Publ. Math. Debrecen
79/3-4 (2011), 675–685
DOI: 10.5486/PMD.2011.5135

The difference graph of S-units

By IMRE Z. RUZSA (Budapest)

To my esteemed friends Kálmán, András, János, Attila

Abstract. For a given finite set S of primes we construct a graph on $\mathbb{Z}$ by connecting two integers if their difference is an S-unit. We investigate finite subgraphs of this graph. We show that the average degree is always small. We also meditate on the possible cycles that arise as induced subgraphs.

1. Introduction

In the Schweitzer competition of 2009, the following problem was proposed by Kálmán Győry:

Let S be a finite set of primes. We make a graph on $\mathbb{Z}$ by connecting m, n if $m - n$ is an S-*unit*, that is, an integer composed exclusively of primes $p \in S$.

Find the possible sizes of

(a) complete subgraphs,

(b) induced paths.

Though this problem was an easy one, it points towards two basic directions in understanding the structure of the S-unit graph defined above: to show that certain subgraphs are impossible, and to find some subgraphs that do exist. We make some further steps in both directions. We find the possible lengths of

Mathematics Subject Classification: 11B50, 11B75, 11P70.
Key words and phrases: S-unit, difference graph, average degree.
Author was supported by ERC–AdG Grant No. 228005 and Hungarian National Foundation for Scientific Research (OTKA), Grants No. 81658, 72731.

induced cycles, and we show that in a subgraph on n points the average degree is $O(n^{\varepsilon})$ for every $\varepsilon > 0$.

While my direct motivation came from the competition problem, it should be remarked that these graphs were introduced by Győry earlier and in a more general form, see [3], [4] and references therein, for a general finitely generated integral domain R over $\mathbb{Z}$ and more general multiplicative subgroups of R in the place of S-units. He described the component structure of the complement of our graph. Among others, he showed that these more general graphs do not contain large cliques either. He (and others) applied these results and this graph method to certain diophantine problems.

2. Preparation: equations in S-units

We fix our concepts and notations as follows. Given a finite nonempty set S of primes, $U \subset \mathbb{Q}$ will be the set of S-units, that is, those rational numbers whose numerator and denominator is composed only of primes $p \in S$. We make a graph $\mathcal{G}$ on $\mathbb{Q}$ by connecting x, y if $x - y \in U$.

Since we are interested only in finite subgraphs, we could restrict our attention to integers; indeed, given a finite collection of rational numbers, we can multiply them by a common denominator denominator to get a set of integers which induces an isomorphic graph. Sometimes it will be more convenient to work with rational numbers.

We shall consider the equation

$$x_1 + x_2 + \cdots + x_k = a,\ x_i \in U, \tag{2.1}$$

for $a \in \mathbb{Q}$. We call a solution $(x_1, \ldots, x_k)$ *degenerate*, if there is a nonempty proper zero subsum

$$x_{i_1} + \cdots + x_{i_m} = 0, \quad 1 \le i_1 < \cdots < i_m \le k,\ 0 < m < k,$$

and *nondegenerate* otherwise.

There are many results on equations in S-units. For us the following form will be most convenient.

Lemma 2.1. *The number of nondegenerate solutions of equation* (2.1) *is bounded by a number $C(k, S)$ for all $a \neq 0$, independent of the number a on the right hand side.*

This is essentially a deep theorem of Evertse[1] and van der Poorten and Schlickewei [5]. More general results and explicit bounds are available; we shall stay with this simple form.

In the above papers and other works dealing with this subject often superficially different versions are stated. It may be useful to clarify the connection between them. These versions sound as follows.

Version 1. The number of nondegenerate solutions of the equation

$$x_1 + x_2 + \dots + x_k = 1, \quad x_i \in U \tag{2.2}$$

is finite.

We denote this number by $C_1(k, S)$.

Version 0. The nondegenerate solutions of the equation

$$x_1 + x_2 + \dots + x_k = 0, \quad x_i \in U \tag{2.3}$$

lie in finitely many homothety classes. (The class of a solution $(x_1, \dots, x_k)$ contains the solutions $(tx_1, \dots, tx_k)$ with $t \neq 0$.)

We denote the number of these classes by $C_0(k, S)$.

Statement 2.2. *Version 1, Version 0 and Lemma 2.1 are equivalent.*

Proof. A possible way to select a representative from a homothety class is to fix $x_k = -1$. In this way equation (2.3) becomes

$$x_1 + x_2 + \dots + x_{k-1} = 1. \tag{2.4}$$

A solution of (2.3) is nondegenerate exactly if the corresponding solution of (2.4) is nondegenerate. This shows the equivalence of Versions 1 and 0 with

$$C_0(k, S) = C_1(k - 1, S).$$

Version 1 is the case $a = 1$ of Lemma 2.1, so an implication between them and the inequality $C_1(k, S) \le C(k, S)$ is clear. Now we deduce the Lemma from Version 0.

To this end fix a nondegenerate solution of (2.1), say $x_i = b_i$. The equation can be rewritten as

$$x_1 + \dots + x_k = b_1 + \dots + b_k.$$

This is equivalent to the equation

$$x_1 + x_2 + \dots + x_k + x_{k+1} + \dots + x_{2k} = 0, \quad x_i \in U \tag{2.5}$$

under the restrictions $x_{k+j} = -b_j$ and the assumption that no subsum of $x_1, \dots, x_k$ vanishes.

Such a solution may be degenerate. Let $I_1, \dots, I_m$ be a partition of the set of subscripts $\{1, \dots, 2k\}$ into disjoint nonempty sets such that

$$\sum_{i \in I_j} x_i = 0 \tag{2.6}$$

for all j, with the maximal possible value of m. We show that each fixed partition induces only a finite number of solutions.

The maximality of m implies that the induced solutions of (2.6) are nondegenerate, so they lie in $C_0(|I_j|, S)$ homothety classes. As $(x_1, \dots, x_k)$ has no zero subsum, each I_j must intersect $\{k+1, \dots, 2k\}$. If $r \in I_j$, $r > k$, then the fixed value of $x_r = b_{r-k}$ admits at most one solution from each class. This gives at most

$$\prod_{j=1}^{m} C_0(|I_j|, S)$$

solutions for a given partition, and summing this over all partitions we obtain a bound for $C(k, S)$. $\square$

3. Cycles

As above, S is a fixed finite set of primes, $U \subset \mathbb{Q}$ is the set of S-units and $\mathcal{G}$ is the graph on $\mathbb{Q}$ obtained by connecting x, y if $x - y \in U$.

Theorem 3.1. *If $2 \in S$, then there are cycles of every length among induced subgraphs of $\mathcal{G}$.*

If $2 \notin S$, then there are cycles of every even length (and only even, as the graph is bipartite).

Proof. We will work with integers. Observe first that always there is a cycle of length 4, say

$$0 \to 1 \to u + 1 \to u \to 0,$$

with an arbitrary $u \in U$. This will be an induced cycle if $u - 1 \notin U$; such values of u exist simply because both U and its complement are infinite. If $2 \in S$, we can also find a cycle of length 3, namely

$$0 \to 1 \to 2 \to 0.$$

Next we show how to make a cycle longer by 2. Let the cycle be

$$a_1 \to a_2 \to \cdots \to a_n \to a_1.$$

Take an $u \in U$, and shift a segment by u; the new cycle will be

$$a_1 \to a_1 + u \to a_2 + u \to \cdots \to a_k + u \to a_k \to \cdots \to a_n \to a_1.$$

To avoid coincidences we need that $u \neq a_i - a_j$, and to ensure that it is an induced subgraph we need that $u + a_i - a_j \notin U$ for all $i \neq j$. By Lemma 2.1 all but finitely many $u \in U$ are good (in fact, this follows already from a much older result of Pólya). I remark that for our aims the weaker fact that there are good values of u is sufficient, and this can be shown elementarily, using nothing more advanced than the Fermat–Euler congruence theorem. □

Given a cycle

$$a_1 \to a_2 \to \cdots \to a_n \to a_1.$$

the numbers "sitting" on the edges, $u_i = a_{i+1} - a_i$ for $i = 1, \dots, n-1$ and $u_n = a_1 - a_n$ satisfy

$$u_1 + u_2 + \cdots + u_n = 0.$$

We say that the cycle is *nondegenerate,* if the above representation is nondegenerate. The description of nondegenerate cycles seems to be less obvious.

Theorem 3.2. *If $2 \in S$, then there are nondegenerate cycles of every length among induced subgraphs of $\mathcal{G}$.*

If $2 \notin S$ but $3 \in S$, then there are nondegenerate cycles of every even length (and only even, as the graph is bipartite).

Theorem 3.2 follows from the following result.

Theorem 3.3. *If $2 \in S$, then for every k there are S-units $u_1, \dots, u_k > 0$ such that*

$$u_1 + \cdots + u_k = 1 \tag{3.1}$$

and

$$u_i + u_{i+1} + \cdots + u_j \notin U, \quad 1 \le i < j \le k,\ (i,j) \neq (1,k). \tag{3.2}$$

If $2 \notin S$ but $3 \in S$, then for every odd k there are S-units $u_1, \dots, u_k > 0$ satisfying (3.1) *and* (3.2)*.*

Proof. There is at least one k-tuple satisfying (3.1) with positive elements:

$$1 = 1/2 + 1/4 + \cdots + 1/2^{k-1} + 1/2^{k-1}$$

for $2 \in S$, k arbitrary, and

$$1 = 1/3 + 1/3 + 1/9 + 1/9 \cdots + 1/3^{(k-1)/2} + 1/3^{(k-1)/2} + 1/3^{(k-1)/2}$$

for $2 \notin S$, $3 \in S$, k odd.

Arrange each k-tuple decreasingly:

$$u_1 \geq u_2 \geq \dots \geq u_k > 0.$$

From the k-tuples select the lexicographically last (maximal u_1; for this u_1, maximal u_2, ...). Such a last one exists, since the total number of these tuples is finite by Lemma 2.1 (now best used as Variant 1).

This may not satisfy our requirement (3.2), but it has the following related property: if a subsum is in U, say

$$v = \sum_{i \in I} u_i \in U, \quad |I| \geq 2,$$

then necessarily $I = \{j, j+1, \dots, k\}$, a final segment.

Indeed, if j is the minimal element of I, and there is an $m \notin I$, $j < m \leq k$, then we can make a lexicographically later representation by the following transformations. First replace u_j by v and delete all $u_i, i \in I$; this is a later representation since we increased u_j, but a shorter one, having only $k - |I| + 1$ summands. Next we make it longer, via replacing u_m by $u_m/2, u_m/4, \dots$, if $2 \in S$, and by $u_m/3, u_m/3, u_m/9, u_m/9, \dots$, if $2 \notin S$, $3 \in S$. In this way we obtain a new representation of the same length which is lexicographically later, contradiction.

To ensure property (3.2) we need only to rearrange this sequence so as to separate u_k from u_{k-1}. A possible new sequence is $u'_1 = u_2$, $u'_2 = u_3$, ..., $u'_{k-1} = u_k$, $u'_k = u_1$. □

If the smallest element of S is at least 5, then the above simple arguments do not work.

Conjecture 3.4. If $2 \notin S$, then there are nondegenerate induced cycles of every sufficiently large even length.

Definitely a proof of this conjecture cannot proceed as that of Theorem 3.2. Indeed, write

$$m = \gcd\{p - 1 : p \in S\}. \tag{3.3}$$

By considering congruences modulo m we see immediately that a representation of the form (3.1) is possible only if $m|k$.

Conjecture 3.5. Assume $2 \notin S$ and $3 \notin S$, and define m by (3.3). For every sufficiently large k such that $m|k$ there are S-units $u_1, \dots, u_k > 0$ such that $u_1 + \dots + u_k = 1$ and

$$u_i + u_{i+1} + \dots + u_j \notin U, \quad 1 \leq i < j \leq k, \ (i, j) \neq (1, k).$$

Problem 3.6. In the above arguments the main difficulty arose because we wanted our cycles to be induced subgraphs. Is this necessary? Is there a finite subgraph of $\mathcal{G}$ such that there is no isomorphic induced subgraph?

I expect a positive answer.

The next step beyond cycles would be graphs where the degree of each vertex is at most 3.

Problem 3.7. Assume $2 \in S$. Is there a constant k with the property that every finite graph with maximal degree 3 and girth $> k$ occurs among the subgraphs of $\mathcal{G}$?

4. Upper estimate for the average degree

It is easy to see, and was part of the Schweitzer problem mentioned in the Introduction, that a complete subgraph of $\mathcal{G}$ cannot have more than p elements, where p is the smallest prime outside S. We show that large subgraphs will be very far from complete.

Theorem 4.1. *For any $\varepsilon > 0$, any subgraph of $\mathcal{G}$ on n vertices has average degree $< c_{\varepsilon,S} n^{\varepsilon}$.*

Proof. Let D_n be the maximum of the average degrees of n-point subgraphs of $\mathcal{G}$, and let d_n be the maximum of the *minimal* degrees of n-point subgraphs of $\mathcal{G}$. We first estimate d_n, and use this estimate to find a bound for D_n. This will be achieved by giving a lower and upper estimate for the number of certain paths in such graphs.

Let $\mathcal{H}$ be a subgraph of $\mathcal{G}$ on n vertices. Given an integer $k \le n$, by a *path* of length k we mean a sequence $a_0, a_1, \dots, a_k$ of distinct vertices (rational numbers) such that each a_i is connected to a_{i+1}, that is, $u_i = a_i - a_{i-1} \in U$. These S-units u_i clearly satisfy

$$u_1 + u_2 + \cdots + u_k = a_k - a_0 \neq 0.$$

We call this path *nondegenerate*, if the above equality is a nondegenerate representation of $a_k - a_0$.

By Lemma 1, the number of nondegenerate paths of length k between any two fixed vertices is at most $C(k, S)$ in $\mathcal{G}$, and a fortiori the same estimate holds in $\mathcal{H}$. Since there are $< n^2$ ways to fix the endpoints a_0, a_k, the total number of nondegenerate paths of length k in $\mathcal{H}$ is

$$< C(k, S) n^2.$$

Next we give a lower estimate for this quantity. Let $\mathcal{H}$ be a subgraph for which the minimal degree is equal to its maximal possible value d_n. We claim that if $d_n > 2^{k-1}$, then from each start-point a_0 there are at least

$$d_n(d_n-1)(d_n-3)\dots(d_n-(2^{k-1}-1))$$

nondegenerate paths of length k in $\mathcal{H}$.

Indeed, as a_0 has degree $\geq d_n$, there are at least d_n choices for a_1. We show that given a nondegenerate path $a_0,\dots,a_i$ with $i<k$, there are at least $d_n-(2^i-1)$ possible choices for a_{i+1}. There are altogether at least d_n edges from a_i to some a_i+u, $u\in U$. Some choices will not produce a path as $a_i+u=a_j$ for some $j<i$, and some choices make the representation degenerate. These events happen exactly if there is a vanishing nonempty subsum of the sum

$$(a_1-a_0)+\dots+(a_i-a_{i-1})+u.$$

This means that u is equal to the negative of one of the 2^i-1 nonempty subsums formed from the numbers a_j-a_{j-1}, $1\leq j\leq i$.

By comparing the lower and upper estimate we see that

$$\left(d_n-2^{k-1}\right)^k < C(k,S)n^2,$$

consequently

$$d_n < C(k,S)^{1/k}n^{2/k}+2^{k-1}.$$

This estimate also trivially holds if the assumption $d_n > 2^{k-1}$ fails, so this is a universal bound for every value of n.

Now we estimate D_n. Take an n-point graph with this average degree. This means that the total number of edges is $nD_n/2$. Take a vertex with minimal degree; removing this vertex decreases the number of edges by at most d_n. Removing a vertex from the remaining graph again decreases the number of vertices by at mos d_{n-1}. Repeating this operation we see that the number of vertices is at most

$$d_n+d_{n-1}+\dots+d_1.$$

Consequently we have

$$D_n \leq \frac{2}{n}\left(d_n+d_{n-1}+\dots+d_1\right) < 2C(k,n)^{1/k}n^{2/k}+2^k.$$

Theorem 4.1 follows by taking $k=1+[2/\varepsilon]$. □

If for the vertices we take the first n integers, the average degree will be of order $(\log n)^{|S|}$. One is inclined to expect that this is not very far from the maximum, which supports the following conjecture.

Conjecture 4.2. With a suitable constant c depending on the set S we have

$$D_n = O\left((\log n)^c\right).$$

Remark 4.3. A referee directed my attention to the following way to make the ε explicit. Theorem 3 of Evertse [2] gives the estimate

$$C_1(k, S) \le \left(2^{35} k^2\right)^{k^3(|S|+1)}.$$

Following the argument described in Section 2 one can obtain an estimate of type

$$C_0(k, S) \le k^{ck^3},$$

with a constant c depending on $|S|$. With the choice

$$k \sim (\log n / \log\log n)^{1/3}$$

this leads to a bound

$$D_n < \exp\left(c'(\log n)^{2/3}(\log\log n)^{1/3}\right).$$

If Evertse's bound could be improved to c^{k^2}, the bound of D_n would also improve to $e^{c'\sqrt{\log n}}$. Any further improvement would, however, only minimally affect our bound, so this conjecture definitely cannot be established by this simple path-counting argument.

We can establish the simplest case when S consists of a single prime.

Theorem 4.4. *Let p be a prime and $S = \{p\}$. The number of edges of an n-point subgraph of $\mathcal{G}$ is at most*

$$f(n) = \sum_{i=0}^{[(\log n)/\log 2]} (n - 2^i) \sim \frac{n \log n}{\log 2}.$$

Equality holds if $p = 2$ and the vertices are n consecutive integers.

Proof. Let $g(n)$ be the maximal possible number of edges. Our aim is to prove $g(n) \le f(n)$. Clearly $g(1) = f(1) = 0$, $g(2) = f(2) = 1$, $g(3) \le f(3) = 3$. For larger values we shall use induction on n.

Let A be a set for which this maximum is attained. By multiplying with a proper power of p we can ensure that the smallest power of p occurring as a difference between elements of A is $p^0 = 1$. Divide A into congruence classes modulo p; there must be at least two classes. Let A_1 be the smallest class, and A_2 the union of the others, and write $|A_i| = n_i$. We have $n_1 \le n_2$.

Within each set A_i there are at most $f(n_i)$ edges by the induction hypothesis. Between A_1 and A_2 the only possible edges come from a difference 1. From an $a \in A_1$ there are two possibilities, $a+1$ and $a-1$, so we get $2n_1$ as an upper bound for the number of such edges. If $n_1 = n_2$, we can reduce this bound by 1. Indeed, let a^* be the smallest element of A. We have $a^* \in A_i$ with $i = 1$ or 2. From each $a \in A_i$ there may be at most 2 edges, and from a^* only 1, as $a^* - 1 \notin A$. This gives $2n_i - 1 = 2n_1 - 1$ possibilities.

Hence we have

$$g(n) \le f(n_1) + f(n_2) + 2n_1 - \delta$$

where n_1, n_2 are integers such that $1 \le n_1 \le n_2$, $n_1 + n_2 = n$, and $\delta = 1$ if $n_1 = n_2$, $\delta = 0$ otherwise. This easily implies $g(n) \le f(n)$. $\square$

I proposed a version of this problem for the 2008 Kürschák competition (this is for freshmen and secondary school students).

Problem 4.5. Determine the exact value of $g(n)$ for primes $p > 2$.

For $p \ge 5$ the extremal configuration is not the first n integers. The first integers yield about $(n \log n)/\log p$ edges. The following configuration is better. For $1 \le i \le n$, write i in base 2 as

$$i = \sum e_j 2^j,$$

and put

$$a_i = \sum e_j p^j.$$

The set $A = \{a_i : 1 \le i \le n\}$ produces about $(n \log n)/\log 4$ edges.

Acknowledgements. I profited much from discussing this subject with Prof. Kálmán Győry, besides being inspired by his Schweitzer problem. I am also indebted to an anonymous referee for pointing out the bound described as Remark 4.3 and for numerous corrections.

References

[1] J.-H. Evertse, On sums of s-units and linear recurrences, *Comp. Math.* **53** (1984), 225–244.

[2] J.-H. Evertse, The number of solutions of decomposable form equations, *Invent. Math.* **122** (1995), 559–601.

[3] K. Győry, On certain graphs associated with an integral domain and their applications to diophantine problems, *Publ. Math. Debrecen* **29** (1982), 79–94.

[4] K. Győry, On certain arithmetic graphs and their applications to diophantine problems, *Funct. Approx. Comment. Math.* **39** (2008), 289–314.

[5] A. J. van der Poorten and H. P. Schlickewei, The growth condition for recurrence sequences, *Macquarie Math. Reports*, no. 82–0041 (1982).

IMRE Z. RUZSA
ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS
H-1364 BUDAPEST, P.O. BOX 127
HUNGARY

E-mail: ruzsa@renyi.hu

(Received February 7, 2011; revised April 18, 2011)