On the Number of Sets of Cycle Lengths

Jacques Verstraëte

Department of Pure Mathematics and Mathematical Statistics,
Centre for Mathematical Sciences,
Wilberforce Road,
Cambridge CB3 OWB,
United Kingdom

E-mail:j.b.a.verstraete@dpmms.cam.ac.uk

Abstract

A set S of integers is called a *cycle set on* $\{1, 2, \ldots, n\}$ if there exists a graph G on n vertices such that the set of lengths of cycles in G is S. Erdős conjectured that the number of cycle sets on $\{1, 2, \ldots, n\}$ is $o(2^n)$. In this paper, we verify this conjecture by proving that there exists an absolute constant $c \geq 0.1$ such that the number of cycle sets on $\{1, 2, \ldots, n\}$ is $o(2^{n-n^c})$.

1 Introduction.

For any graph G, the set of lengths of cycles in G is denoted by $C(G)$. A set S of integers is called a *cycle set on* $\{1, 2, \ldots, n\}$ if there exists a graph G on n vertices such that $C(G) = S$. There are numerous results relating the edge density and arithmetic properties of cycle sets in graphs. Possibly the simplest statement is that if G is a graph of minimum degree at least k, then $|C(G)| \geq k - 1$. The residue properties of $C(G)$ are also well-researched: for example, Bondy and Vince [3] proved that if G is a graph of minimum degree at least three, then $C(G)$ contains a pair of consecutive integers or a pair of consecutive even integers. The following result of the author [15] partly

generalises this: any graph of average degree at least $8k$ contains cycles of k consecutive even lengths. Very general results pertaining to the density of the set of cycle lengths were established by Gyárfás, Komlós and Szemerédi [11]: they proved that there exists an absolute positive constant ε such that for every graph G of average degree at least d, the sum of reciprocals of elements of $C(G)$ is at least $\varepsilon \log d$. Although these results suggest that cycle sets on $\{1, 2, \ldots, n\}$ differ significantly from ordinary subsets of $\{1, 2, \ldots, n\}$, they do not imply the following conjecture of Erdős [5]:

Conjecture 1.1. *The number of cycle sets on* $\{1, 2, \ldots, n\}$ *is* $o(2^n)$.

Conjecture 1.1 has been studied for certain classes of graphs by Denley [4]. From the results of the author [13], it can be deduced that the number of cycle sets on $\{1, 2, \ldots, n\}$ arising from the class of graphs on n vertices and of average degree at least $9 \log_2 \log_2 n$ is $o(2^n)$. In this paper, we will verify Conjecture 1.1 by proving the following theorem:

Theorem 1.2. *There exists an absolute positive constant* c *such that the number of cycle sets on* $\{1, 2, \ldots, n\}$ *is* $o(2^{n-n^c})$.

A simple lower bound for the number of cycle sets on $\{1, 2, \ldots, n\}$ is given approximately by the partition function

$$p(n) \sim \frac{1}{4\sqrt{3}n} \mathrm{e}^{\pi\sqrt{2n/3}},$$

by considering the class of graphs on n vertices comprising vertex-disjoint cycles. A more effective example is as follows: let $A \subset \{3, 5, \ldots, 2n-3\}$, $C = v_0 v_1 \ldots v_{2n-1} v_0$ and define $G_{2n}^A = C + \{v_0 v_i : i+1 \in A\}$. Then $C(G_{2n}^A) \cap (2\mathbb{Z}+1) = A$. So the number of distinct sets $C(G_{2n}^A)$ is at least 2^{n-2}. A slightly better example was found by Faudree: let $A \subset \{n+1, \ldots, 2n-1\}$ and $G_{2n}^A = C + \{v_0 v_{i-1} : i \in A\}$. Then $C(G_A) \cap \{n, n+1, \ldots, 2n\} = A$ and therefore the number of distinct sets $C(G_A)$ is at least 2^{n-1}. Let $C(n)$ be the number of cycle sets on $\{1, 2, \ldots, n\}$. It would be interesting to determine the limit (if it exists)

$$\lim_{n\to\infty} n^{-1}\log_2 C(n).$$

The above examples show that the limit is at least $\frac{1}{2}$.

This paper is organised as follows: in section 2, we count the number of sets in $\{1,2,\ldots,n\}$ containing a certain strong additive structure. In section 3, we prove that all but a negligible number of graphs G_n contain subgraphs each of which guarantees that $C(G_n)$ contains this strong additive structure. The proof of Theorem 1.2 is then completed in section 4.

NOTATION. In this paper, G_n denotes a generic graph of order n, $V(G_n)$ its vertex set and $E(G_n)$ its edge set. A path of length k with vertex set $\{u_1,u_2,\ldots,u_{k+1}\}$ is represented by $u_1u_2\ldots u_{k+1}$. Similar notation is used for cycles. As we have mentioned above, $C(G)$ denotes the set of lengths of cycles in a graph G, and we write $C(n)$ for the number of cycle sets on $\{1,2,\ldots,n\}$. The notation $\mathbb{Z}_n$ is used for the integers modulo n and $\mathbb{N}$ denotes the set of natural numbers. If S is a set, then its complement is denoted S^c.

2 Counting Sums of Sets.

Let A,B be sets of integers and define the *positive difference set of A and B* by $(A-B)^+ = \{a-b : a\in A, b\in B, a>b\}$. The *Fourier transform* of a complex-valued function f on $\mathbb{Z}_{2n+1}$ is:

$$\hat{f}(x) = \sum_{r=0}^{2n} f(r)\,\exp(2\pi irx/(2n+1)).$$

For convenience, we replace $\exp(2\pi irx/(2n+1))$ by ω^{rx}. If $A\subset\mathbb{Z}_{2n+1}$, then $A(r)=1$ if $r\in A$ and $A(r)=0$ otherwise. References on set summation and properties of Fourier transforms can be found in Nathanson [14] and

Freiman [8]. We will count the number of subsets of $\{1, 2, \ldots, n\}$ containing large (in a sense made precise later) positive difference sets. If a subset of $\{1, 2, \ldots, n\}$ contains such a set, then its Fourier transform is large in modulus at some point in $\{1, 2, \ldots, n\}$. The probability that a random subset of $\{1, 2, \ldots, n\}$ has a large transform on some point is then estimated. The only non-elementary tool required is the following concentration result derived from Hoeffding's Inequality (see for example McDiarmid's formulation [13], Proposition 7.11, page 175):

Proposition 2.1. *Let $f : \{0,1\}^n \to \mathbb{R}$ be such that f changes by at most c when the ith co-ordinate changes. Let $Y = f(X_1, X_2, \ldots, X_n)$ where X_i are chosen independently and at random from $\{0,1\}$. Then $Pr[Y \geq E[Y] + t] \leq exp(-t^2/2c^2n)$.*

Lemma 2.2. *Suppose $A, S \subset \{1, 2, \ldots, n\}$ and $S \supset (A - A)^+$. Let $T = S^c \cup (-S^c)$ (mod $2n+1$). Then $\max_{r \neq 0} |\hat{T}(r)| \geq |S^c||A|/n$.*

Proof. As $t = a - a'$ has no solution with $t \in T^c$ and $a, a' \in A$, we have:

$$\sum_{r=0}^{2n} \sum_{t=0}^{2n} \sum_{a=0}^{2n} \sum_{a'=0}^{2n} A(a)A(a')T(t)\omega^{r(t-a+a')} = 0.$$

Therefore $\sum_{r=0}^{2n} \hat{T}(r)|\hat{A}(r)|^2 = 0$. For $r = 0$, the summand is $2|S^c||A|^2$. So if $m = \max_{r \neq 0} |\hat{T}(r)|$ then

$$\begin{aligned} 2|S^c||A|^2 &= \left| \sum_{r=1}^{2n} \hat{T}(r)|\hat{A}(r)|^2 \right| \\ &\leq m \sum_{r=1}^{2n} |\hat{A}(r)|^2 = m((2n+1)|A| - |A|^2). \end{aligned}$$

Since $|A| \geq 1$, we find $m \geq |S^c||A|/n$ as required. ∎

Using Proposition 2.1 and Lemma 2.2, it is relatively straightforward to count the number of sets containing large positive difference sets:

Lemma 2.3. *The number of subsets of* $\{1,2,\ldots,n\}$ *containing a positive difference set* $(A-A)^+$ *with* $A \subset \{1,\ldots,n\}$ *and* $|A| > a$ *is at most* $(2n+1)2^{n-a^2/18n}$.

Proof. Let S be a random set whose elements are chosen uniformly with probability $\frac{1}{2}$ from $\{1,2,\ldots,n\}$. Let $X_i = S(i)$ and $f(X_1,X_2,\ldots,X_n) = \sum X_i = |S|$. By Proposition 2.1, $\Pr[|S| \geq 2n/3] \leq \exp(-n/18)$. The transform of T in Lemma 2.2 is

$$\hat{T}(r) = \sum_{i=1}^{n}(1-X_i)(\omega^r + \omega^{-r}).$$

Thus $\hat{T}(r)$ is real, $\hat{T}(r) = \hat{T}(2n+1-r)$ and $\mathrm{E}[\hat{T}(r)] = 0$. By Lemma 2.2 and Proposition 2.1 applied to $f = \hat{T}(r)$, $\Pr[\hat{T}(r) \geq t] \leq \exp(-t^2/2n)$. By applying Proposition 2.1 to $1-\hat{T}(r)$, we obtain $\Pr[|\hat{T}(r)|] \leq 2\exp(-t^2/2n)$. Now define the events

$$\begin{aligned} E_1 &= \{\ |S^c| \leq n/3\ \} \\ E_2 &= \{\ \exists r \neq 0 : |\hat{T}(r)| \geq n/3\ \}. \end{aligned}$$

Then $\Pr[E_1] \leq \exp(-n/18)$ and $\Pr[E_2] \leq 2n\exp(-n/18)$. Consequently, the number of sets S in $\{1,2,\ldots,n\}$ containing $(A-A)^+$ for some $A \subset \{1,2,\ldots,n\}$ with $|A| \geq a$ is at most $(2n+1)2^{n-a^2/18n}$. ∎

This lemma is effective when $a/\sqrt{n} \to \infty$. With this in mind, we call a positive difference set $(A-A)^+$ in $\{1,2,\ldots,n\}$ $\eta-$*large* if $|A| \geq \sqrt{18}n^{(1+\eta)/2}$ and $\eta > 0$. By Lemma 2.3, the number of sets $S \subset \{1,2,\ldots,n\}$ containing an $\eta-$large positive difference set is $O(2^{n-n^\eta})$.

A more difficult problem, of independent interest, is to determine the asymptotic number of subsets of $\{1,2,\ldots,n\}$ or of $\mathbb{Z}_n$ of the form $A+A$ for some $A \subset \{1,2,\ldots,n\}$. Green and Ruzsa [10] solved this problem by showing that the number of subsets of $\{1,2,\ldots,n\}$ of the form $A+A$ is $\Theta(2^{n/2})$. They also showed that the number of subsets of $\mathbb{Z}_p$ of the form $A+A$ is $\Theta(p2^{p/3})$

when p is prime. This is related problem to the well-known Erdős-Cameron conjecture [8], that the number of sum-free subsets of $\{1, 2, \ldots, n\}$ is $O(2^{n/2})$. This problem was studied by many authors, and was finally verified by Green [9].

Let A be a multiset with elements $a_1, a_2, \ldots, a_n$. We define the *set of subset sums* of A to be the set

$$A^* = \{\varepsilon_1 a_1 + \ldots + \varepsilon_n a_n : \varepsilon_i \in \{0, 1\}, 1 \leq i \leq n\}.$$

A beautiful theorem of Erdős and Sárközy [7] states that if A and A^* are subsets of $\{1, 2, \ldots, n\}$ and $|A| \geq 18(\log_2 n)^2$, then A^* contains an arithmetic progression of length $|A|/18(\log_2 n)^2$. We outline their proof, which also works for multisets:

Proposition 2.4. *Suppose A is a multiset with elements $a_1, a_2, \ldots, a_t \in \{1, 2, \ldots, n\}$ and $t \geq 18(\log_2 n)^2$. Then A^* contains at least $t/18(\log_2 n)^2$ consecutive multiples of some positive integer.*

Proof. Set $q = \lfloor \log_2 n \rfloor$ and $p = \lfloor t/18(\log_2 n)^2 \rfloor$. By the pigeonhole principle, there exists a number m such that at least

$$\frac{1}{qn}\binom{t}{q} > q! p^{q+1}$$

sets of q elements of A sum to m. If $\mathcal{A}$ is the set system comprising those sets $I \subset \{1, 2, \ldots, t\}$ of size q such that $\sum_{i \in I} a_i = m$, then, by a result of Erdős and Rado [6], there exists a set system $\mathcal{B} \subset \mathcal{A}$ such that $|\mathcal{B}| > p$, and all intersections of distinct sets in $\mathcal{B}$ are equal. Let B be the common intersection of the elements of $\mathcal{B}$. Then more than p consecutive multiples of $m - \sum_{a_i \in B} a_i$ appear in A^*. ∎

We will use the following terminology: a set of subset sums $A^* \subset \{1, 2, \ldots, n\}$ is called $\eta-$*large* if $|A| \geq 20n^\eta(\log_2 n)^2$ and $\eta > 0$. The following lemma shows that few sets contain a translate of an $\eta-$large set of subset sums:

Lemma 2.5. *The number of sets $S \subset \{1, 2, \ldots, n\}$ containing a translate of an $\eta-$large set of subset sums is $O(2^{n-n^\eta})$.*

Proof. If A^* is an $\eta-$large set of subset sums, then by Proposition 2.4, A^* contains at least $\frac{10}{9}n^\eta$ consecutive multiples of some positive integer. In any translate of A^*, this set corresponds to an arithmetic progression. There are at most n^2 arithmetic progressions in $\{1, 2, .., n\}$, as there are at most n choices for the first term and n choices for the common difference. Let S be a random set whose elements are selected uniformly from $\{1, 2, \ldots, n\}$ with probability $\frac{1}{2}$. Then the probability that S contains a given arithmetic progression of length k is 2^{-k}. So the number of sets containing such an arithmetic progression is at most $n^2 2^{-k} 2^n$. With $k = \frac{10}{9}n^\eta$, the proof is complete. ∎

Another problem of independent interest is counting the number of subsets of $\{1, 2, \ldots, n\}$ of the form A^* where A is a multiset in $\{1, 2, \ldots, n\}$. Szemerédi and Vu [14] recently solved this problem and showed that the number of sets A^* is $2^{O(\sqrt{n})}$.

3 Cycles in Hamiltonian Graphs

In order to prove Theorem 1.2, we require a number of preliminary lemmas, which concern cycle structure in hamiltonian graphs. In the proof of Theorem 1.2, it will be seen that the assumption of hamiltonicity may be made. In order to apply the results of section 2, we will show that all but a negligible number of hamiltonian graphs G_n contain subgraphs which guarantee that $C(G_n)$ contains an $\eta-$large positive difference set or a translate of an $\eta-$large set of subset sums.

Throughout this section, we will assume that a hamiltonian graph G with a given hamiltonian cycle C is drawn in the plane in such a way that all

the edges of G are straight line segments and the edges of C form a convex polygon.

Let G be a graph with a hamiltonian cycle C. Suppose G is a plane graph, C has k chords, and the bounded regions of G are $F_1, F_2, \ldots, F_{k+1}$. Then G is called a *k-ladder*, denoted $H(k)$, if the boundaries of F_i and F_{i+1} share precisely one chord of C for $i = 1, 2, \ldots, k$. The following lemma shows that such graphs guarantee a large positive difference set of cycle lengths:

Lemma 3.1. *If G contains a k-ladder, then $C(G)$ contains a translate of a positive difference set $(A - A)^+$ with $|A| = k + 1$.*

Proof. Suppose $H(k) \subset G$ and let $F_1, F_2, \ldots, F_{k+1}$ denote the bounded regions of $H(k)$ such that the boundaries of F_i and the F_{i+1} share precisely one edge for $i = 1, 2, \ldots, k$. Let a_i denote the number of edges on the boundary of F_i and let

$$A = \left\{ \sum_{i=1}^{j} (a_i - 2) : 1 \leq j \leq k+1 \right\} \cup \{0\}$$

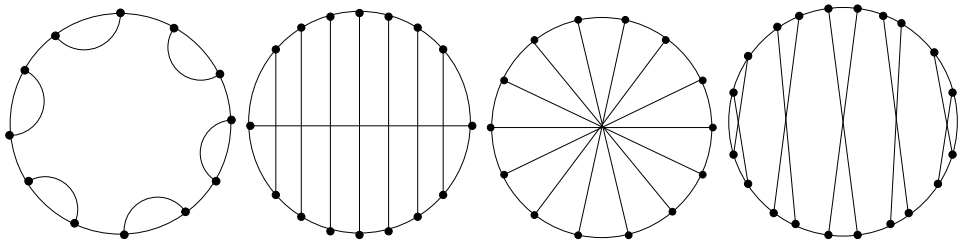
Then $C(G) \supset C(H(k)) \supset 2 + (A - A)^+$. ∎

If G_n is a hamiltonian graph containing a vertex of degree $d \geq \sqrt{18} n^{(1+\eta)/2} \geq 2$, then $G_n \supset H(d-2)$. Therefore, by Lemma 3.1, $C(G_n)$ contains an $\eta-$large positive difference set. This fact will be used later.

We now look at graphs which guarantee a translate of an $\eta-$large set of subset sums of cycle lengths. Let k be a positive integer and let C_m denote a cycle of length m labelled $v_0 v_1 \ldots v_{m-1} v_0$. Define the following graphs:

$$\begin{aligned}
J_k &= C_{3k} \cup \{v_{3i} v_{3i+2} : 0 \leq i \leq k-1\} \\
L_k &= C_{8k+2} \cup \{v_0 v_{4k+1}\} \cup \{v_i v_{8k+2-i} : 1 \leq i \leq 4k\} \\
M_k &= C_{8k+2} \cup \{v_i v_{i+4k+1} : 0 \leq i \leq 4k\} \\
N_k &= C_{4k} \cup \{v_{2i} v_{4k-2i-2}, v_{2i+1} v_{4k-2i-1} : 0 \leq i \leq k-1\}.
\end{aligned}$$

Throughout this section, we will suppose that the vertices of J_k, L_k, M_k and N_k are labelled as in their definitions above. Label the hamiltonian cycle C_{3k}, C_{8k+2} or C_{4k}, respectively, C'. Natural drawings of J_k, L_k, M_k and N_k, with C' as a convex polygon, are illustrated below:



J_k L_k M_k N_k

A graph G obtained by subdividing edges of C' in J_k, L_k,M_k and N_k is is denoted $J(k), L(k), M(k)$ or $N(k)$, respectively.

Let G be a hamiltonian graph with hamiltonian cycle C. Chords e and f of C are *crossing* if C, together with e and f, is not a plane graph. A pair of crossing chords e and f of C are said to cross *trivially* if C, together with e and f, contains a 4-cycle through e and f. Otherwise e and f are said to cross *non-trivially*. If $N(k)$ contains no trivially crossing chords, then we relabel it $N^*(k)$. A *type k graph* is any of the graphs $J(k), L(k), M(k)$ or $N^*(k)$. The proof of the following lemma shows a reason for the additional requirement on $N(k)$:

Lemma 3.2. *Let G be a type k graph. Then $C(G) \supset a + A^*$, where A is a multiset in $\{1, 2, \ldots, n\}$ of size at least k and a is a positive integer.*

Proof. Suppose $G = J(k)$. Let P_i be a v_{3i}-v_{3i+2} subpath of C for $i = 1, 2, \ldots, k$ and suppose P_i and P_j are vertex-disjoint whenever $i \neq j$. Let $a_i = |P_i| - 1$ and set $a = |E(G)| - \sum_{i=1}^{k}(a_i - 1)$. Then, for any $S \subset \{1, 2, \ldots, k\}$, we may find a cycle of length $a + \sum_{i \in S}(a_i - 1)$ in G. So $C(G) \supset a + A^*$ where A is the multiset with elements $a_1 - 1, a_2 - 1, \ldots, a_k - 1$.

Suppose $G = L(k)$. Let P_i and Q_i be a pair of vertex-disjoint v_i-v_{i+1} and v_{8k+1-i}-v_{8k-i} paths, for $i = 1, 2, \ldots, 4k-1$. Let $x_i = |P_i| - 1$ and $y_i = |Q_i| - 1$ and $I = \{1, 3, \ldots, 4k-1\}$. Then, for any $S \subset I$, we may find a cycle in G through $v_0 v_{4k+1}$ which excludes the edges in $\bigcup_{i \in S} P_i$ and includes the edges in $\bigcup_{i \in S} Q_i$. Let a and b be the number of vertices in the two edge-disjoint v_0-v_{4k+1} subpaths of C. Then, for any $S \subset I$,

$$C(G) \supset a + \sum_{i \in S} (y_i + 2 - x_i).$$

Similarly, $C(G) \supset b + \sum_{i \in S} (x_i + 2 - y_i)$. For each $i \in I$, it is not possible that neither summand is positive. Without loss of generality, there exists $I' \subset I : |I'| \geq |I|/2 = k$ such that $a_i = x_i + 2 - y_i > 0$ for $i \in I'$. With $A = \{a_i : i \in I'\}$, we have $a + A^* \subset C(G)$ and $|A| \geq k$.

Suppose $G = M(k)$. Let P_i be a v_i-v_{i+1} subpath of C, chosen so that P_i and P_{i+1} are edge-disjoint for $i = 0, \ldots, 8k$. Let $x_i = |P_i| - 1$ and $y_i = |P_{i+4k+1}| - 1$ for $i = 0, 1, \ldots, 4k$. Let $I = \{1, 3, \ldots, 4k-1\}$ and $S \subset I$. Then

$$C(G) \supset |P_0| + |P_{4k}| - 1 + \sum_{i \in S} (y_i + 2 - x_i)$$

and similarly $C(G) \supset |P_{4k}| + |P_{8k+1}| - 1 + \sum_{i \in S} (x_i + 2 - y_i)$. For $i \in I$, it is not possible that neither summand is positive. Without loss of generality, there is a set $I' \subset I : |I'| \geq |I|/2 = k$ such that $a_i = x_i + 2 - y_i > 0$ for $i \in I'$. With $a = |P_0| + |P_{4k}| - 1$ and $A = \{a_i : i \in I'\}$, $a + A^* \subset C(G)$ and $|A| \geq k$.

If $G = N^*(k)$, let P_i and Q_i denote the pair of vertex-disjoint v_{2i-2}-v_{2i-1} and v_{4k-2i}-$v_{4k-2i+1}$ subpaths of C, respectively, for $i = 1, \ldots, k$. Let $x_i = |P_i| - 1$ and $y_i = |Q_i| - 1$ and $a = |C| - \sum_{i=1}^{k} (x_i + y_i)$. Then, for any $S \subset \{1, 2, \ldots, k\} = I$,

$$C(G) \supset a + \sum_{i \in S} (x_i + y_i - 2).$$

As no chords cross trivially, $a_i = x_i + y_i - 2 > 0$ for $i \in I$. So $C(G) \supset a + A^*$ where A has elements $a_1, a_2, \ldots, a_k$. ∎

It follows that if G_n contains a type k-graph and $k \geq 20n^{\eta}(\log_2 n)^2$, then $C(G_n)$ contains a translate of an $\eta-$large set of subset sums. The following lemma, which will be used to find such structure in hamiltonian graphs, generalizes a result due to Aldred and Thomassen [1], who proved it for $d = 1$:

Lemma 3.3. *Let k be a positive even integer and let G be a graph of maximum degree at most $d+2$, containing a hamiltonian cycle C. Suppose C has at least $(4k)^4 d$ chords and each chord of C is non-trivially crossed by some other chord of C. Then G contains a type k-graph.*

Proof. Since any graph with m edges and maximum degree at most d contains a set of at least $m/2d$ independent edges, there is a set I of $128k^4$ independent chords of C. Partition C into k subpaths, each containing $256k^3$ end vertices of the edges in I. If for every subpath there is a chord of G with both ends in the subpath, then $J(k) \subset G$. Otherwise, choose a subpath P, containing no edge in I with both ends in P, and let the edges of I meeting P be $e_1, e_2, \ldots, e_{256k^3}$, numbered according to the order in which their ends in P appear in some orientation of C. In the same orientation of C, these edges are re-ordered $e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(256k^3)}$ according to the order of appearance of their ends on $C - V(P)$. If there exists an increasing sequence of $\sigma(i)$ of length at least $8k+2$, then $M(k) \subset G$. Otherwise, by an old result of Erdős and Szekeres, there exists a decreasing subsequence of numbers $\sigma(i)$ of length at least $256k^3/(8k+2) \geq (2k-1)(16k+4)$. Let $f(i) = u_i v_i$ denote the edges $e_{\sigma(i)} : 1 \leq i \leq (2k-1)(16k+4)$.

For $i = 0, 1, \ldots, 2k-2$, let

$$\begin{aligned} E_i &= \{f((16k+4)i + j) : 1 \leq j \leq (16k+4)\} \text{ and} \\ V_i &= \{u_j, v_j : f_j \in E_i\}. \end{aligned}$$

Since every chord of C is non-trivially crossed by some other chord of C, for each $i : 0 \leq i \leq 2k-2$, $f((8k+2)(2i+1)) = e(i)$ is non-trivially crossed by some chord $e'(i)$ of C. If some end of $e'(i)$ is not in V_i, then

$L(k) \subset G$. So we suppose the ends of $e'(i)$ are in V_i for $i = 0, 1, 2, \ldots, 2k-2$. If both ends of $e'(i)$ are not in P for at least k values of i, then $J(k) \subset G$. Therefore we suppose $e'(i)$ has one end in P and one end in $V_i \backslash V(P)$ for $i \in I \subset \{0, 1, 2, \ldots, 2k-2\}$ with $|I| \geq k$. The edges $e(i)$ and $e'(i)$ for $i \in I$, together with C, now comprise $N^*(k)$ in G. ∎

The next lemma deals with hamiltonian graphs do not necessarily satisfy the hypotheses of Lemma 3.3:

Lemma 3.4. *Let k be a positive integer and let G be a graph of maximum degree at most $d+2$. Suppose G contains a hamiltonian cycle C with m chords. Then either*

(1) *G contains a type k graph or*
(2) *G contains a plane graph containing C and at least $\frac{1}{2}(m-(4k)^4)$ chords of C.*

Proof. Colour a chord of C blue if some chord of C non-trivially crosses it, and red otherwise. Then the set of blue edges and the edges of C span a subgraph of G satisfying the hypotheses of Lemma 3.3. We therefore assume that G contains at most $(4k)^4 d$ blue edges and at least $m-(4k)^4 d$ red edges. The set of red edges, together with the edges of C, span a graph in which any crossing edges cross trivially. So we may remove a set of at most half the red chords of C to obtain a plane subgraph of G satisfying (2). ∎

The following elementary lemma will be used in the next corollary:

Lemma 3.5. *Let T be a tree. Then T contains at least s leaves or a path of length at least $|T|/2s-2$.*

Proof. If T contains at most s leaves, then T contains at least $|T|-2s$ vertices of degree two. This implies that some set of at least $|T|/2s-1$ vertices of degree two in T induce a path in T. ∎

Corollary 3.6. *Let G satisfy the hypotheses Lemma 3.3. Then G contains a type k graph or $G \supset H(r)$ where $r \geq \max\{\frac{1}{2k}(m-(4k)^4 d)-2, d\}$.*

Proof. If G does not contain a type k graph then, by (2) in Lemma 3.4, G contains a plane graph F containing C and at least $\frac{1}{2}(m-(4k)^4 d)$ chords of C. By Euler's formula, F has at least $\frac{1}{2}(m-(4k)^4 d)$ bounded faces. The vertices of the plane dual graph of F corresponding to these bounded faces induce a tree T. By Lemma 3.5, T contains at least k leaves or a path of length at least $\frac{1}{2k}(m-(4k)^4 d)-2$. If the former case, $J(k) \subset F$ and in the latter case, C, together with the edges on the boundaries of faces in F corresponding to vertices of P, forms an $H(r)$, where $r \geq \frac{1}{2k}(m-(4k)^4 d)-2$. By Lemma 3.1, G also contains an $H(d)$. This completes the proof. ∎

This corollary is fundamental in proving Theorem 1.2.

4 Proof of Theorem 1.2

Using the results of section 2 and 3, we prove Theorem 1.2:

Proof of Theorem 1.2. If $\mathcal{H}$ is any class of graphs, we write $C_{\mathcal{H}}(n)$ for the number of cycle sets on $\{1,2,\ldots,n\}$ arising from graphs on n vertices in $\mathcal{H}$. When $\mathcal{H}$ is the class of all graphs, we write $C(n)$ instead of $C_{\mathcal{H}}(n)$. If $\mathcal{H}_1$ is the class of graphs of circumference less than $n-n^{1/10}$, then clearly $\log_2 C_{\mathcal{H}_1}(n) < n-n^{1/10}$. Let $\mathcal{H}_2$ be the class of graphs of circumference at least $n-n^{1/10}$ containing less than $n+n/(4\log_2 n)$ edges. Let G_n be a graph of order n in $\mathcal{H}_2$. Fixing a cycle of length at least $n-n^{1/10}$ in G_n, the number of ways of placing the remaining edges is at most

$$\binom{n}{2}^{n/(4\log_2 n)+n^{1/10}} = 2^{n/2+o(n)}.$$

This is an upper bound for $C_{\mathcal{H}_2}(n)$. Now let $\mathcal{H}_3$ denote the class of graphs not in $\mathcal{H}_1$ or $\mathcal{H}_2$. We will show that if $G_n \in \mathcal{H}_3$, then G_n contains a $\frac{1}{10}$–large positive difference set or a translate of a $\frac{1}{10}$–large set of subset sums.

Let C be a longest cycle in a graph $G_n \in \mathcal{H}_3$. If some vertex $v \in V(G_n)\backslash V(C)$ has at least $\sqrt{18}n^{2/3}+2$ neighbours on C, then we easily find a cycle C' in G_n containing v and all the neighbours of v. By Lemma 3.1, this implies $C(G_n)$ contains a $\frac{1}{3}-$large positive difference set. If no vertex of $G_n - V(C)$ has at least $\sqrt{18}n^{2/3}+2$ neighbours on C, then C has at least

$$n + n/(4\log_2 n) - n^{1/10}(\sqrt{18}n^{2/3}+1) - n^{1/5} = n + \Omega(n/\log_2 n)$$

chords. By Corollary 3.6, with $k = 20n^{2/19}(\log_2 n)^2$ and $d = \sqrt{18}n^{1/2+1/19}+2$, H contains a type k-graph or $H(r)$, where $r = \max\{d-2, \frac{1}{2k}(\Omega(n/\log_2 n) - (4k)^4 d)\}$. By the choice of d and k, this maximum is $d-2 = n^{1/2+1/19}$, so G contains $H(d-2)$. By Lemma 3.1 and Lemma 3.2, this implies that $C(H)$ contains a $\frac{2}{19}-$large positive difference set or a translate of a $\frac{2}{19}-$large set of subset sums. By Lemma 2.3 and Lemma 2.5, the number of subsets of $\{1,2,\ldots,n\}$ containing a $\frac{2}{19}-$large positive difference set or a translate of a $\frac{2}{19}-$large set of subset sums is $O(2^{n-n^{2/19}})$. The proof of Theorem 1.2 is now complete with $c \geq \frac{1}{10}$. ∎

Acknowledgements.

I would like to thank Andrew Thomason and Ben Green for helpful discussions.

References.

[1] Aldred, R, Thomassen, C., On the number of cycles in 3-connected cubic graphs, *J. Combin. Theory Ser. B* **71** (1997), no. 1, 79–84.

[2] Alon, N., Spencer, J., The probabilistic method, Wiley-Interscience (1992).

[3] Bondy, J., Vince, A., Cycles in a graph whose lengths differ by one or two, *J. Graph Theory* **27** (1998) 11–15.

[4] Denley, T., Arithmetic progressions of cycles in outer-planar graphs. Combinatorics, graph theory and computing (Louisville, KY, 1999). Discrete Math. 249 (2002).

[5] Erdős, P., Conference communication, Memphis, TN, 1996.

[6] Erdős, P., Rado, R., Intersection theorems for finite sets, (1960).

[7] Erdős, P., Sárközy, A., Arithmetic progressions in subset sums, *Discrete Math.* **102** (1992), no. 3, 249–264.

[8] Freiman, G., Structure theory of set addition, in: *Astérisque* **258** (1999), 1–33.

[9] Green, B., The Cameron-Erdős conjecture. To appear in Bull. London Math. Soc

[10] Green, B., Ruzsa, I., Counting sumsets and sum-free sets modulo a prime. To appear in Studia Sci. Math. Hungarica.

[11] Gyárfas, A., Komlós, J., Szemerédi, E., On the distribution of cycle lengths in graphs, *J. Graph Theory* **8** (1984) 441–466.

[12] McDiarmid, C., On the method of bounded differences, Surveys in Combinatorics, LMS **141** (1980) 148–188.

[13] Nathanson, M., Additive number theory. Inverse problems and the geometry of sumsets, Graduate texts in mathematics, **165**, Springer-Verlag (1996).

[14] Szemerédi, E., Vu, V., Long arithmetic progressions in sumsets and the number of x-free sets. Submitted.

[15] Verstraëte, J., Arithmetic progressions of cycle lengths in graphs, *Combin. Probab. Comput.* **9** (2000) no 4. 369–373.