{1-\varepsilon}}$ .

(ii) if  $k = 2^a$ , then for  $c(a) = e^{-\gamma}/a!$

$$F_k(x) = \frac{(c(a) + o(1))x}{(\log \log \log x)^{a+1}}.$$

Remarks. (1) With a little more care, (i) can be improved to

$$o\left(\frac{x}{\log \log x}\right).$$

(2) It is conceivable that if  $k = 3$ ,

$$F_3(x) \gg \frac{x}{(\log \log x)^{1+o(1)}},$$

if hopeless, but obvious, conjectures concerning the distribution of primes are assumed. It is too early to predict the behaviour of  $F_k(x)$  when  $k$  is not a power of 2.

(3) It should be noted that the constants implicit in (i) and (ii) above depend on  $k$ .

(4) Spiro has recently found an infinite set  $S$ , which includes the Fibonacci numbers, such that for any  $\varepsilon > 0$ , and any  $k \in S$ ,  $F_k(x) \gg_{k,\varepsilon} x(\log x)^{-\varepsilon}$ . In particular, this holds for  $k = 3$ .

## 2. AN $\Omega$ RESULT FOR $G(n)$

In this section, we prove Theorem 1. Let  $N$  and  $D$  be defined by

$$\log N = \sum_{p \leq x} \log p,$$

and

$$\log D = \sum_{p \leq y} \log p,$$

where  $y \leq x$ , and  $x, y$  shall be chosen later. Utilising the explicit formula (1), we find that

$$G(N) \geq \prod_{p|D} \left( \frac{p^{V_p(N/D)} - 1}{p - 1} \right).$$

If we denote by  $\pi(x, q)$ , the number of primes  $p \leq x, p \equiv 1 \pmod{q}$ , then it is easily seen that

$$V_p(N/D) = \pi(x, p) - \pi(y, p).$$

By elementary estimates, it follows that

$$\log G(N) \geq \sum_{p \leq y} \{ \pi(x, p) - \pi(y, p) \} \log p + O(y).$$

We need the following lemmas:

LEMMA 1.

$$\sum_{p \leq x} \pi(x, p) \log p = (1 + o(1))x \quad \text{as } x \rightarrow \infty.$$

*Proof.* It is easy to see that

$$\sum_{q \leq x} \log(q - 1) = \sum_{p \leq x} \{ \pi(x, p) + \pi(x, p^2) + \dots \} \log p,$$

where the sum on the left-hand side of the above equation ranges over primes  $q \leq x$ .

Using the trivial estimate,

$$\pi(x, p^2) \leq x/p^2,$$

we find

$$\sum_{\alpha \geq 2} \sum_{(\log x)^\alpha \leq p^\alpha \leq x} \pi(x, p^\alpha) \log p = O\left(\frac{x}{(\log x)^2}\right).$$

On the other hand, by the Brun–Titchmarsh inequality,

$$\sum_{\alpha \geq 2} \sum_{p^\alpha \leq (\log x)^4} \pi(x, p^\alpha) \log p = O\left(\frac{x}{\log x}\right).$$

This proves that

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

since by the prime number theorem

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

LEMMA 2. *There is an absolute constant  $c > 0$ , such that*

$$\sum_{p \leq x^{1-\varepsilon}} \pi(x, p) \log p \geq (1 - c\varepsilon)x$$

as  $x \rightarrow \infty$ , for any  $\varepsilon > 0$ .

*Proof.* By Lemma 1, we see that it suffices to show that

$$\sum_{x^{1-\varepsilon} < p \leq x} \pi(x, p) \log p \ll \varepsilon x.$$

Indeed,

$$\sum_{x^{1-\varepsilon} < p \leq x} \pi(x, p) \log p \leq \log x \sum_{x^{1-\varepsilon} < p \leq x} \pi(x, p).$$

The last sum can be written as

$$\sum_{t \leq x^\varepsilon} N(x, t),$$

where  $N(x, t)$  is the number of solutions  $p$  of  $p - 1 = qt$ , where  $p$  and  $q$  are prime numbers. By any sieve method,

$$N(x, t) = O\left(\frac{x}{\varphi(t) \log^2(x/t)}\right).$$

This estimate now yields the desired result, as

$$\sum_{t \leq x^\varepsilon} \frac{1}{\varphi(t)} = O(\varepsilon \log x).$$

We can now complete the proof of our theorem. We find by Lemma 1, that

$$\log G(N) \geq \sum_{p \leq y} \pi(x, p) \log p + O(y).$$

Choosing  $y = x^{1-\varepsilon}$ , yields, by Lemma 2,

$$\log G(N) \geq (1 - c\varepsilon)x + O(x^{1-\varepsilon})$$

as  $x \rightarrow \infty$ . By the prime number theorem,

$$\log N = (1 + o(1))x,$$

and hence,

$$\log G(N) \geq (1 - c\varepsilon + o(1)) \log N,$$

as desired.

3. PROOF OF THEOREM 2

We want to establish that  $\log G(n)$  has a distribution function for squarefree  $n$ . Recall that in [7], it was shown that for square-free  $n$ ,

$$G(n) \leq \prod_{p|n} p^{V_p(n)}.$$

Therefore,

$$\log G(n) \leq \sum_{p|n} V_p(n) \log p.$$

Consider the set

$$L_1 = \{n \leq x: V_p(n) \geq 1 \text{ for some } p | n, p > (\log \log x)\}.$$

We begin by showing that  $|L_1| = o(x)$ . We need

LEMMA 3. *Let  $p$  be a prime. Then*

$$\sum_{\substack{q < x \\ q \equiv 1 \pmod{p}}} \frac{1}{q} \leq \frac{C(\log \log x + \log p)}{p}$$

for some absolute constant  $C$ .

*Proof.* See [3].

LEMMA 4.

$$|L_1| = O\left(\frac{x}{\log \log \log x}\right).$$

*Proof.* Clearly, the size of  $L_1$  is bounded by

$$\sum_{\substack{q \equiv 1 \pmod{p} \\ p > (\log \log x)}} \frac{x}{pq}$$

and by Lemma 3, this is dominated by

$$\sum_{p > (\log \log x)} \frac{x(\log \log x + \log p)}{p^2} = O\left(\frac{x}{\log \log \log x}\right)$$

as desired.

We may therefore assume that  $V_p(n) = 0$  for all  $p \mid n$ , with  $p > \log \log n$ , so that for almost all squarefree  $n$ ,

$$G(n) \leq \sum_{\substack{p \mid n \\ p < \log \log n}} V_p(n) \log p. \quad (2)$$

We show next that  $V_p(n) \leq 2(\log \log n)/p$  for almost all  $n$ , uniformly for  $p < \log \log n$ .

LEMMA 5. Let  $P$  be a set of primes satisfying  $\sum_{p \leq x, p \in P} (1/p) = t_x$ , with  $t_x \rightarrow \infty$  as  $x \rightarrow \infty$ .

Set  $\omega_p(n) = \sum_{p \mid n, p \in P} 1$  and fix  $\varepsilon > 0$ . The number of integers  $n < x$  for which the inequalities:

$$(1 - \varepsilon)t_x < \omega_p(n) < (1 + \varepsilon)t_x$$

do not hold is less than

$$x \exp(-\eta t_x),$$

where  $\eta = \eta(\varepsilon)$  is a positive constant which depends only on  $\varepsilon$ .

*Proof.* This result follows easily from the method of Hardy and Ramanujan [5] combined with Brun's method. As the derivation closely follows the method of [5], we suppress the details. (The referee informs us that a sharper version of this lemma appears in K. K. Norton, *Illinois J. Math.* **20** (1976), 681–705.)

COROLLARY. Uniformly for  $p < (\log \log x)^{1-\varepsilon}$ ,

$$(1 - \varepsilon) \frac{\log \log x}{p-1} < V_p(n) < (1 + \varepsilon) \frac{\log \log x}{p-1}$$

is satisfied for all  $n \leq x$  with at most  $O(x/(\log \log x)^A)$  exceptions, for any  $A > 0$ .

*Proof.* By Lemma 5,

$$(1 - \varepsilon) \frac{\log \log x}{p-1} < V_p(n) < (1 + \varepsilon) \frac{\log \log x}{p-1}$$

is satisfied for all  $n \leq x$  apart from

$$O\left(x \exp\left(-\eta \frac{\log \log x}{p}\right)\right)$$

exceptions. We sum this over  $p < (\log \log x)^{1-\varepsilon}$  to obtain the desired result.

LEMMA 6. *The number of  $n \leq x$  divisible by a prime  $p$  in the range  $(\log \log x)^{1-\varepsilon} < p < \log \log x$  is  $O(\varepsilon x)$ .*

*Proof.* The number of such  $n \leq x$  is clearly bounded by

$$\sum'_p \frac{x}{p},$$

where the dash on the sum indicates that  $p$  is in the specified range.

Using the elementary fact

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

The result follows immediately.

From (2), we find from the preceding that apart from  $O(\varepsilon x)$  squarefree numbers  $n \leq x$ , we have

$$\log G(n) \leq \sum_{\substack{p|n \\ p < (\log \log n)^{1-\varepsilon}}} (1 + o(1)) \frac{\log \log n}{p-1} \log p.$$

For the lower bound, set

$$d = \prod_{\substack{p|n \\ p < (\log \log n)^{1-\varepsilon}}} p.$$

With the exception of  $o(x)$  of the  $n \leq x$ ,

$$\sum_{\substack{q|n \\ q < (\log \log n)^{1-\varepsilon}}} 1 \ll \log \log \log \log n.$$

Therefore,

$$V_\rho(n/d) = V_\rho(n) + O(\log \log \log \log n).$$

Now, by Hölder's formula (1) and the corollary to Lemma 5, we get

$$\begin{aligned} \log G(n) &\geq \sum_{\substack{p|n \\ p < (\log \log n)^{1-\varepsilon}}} (V_\rho(n/d) - 1) \log p \\ &\geq \sum_{\substack{p|n \\ p < (\log \log n)^{1-\varepsilon}}} V_\rho(n) \log p \\ &\quad + O((\log \log \log n)(\log \log \log \log n)^2) \\ &\geq (1 + o(1))(\log \log x) \sum_{\substack{p|n \\ p < (\log \log n)^{1-\varepsilon}}} (\log p)/(p-1) \end{aligned}$$

except possibly for  $o(x)$  of the squarefree  $n \leq x$ .

Let  $S_\varepsilon(x)$  denote the set of integers  $n$  with  $\sqrt{x} \leq n \leq x$  for which

$$\sum_{\substack{p|n \\ p > (\log \log n)^{1-\varepsilon}}} \frac{\log p}{p-1} > \varepsilon.$$

Then

$$\begin{aligned} \varepsilon |S_\varepsilon(x)| &< \sum_{\sqrt{x} \leq n \leq x} \sum_{\substack{p|n \\ p > (\log \log n)^{1-\varepsilon}}} \frac{\log p}{p-1} \leq x \sum_{p > (\log \log x)^{1-\varepsilon}} \frac{\log p}{p(p-1)} \\ &\leq x \frac{\log \log \log x}{(\log \log x)^{1-\varepsilon}}. \end{aligned}$$

Thus, we have

$$\sum_{p|n} \frac{\log p}{p-1} - \varepsilon \leq \sum_{\substack{p|n \\ p < (\log \log n)^{1-\varepsilon}}} \frac{\log p}{p-1} < \sum_{p|n} \frac{\log p}{p-1}$$

except for  $O(\varepsilon x)$  of the  $n \leq x$ . The function

$$\sum_{p|n} \frac{\log p}{p-1}$$

is additive, and by Theorem 5.1 of [2], it has a *continuous* distribution function. Thus, the same can be said of

$$\sum_{\substack{p|n \\ p < (\log \log n)^{1-\varepsilon}}} \frac{\log p}{p-1}$$

and of  $(\log G(n))/\log \log n$ . This completes the proof of Theorem 2.

#### 4. AN UPPER BOUND FOR $F_k(x)$

Let

$$F_k = \{n \leq x: G(n) = k\}$$

and denote by  $p_n$  the smallest prime divisor of  $n$ . Let  $\varepsilon > 0$  be arbitrary. Let us write

$$F_k = S_1 \cup S_2,$$

where in  $S_1$ ,  $p_n < (\log \log x)^{1-\varepsilon}$  and in  $S_2$ ,  $p_n > (\log \log x)^{1-\varepsilon}$ . We use Brun's method to estimate  $|S_1|$ .

LEMMA 7. Let  $c$  be a fixed positive integer and suppose that  $p < (\log \log x)^{1-\epsilon}$ . Then

$$\text{card}(n \leq x: p \mid n, \text{ and } V_p(n) \leq c)$$

is

$$O\left(\frac{x}{p^{1+c}} (\log \log x)^c \exp\left(-\frac{\log \log x}{p}\right)\right).$$

*Proof.* By Brun's sieve, the number with  $V_p(n) = c$  is

$$\ll \frac{x}{p} \sum_{\substack{q_1, \dots, q_c \equiv 1 \pmod{p} \\ q_1 \cdots q_c < x/p}} \frac{1}{q_1 \cdots q_c} \prod_{r < \xi} \left(1 - \frac{1}{r}\right),$$

where  $\xi = x^{1/\log \log x}$ . It follows that this is bounded by

$$\ll \frac{x}{p} \left\{ \sum_{\substack{q \leq x \\ q \equiv 1 \pmod{p}}} \frac{1}{q} \right\}^c \exp\left(-\frac{\log \log x}{p}\right).$$

By well-known estimates (see [3]) it follows that the above is

$$\ll \frac{x}{p^{1+c}} (\log \log x)^c \exp\left(-\frac{\log \log x}{p}\right).$$

*Proof of Theorem 3(i).* Let  $p = p_n$  and suppose that  $V_p(n) \geq k + 1$ . It follows from Theorem 1.1 of [8] that

$$G(n) \geq \frac{p^{V_p(n)} - 1}{p - 1} \geq 2^{V_p(n) - 1} \geq 2^k > k.$$

Hence, if  $G(n) = k$ , we must have  $V_p(n) \leq k$ . By Lemma 7 (or by the corollary to Lemma 5), it follows that for any  $A > 0$ ,

$$|S_1| = O\left(\frac{x}{(\log \log x)^A}\right).$$

Now we write

$$S_2 = S_3 \cup S_4,$$

where  $S_3$  consists of squarefree numbers and  $S_4$  consists of those elements with a squared prime factor. As  $p_n > (\log \log x)^{1-\epsilon}$ , for elements of  $S_2$ , we find

$$|S_4| \leq \sum_{p > (\log \log x)^{1-\epsilon}} \frac{x}{p^2} = O\left(\frac{x}{(\log \log x)^{1-\epsilon}}\right).$$

It remains to estimate  $S_3$ . If  $n \in S_3$ , then for any  $p \mid n$ ,  $V_p(n) \leq 1$ . For if  $V_p(n) \geq 2$  then

$$k = G(n) \geq \frac{p^{V_p(n)} - 1}{p - 1} \geq p_n > (\log \log x)^{1-\epsilon}$$

a contradiction for sufficiently large  $x$ .

For the sake of convenience, we introduce for every natural number  $n$ , the graph of  $n$ , denoted by  $g(n)$ . The vertices of this graph are the prime divisors of  $n$ , and two prime divisors  $p, q$  of  $n$  are joined if  $p \mid (q - 1)$ . If  $g(n)$  has connected components given by  $g(n_i)$ , then it follows from Hölder's formula that

$$G(n) = \prod G(n_i)$$

when  $n$  is squarefree.

The fact that  $V_p(n) \leq 1$ , means that for  $n \in S_3$ ,  $g(n)$  consists of disjoint segments of the type



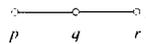
If the subgraph



does not appear in  $g(n)$ , then  $g(n)$  consists of



in which case  $G(n)$  is a power of 2, contrary to hypothesis. Hence, for  $n \in S_3$ ,  $g(n)$  contains the subgraph



Thus,

$$|S_3| \leq \sum_{\substack{p > (\log \log x)^{1-\epsilon} \\ q \equiv 1 \pmod{p} \\ r \equiv 1 \pmod{q}}} \frac{x}{pqr}$$

Lemma 3 applies, and we get

$$|S_3| \ll \sum_{\substack{p > (\log \log x)^{1-\epsilon} \\ q \equiv 1 \pmod{p}}} \frac{x}{pq^2} (\log \log x + \log q) = \Sigma_1 + \Sigma_2 \quad (\text{say}).$$

Then,

$$|\Sigma_1| \ll \sum_{\substack{p > (\log \log x)^{1-\epsilon} \\ t \geq 1}} \frac{x \log \log x}{p^3 t^2}.$$

The latter sum is easily seen to be

$$O\left(\frac{x}{(\log \log x)^{1-2\epsilon}}\right).$$

The term  $\Sigma_2$  with the  $(\log q)/q^2$  term is handled similarly. This completes the proof of (i).

*Remark.* The above argument shows that

$$|S_1| = O\left(\frac{x}{(\log \log x)^A}\right)$$

for any  $k > 1$ , and any  $A > 0$ . It is equally clear that

$$|S_4| = O\left(\frac{x}{(\log \log x)^{1-\epsilon}}\right)$$

for any  $k \geq 1$ .

### 5. THE ASYMPTOTIC FORMULA FOR $F_k(x)$ , $k = 2^a$

By the remarks following the proof of (i), it is clear that we need to establish an asymptotic formula for the number of squarefree  $n \leq x$ , whose graph  $g(n)$  has exactly  $a$  connected components of the form



together with a finite set of disjoint vertices. The case  $k = 1$ , when  $a = 0$ , has already been dealt with in Erdős [3]. We begin by considering the case  $k = 2$ , corresponding to  $a = 1$ . We must enumerate squarefree numbers  $n \leq x$  of the form  $n = pqm$ , where  $(m, \varphi(m)) = 1$ ,  $q \equiv 1 \pmod p$  and  $(pm, \varphi(pm)) = 1$ ,  $(qm, \varphi(qm)) = 1$ . For any fixed pair of primes  $p, q$ , with  $q \equiv 1 \pmod p$ , let  $A_{pq}(x)$  denote the number of squarefree  $n \leq x$  satisfying the above conditions. It is also clear that we need only consider  $n \in S_3$ , by the remarks at the end of the last section. Hence we may take  $p > (\log \log x)^{1-\epsilon}$ , and assume that all prime divisors of  $m$  are greater than  $(\log \log x)^{1-\epsilon}$ .

LEMMA 8.

$$\sum_{p > (\log \log x)^{1+\epsilon}} A_{pq}(x) = O\left(\frac{x}{(\log \log x)^\epsilon}\right).$$

*Proof.* By Lemma 3, we have

$$\sum_{p > (\log \log x)^{1+\epsilon}} \frac{x}{pq} \ll \sum_{p > (\log \log x)^{1+\epsilon}} x \frac{(\log \log x + \log p)}{p^2},$$

and this latter sum is  $O(x/(\log \log x)^\epsilon)$  by an easy computation.

The lemma shows that we may take  $p$  satisfying  $(\log \log x)^{1+\epsilon} < p < (\log \log x)^{1+\epsilon}$  in the following discussion. We need:

LEMMA 9. *Let  $p$  be a prime  $< (\log x)^c$  (where  $c$  is an arbitrary constant). Then*

$$\sum_{\substack{q < x \\ q \equiv 1 \pmod{p}}} \frac{1}{q} = \frac{\log \log x}{p-1} + O\left(\frac{\log p}{p}\right).$$

*Proof.* This is a straight forward consequence of the Siegel–Walfisz theorem and the Brun–Titchmarsh inequality on the number of primes in an arithmetic progression.

*Remark.* The referee informs us that the result is true without the restriction on  $p$  (see, e.g., the paper of Norton mentioned earlier or C. Pomerance, *J. Reine Angew. Math.* **293/294** (1977), 217–222).

COROLLARY. *Let  $\xi = x^{1/\log \log x}$ . Then*

$$\sum_{\substack{\xi < q < x \\ q \equiv 1 \pmod{p}}} \frac{1}{q} = O\left(\frac{\log \log \log x}{p}\right)$$

*uniformly for  $p < (\log \log x)^A$ , for any constant  $A > 0$ .*

*Proof.* As

$$\sum_{\substack{q < \xi \\ q \equiv 1 \pmod{p}}} \frac{1}{q} = \frac{\log \log x - \log \log \log x}{p-1} + O\left(\frac{\log p}{p}\right)$$

the result follows easily from Lemma 9.

We may also take  $q < \xi$  as our next lemma shows.

LEMMA 10.

$$\sum_{\substack{q > \xi \\ q \equiv 1 \pmod{p} \\ p > (\log \log x)^{1-\epsilon}}} A_{pq}(x) \ll \left( \frac{x}{(\log \log x)^{1-\epsilon}} \right).$$

*Proof.* Clearly  $A_{pq}(x) \leq x/pq$ . Since  $p < (\log \log x)^{1+\epsilon}$ , the corollary to Lemma 9 implies that the above sum is bounded by

$$\begin{aligned} \sum_{\substack{x \geq q > \xi \\ q \equiv 1 \pmod{p} \\ p > (\log \log x)^{1-\epsilon}}} \frac{x}{pq} &= O \left( \sum_{p > (\log \log x)^{1-\epsilon}} \frac{x \log \log \log x}{p^2} \right) \\ &= O \left( \frac{x}{(\log \log x)^{1-\epsilon}} \right). \end{aligned}$$

We state the following version of Brun’s sieve for the sake of convenience.

LEMMA 11. *The number of  $n \leq x$  not divisible by any of the primes  $p_1, \dots, p_s$ , where  $p_i < \xi$  is*

$$(1 + o(1))x \cdot \prod_{i=1}^s \left( 1 - \frac{1}{p_i} \right).$$

*Proof.* See Halberstam and Richert [4].

Below, we shall sometimes write  $l_m$  for  $\log_m x$ , the  $m$ -fold iterate of  $\log x$ .

LEMMA 12. *For  $(\log \log x)^{1-\epsilon} < p < (\log \log x)^{1+\epsilon}$ ,*

$$\begin{aligned} \sum_{q < \xi} A_{pq}(x) &= (1 + o(1)) \frac{x e^{-\gamma} \log \log x}{p^2 \log \log \log x} \\ &\quad \times \exp \left( -\frac{\log \log x}{p} \right) + O \left( \frac{x l_3 l_2}{p^3} \right). \end{aligned}$$

*Proof.* By Lemma 11, the number of  $pqm \leq x$  with all the prime factors of  $m > (\log \log x)^{1-\epsilon}$  and no prime factor  $s < \xi$  which is  $\equiv 1 \pmod{p}$  is

$$(1 + o(1)) \frac{x}{pq} \prod_{r < (\log \log x)^{1-\epsilon}} \left( 1 - \frac{1}{r} \right) \prod_{\substack{s < \xi \\ s \equiv 1 \pmod{p}}} \left( 1 - \frac{1}{s} \right)$$

and by familiar estimates of number theory, we find

$$(1 - \varepsilon) A_{pq}(x) \lesssim (1 + o(1)) \frac{x}{pq} \frac{e^{-\gamma}}{(\log \log \log x)} \exp\left(-\frac{\log \log x}{p}\right).$$

If we exclude those  $m \leq x/pq$  which have a prime factor  $s \equiv 1 \pmod{q}$  or a prime factor  $r \equiv 1 \pmod{p}$  with  $r > \xi$ , we find that

$$A_{pq}(x) \gtrsim (1 + o(1)) \frac{xe^{-\gamma}}{pq \log \log \log x} \exp\left(-\frac{\log \log x}{p}\right) - E_{pq}(x),$$

where

$$\begin{aligned} E_{pq}(x) &\leq \frac{x}{pq} \left\{ \sum_{s \equiv 1(q)} \frac{1}{s} + \sum_{\substack{r \equiv 1(p) \\ x \geq r \geq \xi}} \frac{1}{r} \right\} \\ &\ll \frac{x}{pq} \left\{ \frac{l_2 + \log q}{q} + \frac{l_3}{p} \right\} \end{aligned}$$

by Lemma 3 and the corollary to Lemma 9. Summing over  $q < \xi$ , we find

$$\sum_{q < \xi} E_{pq}(x) = O\left(\frac{x \log \log x}{p^3}\right) + O\left(\frac{x l_3}{p^3} (l_2 + \log p)\right),$$

and therefore

$$\begin{aligned} \sum_{q < \xi} A_{pq}(x) &= (1 + o(1)) \frac{xe^{-\gamma} \log \log x}{p^2 \log \log \log x} \\ &\quad \times \exp\left(-\frac{\log \log x}{p}\right) + O\left(\frac{x l_3 l_2}{p^3}\right). \end{aligned}$$

This completes the proof of the lemma.

*Proof of Theorem 3(ii).* We can now give the asymptotic formula for  $F_2(x)$ . Indeed, by Lemma 12, we have

$$\begin{aligned} F_2(x) &= \sum'_p (1 + o(1)) \frac{xe^{-\gamma} \log \log x}{p^2 \log \log \log x} \\ &\quad \times \exp\left(-\frac{\log \log x}{p}\right) + O\left(\frac{x}{l_2^{1-2\varepsilon}}\right), \end{aligned}$$

where the dash on the sum indicates that

$$(\log \log x)^{1-\varepsilon} < p < (\log \log x)^{1+\varepsilon}.$$

Let  $y = \log \log x$ . By partial summation, the sum

$$\sum \frac{1}{p^2} \exp(-y/p)$$

can be replaced by the integral

$$\int_{y^{1-\epsilon}}^{y^{1+\epsilon}} \exp\left(-\frac{y}{t}\right) \frac{dt}{t^2 \log t}$$

Making the substitution  $u = y/t$ , it becomes transformed into

$$\frac{1}{y} \int_{y^{-\epsilon}}^{y^{\epsilon}} \exp(-u) \frac{du}{(\log y/u)}$$

and integrating by parts shows that it is

$$\frac{1}{y \log y} + O\left(\frac{1}{y(\log y)^2}\right).$$

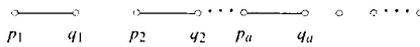
Thus,

$$F_2(x) = (1 + o(1)) \frac{e^{-y} x}{(\log \log x)^2}.$$

In the general case of  $k = 2^a$ , the main contribution comes from squarefree  $n < x$  which have the form

$$n = (p_1, q_1) \cdots (p_a, q_a) m,$$

with  $q_i \equiv 1 \pmod{p_i}$ ,  $(m, \varphi(m)) = 1$  and the graph of  $n$  is isomorphic to



where the last set of disjoint vertices correspond to the prime divisors of  $m$ . By the preceding results, we make take  $(\log \log x)^{1-\epsilon} < p_i < (\log \log x)^{1+\epsilon}$ , for  $1 \leq i \leq a$ . Furthermore, we may take  $q_i < \xi$  for  $1 \leq i \leq a$ , by the method of proof of Lemma 10. Hence, by Brun's sieve, the number of integers  $n \leq x$  of the form

$$n = (p_1, q_1) \cdots (p_a, q_a) m.$$

with  $(m, \varphi(m)) = 1$ ,  $q_i \equiv 1 \pmod{p_i}$ ,  $1 \leq i \leq a$ , and no other relations in  $g(n)$ , is

$$(1 + o(1)) \frac{x e^{-y}}{(p_1 q_1) \cdots (p_a q_a) \log \log \log x} \exp\left(-l_2 \left(\frac{1}{p_1} + \cdots + \frac{1}{p_a}\right)\right)$$

as  $x \rightarrow \infty$ . We sum this over the distinguished pairs of primes  $(p_i, q_i)$  to obtain

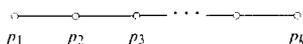
$$(1 + o(1)) \frac{xe^{-\gamma}(\log \log x)^a}{a! (\log \log \log x)} \times \left( \exp \left( -\log \log x \left( \frac{1}{p_1} + \dots + \frac{1}{p_a} \right) \right) / p_1^2 \cdots p_a^2 \right),$$

where the  $a!$  is to take into account the different orderings of the  $a$  prime pairs. In order to evaluate the asymptotic behaviour of the sum of this expression over the primes  $p_i$  in the interval  $(\log \log x)^{1-\epsilon} < p_i < (\log \log x)^{1+\epsilon}$ , we again use partial summation to write the sum as a product of  $a$  integrals. Each of the integrals is of the type considered in the case  $a = 1$ . Applying the same method to each in turn, we find that

$$F_{2^a}(x) = (1 + o(1)) \frac{c(a)x}{(\log \log \log x)^{a+1}}.$$

In fact, the above yields  $c(a) = e^{-\gamma}/a!$ , which holds for  $a \geq 0$ . This completes the proof of Theorem 3(ii).

We include here the following curious observation. Suppose the graph of  $n, g(n)$  has connected component



consisting of a chain of length  $k$ . We claim that  $G(n) = F_k$ , where  $F_k$  denotes the  $k$ th Fibonacci number. Indeed, if  $n = p_1 \cdots p_k$ , then

$$G(n) = \sum_{\substack{d|n \\ p_1 \nmid d}} \prod_{p|d} \left( \frac{p^{V_p(n/d)} - 1}{p - 1} \right) + \sum_{\substack{d|n \\ p_1 | d}} \prod_{p|d} \left( \frac{p^{V_p(n/d)} - 1}{p - 1} \right).$$

The second sum is  $G(n/p_1)$ , whereas the product in the first sum vanishes unless  $d | (n/p_2)$ . As  $p_1 | d$  in the first sum, we find that this sum is  $G(n/p_1 p_2)$ . An easy induction argument utilising  $G(n) = G(n/p_1) + G(n/p_1 p_2)$  now gives the result.

### 6. CONCLUDING REMARKS

If  $n$  is squarefree and  $G(n) = 3$ , then it is easy to see that  $n = pqrm$  where  $q \equiv 1 \pmod{p}$ ,  $r \equiv 1 \pmod{q}$ ,  $(m, \varphi(m)) = 1$  and no other relations hold. If

there are at least  $cx/(\log x)^2$  primes  $p \leq x$  such that  $2p + 1$  is also prime, then it is easy to see from the preceding discussion that for at least

$$\frac{cx}{(\log \log x)^{1-\varepsilon}}$$

numbers  $n \leq x$ , we have  $G(n) = 3$ . This should not be expected for all values of  $k$ .

Concerning the size of  $G(n)$  for squarefree  $n$ , we ask the following: is it true  $G(n) = o(\varphi(n))$  as  $n$  runs over squarefree integers? (see the Note added in proof). In this connection, it will be recalled that in [7], it was shown that if

$$f(n) = \prod_{p|n} (n, p - 1),$$

then

$$G(n) \leq f(n)$$

for all squarefree  $n$ . It is curious to note that  $G(n) = \varphi(n)$  can hold only for finitely many squarefree integers. Indeed, from the above, we find that for each  $p | n$ ,  $(p - 1)$  also divides  $n$  in such a case. We claim that  $n$  must be composed of 2, 3, 6, 43 only and a quick computation yields that  $n = 2, 6, 42, 1806$  are the only solutions. To see this, suppose that a prime  $p \neq 2, 3, 7, 43$  divides such an  $n$ . Then letting  $p$  be the least such prime, we find  $(p - 1) | n$ . But then  $p - 1$  must be composed of 2, 3, 7 or 43.

An immediate check of the corresponding squarefree products gives the result. This elegant elementary result appeared earlier (see Dyer-Bennet [1]) in a different context. It is likely that our question has an affirmative solution.

We have proved that for  $n$  squarefree,

$$\frac{\log G(n)}{\log \log n}$$

has a continuous distribution function. The above function has the same distribution as

$$\frac{\log f(n)}{\log \log n}.$$

It would be desirable to obtain nontrivial upper and lower bounds for

$$\sum_{n \leq x} f(n)$$

and

$$\sum_{n \leq x} \mu^2(n) G(n).$$

The second sum would be more difficult. Using the methods of [7] and [8] and Theorem 2, it follows that for any  $c > 0$ ,

$$x(\log x)^c \ll \sum_{n \leq x} \mu^2(n) G(n) \ll x^2 / \log \log \log x.$$

After this paper was written, Ram Murty and Srinivasan proved that

$$\sum_{n \leq x} \mu^2(n) G(n) \ll x^2 (\log x)^{-c \log \log \log x}$$

for some  $c > 0$ . Carl Pomerance informs us that for  $c > \frac{15}{23}$ , he can prove that

$$x^{1+c} \ll \sum_{n \leq x} \mu^2(n) G(n) \ll x^{2 - (\log \log \log x) / (\log \log x)}.$$

He can also give a heuristic argument to show that the upper bound is essentially best possible.

*Note added in proof.* The fact that  $G(n) = o(\varphi(n))$  has been subsequently proved by M. Ram Murty and S. Srinivasan. In a forthcoming paper entitled, "On the number of groups of squarefree order," they show that for square free  $n$ ,

$$G(n) = O(\varphi(n) / (\log n)^{A \log \log \log n})$$

for some constant  $A > 0$ .

#### REFERENCES

1. J. DYER-BENNET, A theorem on partitions of the set of positive integers, *Amer. Math. Monthly* **47** (1940), 152–154.
2. P. T. D. A. ELLIOTT, "Probabilistic Number Theory I, Mean Value Theorems," Springer-Verlag, New York, 1979.
3. P. ERDŐS, Some asymptotic formulas in number theory, *J. Indian Math. Soc.* **12** (1948), 75–78.
4. H. HALBERSTAM AND H. E. RICHERT, "Sieve Methods," Academic Press, London/New York 1974.
5. G. H. HARDY AND S. RAMANUJAN, The normal number of prime factors of a number  $n$ , *Quart. J. Math.* **48** (1920), 76–92.
6. O. HÖLDER, "Die gruppen mit quadratfreier ordnungszahl," pp. 211–229, *Nachr. Königl. Ges. Wiss. Göttingen Math.-Phys. Kl.*, 1895.
7. M. RAM MURTY AND V. KUMAR MURTY, On the number of groups of a given order, *J. Number Theory* **18** (1984) 178–191.
8. M. RAM MURTY AND V. KUMAR MURTY, On groups of squarefree order, *Math. Ann.* **267** (1984), 299–309.
9. P. NEUMANN, An enumeration theorem for finite groups, *Quart. J. Math. Oxford Ser. (2)* **20** (1969), 395–401.