A Dense Infinite Sidon Sequence

Miklós Ajtai, János Komlós and Endre Szemerédi

INTRODUCTION

A sequence of positive integers is called a Sidon sequence if the pairwise sums are all different (for motivation see [6, 7]). We also say that the elements form a difference-set, since an equivalent condition is that the differences are all different.

If \( f_S(n) \) denotes the number of elements in \( S \) not exceeding \( n \) and 
\[
    f(n) = \max f_S(n), \quad S \text{ is Sidon},
\]
then it is known from the works of Erdös, Singer and Turán that 
\[
    f(n) \sim \sqrt{n}.
\]

For an infinite Sidon sequence one always has \( \lim \inf_n f_S(n)/\sqrt{n} = 0 \) ([5], for a detailed proof see [8]). The greedy method shows that there is an infinite sequence \( S \) with 
\[
    f_S(n) > cn^{1/3} \quad \text{for all } n
\]
(see [2]). Indeed, given \( k \) numbers up to \( n \), each triple kills at most 3 other numbers, thus one can always pick one more number if only 
\[
    k + 3 \binom{k}{3} < n.
\]

But the task of constructing a denser sequence has so far resisted all efforts, both constructive and random methods. Here we use a random construction for giving a sequence that is slightly denser than the above trivial one. (However Erdös conjectures that even \( f_S(n) > n^{1-\epsilon} \) is possible.)

Lemma 2 is of independent interest for a graph-theorist.

We also remark that using random construction Erdős and Rényi [4] proved the existence of an infinite sequence \( S \) with 
\[
    f_S(n) > cn^{1-\epsilon} \quad \text{for all } n
\]
such that the number of solutions of the equation 
\[
    a_i + a_j = m, \quad a_i, a_j \in S,
\]
is bounded, less than \( K = K(\epsilon) \) for all \( m \).

1. THE STATEMENTS

THEOREM. There is an infinite Sidon sequence such that 
\[
    f_S(n) > \frac{1}{1000} (n \log n)^{1/3} \quad \text{for } n > n_0.
\]

Before giving the (random) construction, we make some remarks on the independence number of a graph.

If in a graph on \( n \) vertices the maximal valency is \( d \), and the independence number (the maximal number of independent vertices) is \( \alpha \), then obviously 
\[
    \alpha \geq n/(d + 1).
\]
It is interesting to note that this remains true if we replace maximal valency by average valency.

**Lemma 1.** If the average valency is \( t \), then
\[
\alpha \geq n/(t+1).
\]

This can easily be seen by always picking the vertex with the smallest valency, and deleting its neighbours (greedy algorithm). But it is also an immediate consequence of Turán’s theorem, and is directly stated in [1]. See also [3].

Recall that for random graphs the independence number is much larger than \( n/t \), it is around \( n(\log t)/t \).

The basic tool in this paper will be the following lemma which claims that, from the point of view of the independence number, triangle-free graphs behave just like random graphs.

**Lemma 2.** If in a triangle-free graph on \( n \) vertices the average valency is \( t \), \( t > c_1 \), then the independence number \( \alpha \) can be estimated from below as
\[
\alpha > c_3 n (\log t)/t.
\]

**Remark 1.** The hereby given proof gives 1/100 for \( c_3 \), that could easily be improved to 1/10 or so. But we believe that the proper value is \( c_3 = 1 - o(1) \), where \( o(1) \) is meant as \( t \to \infty \).

**Remark 2.** By setting \( n = (1/c_3) t^2/(\log t) \) we get, as a by-product, the following improvement for a Ramsey number:
\[
R(3, t) \leq (1/c_3) t^2/(\log t).
\]

Indeed, if \( G_n \) is triangle-free, and all valencies are at most \( t \) (otherwise the neighbours of a point would form an empty \( t \)-set), then so is the average valency, and thus, by Lemma 2, \( G_n \) contains
\[
c_3 n (\log t)/t = t \text{ independent points.}
\]
(It has been known that \( c_4 t^2/(\log t)^2 < R(3, t) < c_5 t^2(\log \log t)/(\log t) \).)

**Remark 3.** The condition that the graph is triangle-free, could be replaced by the assertion that the graph contains at most \( c_6 nt^2/(\log t)^2 \) triangles. Apply in (2.15) the inequality
\[
\binom{n}{k}/\binom{n-H}{k} < e^{2Hk/n} \text{ for } H + k < n/2.
\]

A weaker consequence is the following.

**Corollary to Lemma 2.** If \( t > c_1 \), and the number \( H \) of triangles is less than \( \varepsilon nt^2 \), \( 1/t < \varepsilon < \varepsilon_0 \), then
\[
\alpha > c_7 n/t \log 1/\varepsilon.
\]

In particular, \( H < nt^{2-\delta} \) implies \( \alpha > c_7 \delta n(\log t)/t \).

Indeed, define a spanned subgraph by keeping each vertex with probability \( p \) and deleting it with \( 1 - p \). For the new parameters (number of vertices, edges, triangles) we (simultaneously) have with a positive probability
\[
n' > np/2, \quad e < 3ep^2, \quad H' < 3Hp^3,
\]
Dense infinite Sidon sequence

whence \( t' = 2e'/n' < 6tp \). Deleting all vertices in triangles we still have at least \( n'/2 \) points, if only \( 3H'n'/2 \). For the remaining graph, we have

\[
3n'' > np/4, \quad e'' < e' < 3Ep^2, \quad H'' = 0,
\]

and thus \( t'' < 12tp \). According to Lemma 2 this subgraph contains at least \( cn''(\log t'')/t'' \) independent points. Choosing

\[
p = \left( \frac{\log t}{16t} \right)^{1/e} < \frac{\sqrt{n}}{H}
\]

we get the corollary.

**Remark 4.** J. Lehel remarked that we actually prove the following. Make a random numbering of the vertices of the graph (if the original numbering is not random enough), and then use the following trivial algorithm (sort of greedy, but not greedy enough).

Among the vertices left, always pick the one with the smallest index, and delete its neighbours.

Then in a triangle-free graph \( G_n \) this algorithm does not stop before giving \( cn(\log t)/t \) independent points, with a probability near 1. (If it does, start from another point. After a certain number of tries (say \( n \)), give it up and use an arbitrary foolproof slow method to decide whether \( a > cn(\log t)/t \) or not. But this happens with a negligible probability, so that this is an almost surely fast algorithm.)

**Notation**

\( |A| \) denotes the cardinality of the set \( A \).

For two sets \( A \) and \( B \) we define \( \chi(A, B) = 1 \) if \( A \) and \( B \) are disjoint, and 0 otherwise.

\( \overline{A} \) denotes the event complementary to \( A \).

\( E \) stands for expectation and \( D^2 \) for variance.

If \( G \) is a graph, \( \mathcal{V}(G) \) and \( \mathcal{E}(G) \) denote the set of vertices and edges, \( v = v(G) = |\mathcal{V}(G)| \),

\[ e = e(G) = |\mathcal{E}(G)|. \]

\( \mathcal{N}_i \) is the neighbourhood of \( i \) including \( i \) itself.

\( t_i \) is the valency of \( i \), \( \mathcal{N}_i = |\mathcal{N}_i| = t_i + 1, \mathcal{N}_{ij} = \mathcal{N}_i \cup \mathcal{N}_j, \mathcal{N}_{ij} = |\mathcal{N}_{ij}|. \)

\[ t = \sum t_i/v(G) = 2e(G)/v(G) \] is the average valency.

\( \sum f(i, j) \) will stand for summation over the edges \((i, j)\) of the graph, each counted once, i.e. for \( i < j, (i, j) \in \mathcal{E}(G) \).

2. Proof of the Theorem

If \( x \leq y \) are positive integers, then the triple \((x, y, x+y)\) is called a general triangle.

**Lemma 3.** Given arbitrary \( N \) general triangles \((x, y, x+y)\), and an interval \((a, a+n)\), there is a difference-set \( D \) in this interval with at least \( \sqrt{n}/3 \) integers such that no more than \( 10N/\sqrt{n} \) of the given general triangles can be found in \( D \) as consecutive differences, i.e. \( d, d+x, d+x+y \) for some \( d \).

**Proof.** We can assume \( a = 0 \). We prove the lemma by constructing \( m = \sqrt{n}/10 \) difference-sets \( D_1, \ldots, D_m \) of size \( \sqrt{n}/3 \) on \((0, n)\) such that no general triangle can occur in two of them.

Let \( p \) be a prime number of the form \( 4k+1 \) for which \( 0.8\sqrt{n} < p < 0.9\sqrt{n} \). Let \( D_{p,j} = 1, \ldots, m \), have the elements \( px + (jx^2), 1 < x < p \), where \( (t) \) stands for the least positive residue of \( t \) (mod \( p \)), and we allow only those \( x \) for which \((jx^2) < p/2 \).

We now show that these \( D_j \) satisfy the requirements.
1. $D_j$ is a difference-set: let

$$u_i = px_i + (jx_i^2) \in D_j \quad i = 1, 2, 3, 4.$$ 

$u_1 - u_2 = u_3 - u_4$ implies that $x_1 - x_2 = x_3 - x_4$ since $0 < (jx_i^2) < p/2$. So we have $x_1 - x_2 \equiv x_3 - x_4 \pmod{p}$ and $x_1^2 - x_2^2 \equiv x_3^2 - x_4^2 \pmod{p}$, therefore obviously $x_1 \equiv x_3 \pmod{p}$ and $x_2 \equiv x_4 \pmod{p}$.

2. Let $T = (ap + c, bp + d, (a + b)p + (c + d))$ be an arbitrary general triangle where $0 \leq c < p, 0 \leq d < p$, and suppose that $T$ is contained in $D_j$ for some $j$. We show that $T$ determines $j$ uniquely. Indeed, if $x_i + (jx_i^2), i = 1, 2, 3$ are the corresponding elements of $D_j$ then $0 < (jx_i^2) < p/2$ implies that

$$x_1 - x_2 = a, \quad x_2 - x_3 = b, \quad (jx_1^2) - (jx_2^2) = c, \quad (jx_2^2) - (jx_3^2) = d.$$ 

Solving these equations (mod $p$) we get

$$j = (a - b)^{-1}(ca^{-1} + db^{-1}) \pmod{p}.$$ 

By definition $0 < j \leq \sqrt{n}/10 < p$, so $j$ is indeed determined by $T$. $-1$ is a quadratic residue (mod $p$), therefore

$$|D_j| = |\{x|(jx^2) < p/2, 1 < x < p\}| = (p - 1)/2 \approx \sqrt{n}/3.$$ 

We say that a set $A$ generates a general triangle $(x, y, x + y)$ if these three numbers are in $A - A = \{a - a'|a, a' \in A\}$.

For all $i \geq i_0$ we are going to construct a sequence $B_i$ of sets of positive integers with the following properties:

(i) $B_i$ is a subset of the interval $[2 \cdot 10^i, 3 \cdot 10^i]$;

(ii) $|B_i| = \lfloor 10^{i+3/2} \rfloor$;

(iii) $B_i$ is a difference-set;

(iv) the set $A_i = \bigcup_{j \leq i} B_j$ generates less than $10^{1.26i}$ general triangles;

(v) for no pair $b, b' \in B_i, b > b'$, is the difference $b - b'$ in $A_{i-1} - A_{i-1}$.

The sequence of the elements of such sets $B_i$ is obviously a Sidon sequence, as dense as required in the theorem.

We will always assume that $i$ is sufficiently large to justify all approximations.

For a set $R$ we define a graph $G_i$ with vertex-set $R$ by joining $b$ and $b', b, b' \in R$, with an edge if $b - b' \in A_{i-1} - A_{i-1}$.

Assume we have already defined $B_j, j < i$, with the above five properties, and start defining $B_i$.

Take a difference-set $D$ on $[2 \cdot 10^i, 3 \cdot 10^i]$ of $\lfloor 10^{i+3/2} \rfloor$ elements, which contains only $10^{0.77i}$ of the $<10^{1.26i}$ general triangles generated by $A_{i-1}$ (Lemma 3). Set $p = 5 \cdot 10^{0.15i}$. We define a random set $R$ by selecting each integer in $D$ independently, with probability $p$.

We have, with probability $>0.9$, that $10^{0.35i} < |R| < 2 \cdot 10^{0.35i}$. If $GTR_i$ is the number of general triangles generated by $A_i$, then for the expectation of $GTR_i - GTR_{i-1}$ we have

$$E(GTR_i - GTR_{i-1}) < (4 \cdot 10^{0.35i})^3 + (4 \cdot 10^{0.35i})^4 \cdot p < \frac{1}{120} 10^{1.26i}$$

since at least two of the six integers determining the general triangle have to be in $B_i$. Thus

$$GTR_i - GTP_{i-1} < \frac{1}{120} 10^{1.26i} \quad \text{with probability} > 0.9,$$

and hence

$$GTR_i < 10^{1.26(i-1)} + \frac{1}{120} 10^{1.26i} < 10^{1.26i} \quad \text{with probability} > 0.9.$$ 

If $TR$ is the number of triangles in $G_i$, then for the expectation of $TR$ we have

$$E TR < 10^{0.77i} p^3 = 125 \cdot 10^{0.32i}.$$
Dense infinite Sidon sequence

Thus, with probability > 0.9, TR < 10^{0.33i}. Let us leave out all points in triangles, so that 
$G_i$ becomes triangle-free.

For the expected number of edges in $G_i$ we have

\[ E(e(G_i)) < \frac{1}{500i^3} 10^{2/3} \epsilon^2 < \frac{1}{100i^3} 10^{(3-0.3)i} \]

and thus, with probability > 0.9,

\[ e(G_i) < \frac{1}{10i^3} 10^{(3-0.3)i} \]

Since four events, each having a probability > 0.9, always have a non-empty intersection, we can choose $R$ in such a way that it satisfies (i), (iii) and (iv).

Since for the average valency $t$ in $G_i$ we have

\[ t < \frac{1}{26i^3} \frac{1}{10} \frac{1}{10} (3-0.65)i, \]

according to Lemma 2, $G_i$ contains at least

\[ 2 \cdot 10^{0.35i} (\log t)/t > \frac{1}{3} 10^{i/3} \]

independent points. \( \frac{1}{100i^3} 10^{i/3} \) of these will form the set $B_i$ which will have the additional properties (ii) and (v).

3. Two Further Lemmas

For the proof of Lemma 2 we will use two lemmas; the second one (Lemma 5) is similar to the Schwartz inequality for graphs, that says that

\[ \sum (t_i + t_j) \geq nt^2. \]

(Indeed, $\sum (t_i + t_j) = \sum t_i^2 \geq nt^2$.)

**Lemma 4.** If in the graph $G$ on $n$ vertices the average valency $t$ is between $c_8$ and $c_9n$, then there is a spanned subgraph

\[ \mathcal{K} = (\mathcal{V}(\mathcal{K}), \mathcal{E}(\mathcal{K})), \]

such that

\[ v(\mathcal{K}) > \frac{1}{100n}/t, \quad e(\mathcal{K}) < \frac{1}{50} v(\mathcal{K}), \]

and that for the spanned subgraph $\mathcal{M}$ of $G$ whose vertices are connected to no vertex in $\mathcal{V}(\mathcal{K})$, we have $v(\mathcal{M}) > n/2$ and

\[ n'/t' \overset{\text{def}}{=} v^2(\mathcal{M})/(2e(\mathcal{M})) > \vartheta v^2(G)/(2e(G)) = \vartheta n/t \]

where $\vartheta = 1 - 1/t - c_{10} \sqrt{t}/n$.

The following lemma forms the core of the proof.

**Lemma 5.** If in a graph $G$ on $n$ vertices the maximum valency is $T$, and we set

\[ S_1 = \frac{1}{n} \sum_{i=1}^{n} e^{-xt_i}, \]

\[ S_2 = \frac{1}{e(G)} \sum e^{-x(t_i + t_j)}, \]

then $S_2 \leq S_1^2$ for $0 \leq x \leq 1/10T$. 
PROOF OF LEMMA 4. We give a random construction. First we assume that \( \max t_i \leq 10t = T \).

Let us define a random spanned subgraph \( \mathcal{H} \) of \( G \) by choosing \( \mathcal{V}(\mathcal{H}) \) at random from among all subsets of \( \mathcal{V}(G) \) with \( k = n/(100t) \) vertices. For a given \( \mathcal{H} \), let \( \mathcal{M} \) denote the graph spanned by those vertices that are connected to no vertex in \( \mathcal{V}(\mathcal{H}) \), i.e.

\[
\mathcal{V}(\mathcal{M}) = \{i \mid \mathcal{N}_i \cap \mathcal{H} = \emptyset\}.
\]

Then

\[
v(\mathcal{M}) = \sum \chi(\mathcal{N}_i, \mathcal{V}(\mathcal{H}))
\]

whence

\[
Ev(\mathcal{M}) = \frac{1}{(n)^k} \sum_{|\mathcal{V}|=k} \sum_i \chi(\mathcal{N}_i, \mathcal{V}) = \frac{\sum_i (n-N_i)}{k} \binom{n}{k}.
\]  

(3.1)

Similarly,

\[
D^2 v(\mathcal{M}) = \sum_{i,j} \left[ E\chi(\mathcal{N}_i, \mathcal{V})\chi(\mathcal{N}_j, \mathcal{V}) - E\chi(\mathcal{N}_i, \mathcal{V})E\chi(\mathcal{N}_j, \mathcal{V}) \right]
\]

\[
= \frac{1}{(n)^k} \sum_{|\mathcal{V}|=k} \sum_i \chi(\mathcal{N}_i, \mathcal{V}) - \frac{\binom{n-N_i}{k} \binom{n}{k}^2}{k} \binom{n}{k}.
\]

(3.2)

Furthermore

\[
e(\mathcal{M}) = \sum' \chi(\mathcal{N}_i, \mathcal{V}(\mathcal{H})),
\]

whence

\[
Ee(\mathcal{M}) = \sum' \binom{n-N_i}{k} \binom{n}{k}
\]  

(3.3)

and

\[
D^2 e(\mathcal{M}) = \sum_{i \neq j} \left[ \frac{\binom{n-|\mathcal{N}_i \cup \mathcal{N}_j|}{k} \binom{n}{k} - \frac{\binom{n-N_i}{k} \binom{n-N_j}{k}}{k} \binom{n}{k} \right].
\]

(3.4)

For \( e \in \mathcal{E}(G) \), \( e = (i, j) \), define \( \chi_e(\mathcal{V}) = 1 \) if both \( i \) and \( j \) are in \( \mathcal{V} \), and 0 otherwise. Then

\[
e(\mathcal{H}) = \sum_{e \in \mathcal{E}(G)} \chi_e(\mathcal{V}(\mathcal{H}))
\]

whence

\[
Ee(\mathcal{H}) = \frac{1}{(n)^k} \sum_{|\mathcal{V}|=k} \sum e \chi_e(\mathcal{V})
\]

\[
= e(G) \frac{n-2}{k-2} \binom{n}{k}
\]

\[
= \frac{1}{2n} \frac{k(k-1)}{n(n-1)} \frac{k^2 t}{2n}
\]  

(3.5)
Similarly
\[ D^2 e(\mathcal{H}) = \sum_{e_1, e_2 \in \mathcal{H}(G)} \left[ E_{X_{e_1}(\mathcal{V})X_{e_2}(\mathcal{V})} - E_{X_{e_1}(\mathcal{V})} E_{X_{e_2}(\mathcal{V})} \right] \]
\[ = \sum_{e_1, e_2 \in \mathcal{H}(G)} \left[ \frac{\left( n - g(e_1, e_2) \right)}{k} - \frac{k^2(k-1)^2}{n^2(n-1)^2} \right] \]
(3.6)
where \( g(e_1, e_2) \) stands for the number of vertices determining the edges \( e_1 \) and \( e_2 \) (\( g = 2, 3 \) or 4).

According to Chebyshev's inequality, the quantities \( v(M), e(M) \) and \( e(H) \) are near their expectations if only the respective \( D^2/E^2 \) are small. Namely, for any random variable \( X \) and positive number \( \lambda \)
\[ P(|X - EX| > \lambda) \leq D^2 X/\lambda^2, \]
in particular
\[ P(|X - EX| < 2DX) > 3/4. \]
But three sets, each having a measure exceeding \( 3/4 \), always have a non-empty intersection, thus there is a choice for \( \mathcal{H} \) such that the inequalities
\[ v(M) > EV(M) - 2Dv(M) \]
\[ e(M) < EE(M) + 2De(M) \]
\[ e(\mathcal{H}) < EE(\mathcal{H}) + 2De(\mathcal{H}) \]
(3.7)
simultaneously hold.

Let us start estimating \( D^2 v(M) \). We split the summand in (3.2) into two terms
\[ \left( \binom{n-N_i}{k} - \binom{n-N_i-N_j}{k} \right)/\binom{n}{k} \]
and
\[ \left( \binom{n-N_i-N_j}{k}/\binom{n}{k} \right) - \left( \binom{n-N_i}{k}/\binom{n}{k} \right)^2 \]
The second term is easily seen to be negative.

Now for \( 0 \leq N_3 \leq N_1 + N_2 \)
\[ \left( \binom{n-N_3}{k} - \binom{n-N_1-N_2}{k} \right)/\binom{n}{k} < \frac{2k(N_1+N_2-N_3)}{n-k-N_1-N_2} < \frac{2k(N_1+N_2-N_3)}{n-k-2T-2}. \]
Thus
\[ D^2 v(M) = \frac{2k}{n-k-2T-2} \sum_{i,j} (N_i + N_j - N_{ij}) \]
\[ \leq \frac{2k}{n-k-2T-2} \sum_i (t_i + 1)^2 \leq \frac{2k}{n-k-2T-2} n(t+1)(T+1) < nt. \]
(3.8)
Now
\[ Ev(M) = \sum_t \left( \binom{n-N_t}{k}/\binom{n}{k} \right) > n\left( \frac{1 - \frac{k}{n-t}}{1} \right)^{t+1} > 0.9n. \]
(3.9)
Thus
\[ Dv(M)/EV(M) < 2\sqrt{t/n}. \]
(3.10)
Now we estimate $D^2 e(\mathcal{M})$. Just as before,

$$D^2 e(\mathcal{M}) < \frac{4k}{n} \sum_{i_1 < i_2 < j_1 < j_2 \in \mathcal{E}(G)} \left| [N_{i_1} \cup N_{j_1}] + [N_{i_2} \cup N_{j_2}] - |N_{i_1} \cup N_{j_1} \cup N_{i_2} \cup N_{j_2}| \right|$$

$$< \frac{4k}{n} \sum_{i} \sum_{t_1, t_2 \in N_i} t_1 t_2 < \frac{4k}{n} T^3 nt = 40nt^3.$$  

Now since $N_{ij} = N_i + N_j$, we have

$$\frac{1}{e(G)} \sum_i N_{ij} < 2 \sum_i t_i N_i < 2T = 20t.$$  

Hence

$$Ee(\mathcal{M}) = \sum \left( \frac{n - N_i}{k} \right) / \binom{n}{k} \geq \frac{nt(n - 2T - 1)}{2 \binom{n}{k}} > \frac{nt}{2} \left( 1 - \frac{k}{n - 2T} \right)^{2T+1} > nt/10.$$  

Thus

$$D e(\mathcal{M}) / (E e(\mathcal{M})) < 100 \sqrt{t/n}. \quad (3.11)$$

Finally we estimate $D^2 e(\mathcal{X})$. The number of pairs $(e_1, e_2)$ of edges for which $g(e_1, e_2) = 3$ is $\sum_i t_i^2 < ntT$, thus

$$D^2 e(\mathcal{X}) < ntT \frac{k^3}{n^2} + nt \frac{k^2}{n^2}$$

and

$$E e(\mathcal{X}) = \frac{1}{2} nt \frac{k(k - 1)}{n(n - 1)}.$$  

Whence

$$D e(\mathcal{X}) / (E e(\mathcal{X})) < 400 \sqrt{t/n}. \quad (3.12)$$

According to (3.7), (3.10), (3.11) and (3.12) there is a $\mathcal{X}$ such that

$$v(\mathcal{M}) > Ev(\mathcal{M})(1 - \delta)$$

$$e(\mathcal{M}) < E e(\mathcal{M})(1 + \delta)$$

where $\delta = 800 \sqrt{t/n}.$

Now we have

$$\frac{(Ev(\mathcal{M}))^2}{2E e(\mathcal{M})} = \frac{\left( \sum \left( \frac{n - N_i}{k} \right) / \binom{n}{k} \right)^2}{2 \sum \left( \frac{n - N_i}{k} \right) / \binom{n}{k}} \geq \frac{n^2}{2e(G) \frac{1}{e(G)} \sum e^{-k(t_i + t_j)/n}} e^{-\left(10\delta/n\right) - \left(1/t\right)}$$

$$> (1 - \delta) \left( 1 - \frac{1}{t} \right)^n \quad (3.14)$$

by Lemma 5 (we used that $N_{ij} = t_i + t_j$, a consequence of triangle-freeness).
Thus by (3.13) there is a $\mathcal{K}$ such that

$$\left(\nu(\mathcal{M})^2/(2e(\mathcal{M}))\right) > (1-\delta)^4 \left(1 - \frac{1}{t}\right)^n > \delta \cdot n/t.$$  

Also,  

$$\nu(\mathcal{M}) > (1-\delta)E \nu(\mathcal{M}) > \delta \cdot 0.9n.$$  

It remains to show that  

$$e(\mathcal{K}) < \frac{1}{50} \nu(\mathcal{K}),$$

but this is a consequence of (3.5) and (3.13).

So far we have assumed that $\max_p t_i \leqslant 10t$. If not, we start with deleting all vertices of valency $>10t$, and then we apply the above random construction for the remaining graph of $n'$ vertices and average valency $t'$. Obviously, we have $n' \geqslant 0.9n$, and it is easy to see that $n'/t' \geqslant n/t$. Thus the lower bound for $\nu(\mathcal{M})$ will change to  

$$\nu(\mathcal{M}) > \theta(0.9)^2 n > u/2,$$

the others remain unchanged.

**Proof of Lemma 5.** Set  

$$F(x) = \frac{1}{e(G)} \sum e^{-x(t_i + t_j) - x(t_i + t_j)}.$$  

Then  

$$F(x) = \sum_{r=0}^{\infty} F^{(r)}(0) x^r/r!.$$  

Now  

$$F^{(r)}(0) = (-1)^r G_r,$$

where  

$$G_r = \frac{1}{e(G)} \sum (t_i + t_j)^r - \frac{1}{n^t} \sum (t_i + t_j)^r.$$  

Set  

$$H_r = \frac{1}{e(G)} \sum (t_i + t_j - 2t)^r - \frac{1}{n^t} \sum (t_i + t_j - 2t)^r;$$

then by the binomial theorem  

$$G_r = \sum_{l=0}^{r} \binom{r}{l} H_l (2t)^{r-l}.$$  

Now  

$$H_0 = 0, \quad H_1 = \frac{2}{nt} \sum t_i(t_i - t) = 2\sigma^2/t$$

where  

$$\sigma^2 = \frac{1}{n} \sum t_i^2 - t^2$$

and  

$$H_l \leqslant (2T)^{l-2} \left(4 + \frac{T}{t}\right) \sigma^2 \quad \text{for } l \geqslant 2.$$
Thus \( G_1 = 25^2/t \) and

\[
|G_r| \leq \frac{2}{Tt}(2T + 2t)^\sigma r^2 < \frac{2}{Tt}(4T)^\sigma r^2 \quad \text{for } r \geq 2
\]

Since \( F(0) = 0, F'(0) < 0 \) (this is Schwartz's inequality), and

\[
|F'(0)|x \geq \sum_{r=2}^{\infty} |F^{(r)}(0)|x^2/r!
\]

for \( 0 < x < 1/10T, F(x) \) is negative (or 0 if \( \sigma^2 = 0 \)) on this interval.

4. THE PROOF OF LEMMA 2

We can assume that \( t \geq \sqrt{n \log n} \), since the triangular freeness always implies that \( \alpha \geq t \), and that would give \( \alpha \geq t \geq n (\log t)/t \).

By Lemma 4 we can pick a set \( K \) of \((1/110)n/t\) independent vertices, and a set \( M \) independent of them with

\[
|M| > n/2
\]

such that the quantity \( n't' \) within \( M \) is almost as large as in the original graph:

\[
n'/t' > \theta n/t, \quad \theta = 1 - 1/t - c_{10}\sqrt{t/n} > 1 - t^{-3/2}.
\]

We repeat this procedure for the subgraph \( M \), picking \( K', M' \) etc. We denote the corresponding parameters in the \( r \)th step by \( n_r, t_r, \theta_r \). If we can go on for \( R = (\log t)/2 \) steps and always have

\[
\theta_r > 1 - 1/(\log t)
\]

then we are home, for

\[
n_r/t_r > (1 - 1/(\log t))^{(\log t)/2} > n/4, \quad r = 1, \ldots, R,
\]

and thus we get \( 1/600n \log t)/t \) independent points.

If we get stuck in the \( r \)th step, i.e. \( \theta_r \leq 1 - 1/(\log t) \) for the first time, then

(a) \( t_r^{-1} \geq 1/(\log t) \).

But (a) implies that in the last step alone we obtained

\[
n_r/(t_r + 1) > (\log t)^{-3} n/2' > n (\log t)^{-3}/\sqrt{t} > n (\log t)/t
\]

independent points.

REFERENCES


Received 14 March 1980 and in revised form 24 November 1980

M. Ajtai, J. Komlós and E. Szemerédi
Mathematical Institute of the Hungarian Academy of Sciences,
Reáltanoda uica 13–15, Budapest 1053, Hungary