QUASI-RANDOM GRAPHS AND PSEUDO-RANDOM BINARY SEQUENCES

András Sárközy
(Joint work with József Borbély)

Eötvös Loránd University, Faculty of Sciences,
Department of Algebra and Number Theory,
H-1117 Budapest, Pázmány Péter sétány 1/C, Hungary

sarkozy@cs.elte.hu

Quasi-random graphs and pseudo-random binary sequences

In 1997 C. Mauduit and Sárközy proposed a new, constructive and quantitative approach to study pseudo-randomness of binary sequences. In particular, we introduced *measures of pseudo-randomness* of binary sequences

$$E_N = (e_1, e_2, \dots, e_N) \in \{-1, +1\}^N. \tag{1}$$

The most important of them is:

DEFINITION 1. If $k \in \mathbb{N}$, $N \in \mathbb{N}$, then the *correlation of order k* of the sequence E_N of form (1) is defined as

$$C_k(E_N) = \max_{D,M} \left| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \dots e_{n+d_k} \right|$$

where $D = (d_1, d_2, \dots, d_k)$ with $0 < d_1 < d_2 < \dots < d_k \leq N - M - d_k$.

Cassaigne, Mauduit and Sárközy showed that if $N \to \infty$, then for a truly random binary sequence E_N of type (1) the value of $C_k(E_N)$ is around $N^{1/2}$.

Thus if $C_k(E_N)$ is "not much greater than $N^{1/2}$", say, $C_k(E_N) < N^{1/2+o(1)}$, then this is considered as a strong PR ($=$ pseudorandom) property of E_N.

Since that more than 200 papers have been written along these lines by different authors, and in these papers many constructions have been presented for large families of binary sequences with strong pseudorandom properties.

Recently, Borbély and Sárközy realized that these results can be also utilized for constructing *graphs* with strong PR properties. In my talk I will speak on our first results (the paper is submitted to the journal Combinatorica).

The importance of giving *explicit* constructions for graphs with strong pseudorandom properties is explained, e.g., in [B. Bollobás, Random Graphs, Chapter XIII], [N. Alon and J. H. Spencer, The Probabilistic Method, Chapter 9], [F, R, K. Chung, R. L. Graham and R. M. Wilson, Quasi-random graphs, Proc. Natl. Acad. Sci. USA 85 (1988), 969–970], [A. Thomason, Pseudo-random graphs, Annals Discrete Math. 33 (1987), 307–331]. Indeed, an explicit construction "may shed more light on the corresponding problem" [Alon–Spencer], explicit constructions "are more illuminating than existence proofs" [Bollobás], "often happens that a random looking structure is useful for a certain algorithmic procedure" [Alon–Spencer].

In spite of this, in the most related papers and monographs only a few explicit constructions are mentioned (mostly the *Payley graph* and also a few further constructions using linear algebra).

The intensive study of PR graphs started in 1987–89 in two papers of Chang, Graham and Wilson (who introduced the notion of quasi-random graphs) and two papers of Thomason (who introduced the notion of "(p, α)-jumbled graph", and they also presented a few more constructions.

In our paper(s) with Borbély our goal is to propose *methods* for constructing explicitly given pseudo-random graphs and to present *large families* of them. In particular, in our first paper we focused on the "quasi-randomness" approach of Chang, Graham and Wilson since it is simpler and more transparent than Thomason's approach (which we will study in the sequels).

We will need the following definitions, notations and facts.

DEFINITION 2. The *adjacency matrix* of the (undirected) graph G_n on the n vertices $V = \{v_1, v_2, \ldots, v_n\}$ is defined as the $n \times n$ matrix $A(G_n) = \left[a(i,j)\right]_{i,j=1}^{n}$ where

$$a(i,j) = \begin{cases} 1 & \text{if the vertices } v_i, v_j \text{ are joined,} \\ 0 & \text{if the vertices } v_i, v_j \text{ are not joined.} \end{cases}$$

The adjacency matrix has the following properties:

(i) every element of it is 0 or 1,

(ii) the elements in the main diagonal are 0,

(iii) it is symmetric: $a(i,j) = a(j,i)$.

Every (undirected) graph uniquely determines its adjacency matrix and vice versa: every square matrix with these properties uniquely determines the graph whose adjacency matrix it is.

In their papers Chang, Graham and Wilson studied graphs G_n with n vertices and edge density $1/2$ (i.e., roughly speaking, two vertices are joined with probability $1/2$). They "…introduce a large equivalence class of graph properties, all of which are shared by so-called random graphs. Unlike random graphs, however, it is often relatively easy to verify that a particular family of graph possesses some property in this class."

They list 7 such properties of graphs G_n with $n \to \infty$. Here we will use only the sixth property P_6 which says:

P_6: writing

$$s(i,j) = \left|\left\{x \in \{1,2,\ldots,n\} :\ a(i,x) = a(j,x)\right\}\right| \quad \text{for } i,j \in \{1,2,\ldots,n\}$$

we have

$$\sum_{i,j\in\{1,2,\ldots,n\}} \left|s(i,j) - \frac{n}{2}\right| = o(n^3).$$

(For almost every pair i,j the number of vertices v_x such that they are joined either to both v_i, v_j or to none of them is about $\frac{n}{2}$ as expected.)

Chung et al. proved:

THEOREM A. *These 7 properties are equivalent.*

Based on this theorem they define the notion of *quasi-random graph* in the following way:

DEFINITION 3. Graphs having any (and *therefore, all*) of the above properties will be called *quasi-random*.

Our goal was to construct quasi-random *graphs* starting out from PR binary *sequences*. The simplest way to define a graph is to define its adjacency *matrix*. But this would mean that we want to make "random type" two-dimensional structures (matrices) from one-dimensional ones (sequences) which may seem a wrong approach. However, taking a look at the *Payley graph* (which is the most important example for quasi-random graph) may put us on the right track. The definition of the Payley graph is: we take a prime q of form $4k+1$, and then the adjacency matrix $[a(i,j)]_{i,j=1}^{q}$ of it is defined by

$$a(i,j) = \begin{cases} 1 & \text{if } \left(\frac{j-i}{q}\right) = +1, \\ 0 & \text{otherwise} \end{cases}$$

(where $\left(\frac{\cdots}{q}\right)$ is the *Legendre symbol*).

E.g., for $q = 13$ this matrix is

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Observe:

(i) in the first row we have 1 if and only if the column index j is such that $j - 1$ is a quadratic residue,

(ii) we get every row from the preceding one by shifting it to the right by one position *cyclically*.

About (i): the first row is closely related to the mod 13 Legendre symbol sequence which is a binary sequence with strong PR properties:

$$\begin{array}{cccccccccccccl}
0, & \left(\frac{1}{13}\right), & \left(\frac{2}{13}\right), & \left(\frac{3}{13}\right), & \left(\frac{4}{13}\right), & \left(\frac{5}{13}\right), & \left(\frac{6}{13}\right), & \left(\frac{7}{13}\right), & \left(\frac{8}{13}\right), & \left(\frac{9}{13}\right), & \left(\frac{10}{13}\right), & \left(\frac{11}{13}\right), & \left(\frac{12}{13}\right) & \\
 & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \\
0, & +1 & -1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 & \\
 & & \downarrow & & & \downarrow & \downarrow & \downarrow & \downarrow & & & \downarrow & & \\
0, & 1, & 0, & 1, & 1, & 0, & 0, & 0, & 0, & 1, & 1, & 0, & 1 & \begin{array}{l}(=\text{first}\\ \text{row of the}\\ \text{matrix})\end{array}
\end{array}$$

So that replacing the -1's in the Legendre symbol sequence by 0's we get the first row.

About (ii): it is a so-called “circulant matrix”:

DEFINITION 4. A *circulant matrix* is a square matrix whose each row vector can be obtained from the preceding vector by rotating it one element to the right, i.e., a matrix Z of the form

$$Z = \begin{pmatrix} z_0 & z_1 & z_2 & \dots & z_{n-2} & z_{n-1} \\ z_{n-1} & z_0 & z_1 & \dots & z_{n-3} & z_{n-2} \\ z_{n-2} & z_{n-1} & z_0 & \dots & z_{n-4} & z_{n-3} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ z_1 & z_2 & z_3 & \dots & z_{n-1} & z_0 \end{pmatrix}. \tag{2}$$

Such a circulant matrix is uniquely determined by its first row

$$Z_n = \big(z_0, z_1, z_2, \dots, z_{n-2}, z_{n-1}\big). \tag{3}$$

Thus we will say

DEFINITION 5. The *circulant matrix* Z in (2) is *generated* by the sequence Z_n in (3), and the matrix generated by the sequence Z_n is denoted by $Z = Z(Z_n)$.

Based on these observations we proposed the following steps for constructing a quasi-random graph:

(i) take a family of binary sequences $E_N = (e_0, e_1, \ldots, e_{N-1}) \in \{-1, +1\}^N$ with strong PR properties (in particular, with small $C_2(E_N)$);

(ii) transform them into binary sequences $F_N = (f_0, f_1, \ldots, f_{N-1}) \in \{0, 1\}^N$;

(iii) adjust these sequences so that $f_0 = 0$ should hold;

(iv) keep a subfamily of these sequences consisting of *symmetric* sequences, i.e., sequences such that extending them in both directions periodically modulo N: $(\ldots, f_{-2}, f_{-1}, f_0, f_1, f_2, \ldots)$ we have $f_i = f_{-i}$ for all $i \in \mathbb{Z}$;

(v) show that it follows from the PR properties considered (more precisely, from the small C_2 of them) that for all the binary sequences belonging to the subfamily described in (iv) are such that the circulant matrices generated by them are the adjacency matrices of graphs possessing property P_6 defined earlier, thus by the theorem of Chung, Graham and Wilson these graphs are quasi-random.

Indeed, we have proved the following theorem which shows that this method works:

THEOREM 1. *Assume that $n \in \mathbb{Z}$, $n \to \infty$,*

$$F_n = (f_0, f_1, \dots, f_{n-1}) \in \{-1, +1\}^n \tag{4}$$

is such that (i) $f_0 = -1$, (ii) F_n *is symmetric,* (iii) $C_2(F_n) = o(n)$. *Define the mapping* $\varphi : \{-1, +1\} \to \{0, 1\}$ *by*

$$\varphi(e) = \frac{1+e}{2} \quad (\textit{for } e \in \{-1, +1\}),$$

and then define the bijection $\Phi : \{-1, +1\}^n \leftrightarrow \{0, 1\}^n$ *by*

$$\Phi(E_n) = \Phi\big((e_0, e_1, \dots, e_{n-1})\big) = \big(\varphi(e_0), \varphi(e_1), \dots, \varphi(e_{n-1})\big)$$

(for $E_n = (e_0, e_1, \dots, e_{n-1}) \in \{-1, +1\}^n$*). Transform the* $-1, +1$ *sequence* F_n *in* (4) *into a bit sequence by using the transformation* Φ*, i.e., write*

$$F'_n = (f'_0, f'_1, \dots, f'_{n-1}) = \Phi(F_n) = \big(\varphi(f_0), \varphi(f_1), \dots, \varphi(f_{n-1})\big),$$

and consider the circulant matrix $Z(F'_n)$. Then this matrix is a bit matrix, its elements in the main diagonal are 0*, and it is symmetric. Thus there is a uniquely determined graph $G_n(F'_n)$ whose adjacency matrix is $Z(F'_n)$. Then this graph $G_n(F'_n)$ possesses property P_6 of Chung, Graham and Wilson, thus it is* quasi-random *by their Theorem A.*

The most important part of the proof is that the sum in P_6 is small $\big(o(n^3)\big)$; this can be derived from assumption (iii) $(C_2(F_n) = o(n))$ relatively easily by a one and half page computation.

Note that the crucial assumptions are (ii) (symmetry) and (iii) $\big(C_2(\ldots) = o(n)\big)$; thus to construct quasi-random graphs we may start out from $-1, +1$ sequences with small C_2, and then we have to look for *symmetric* ones among them.

Finally, in our paper we presented 6 examples for applications of our Theorem 1. First we showed that the quasi-randomness of the Payley graph (which is a well-known fact) is just a (very) special case of our theorem:

COROLLARY 1. *If $q = 4k + 1$ is a prime, then the Payley graph (defined earlier) is quasi-random.*

SKETCH OF THE PROOF. One takes $F_q = \{f_0, f_1, \dots, f_{q-1}\} = \left\{-1, \left(\frac{1}{q}\right), \left(\frac{2}{q}\right), \dots, \left(\frac{q-1}{q}\right)\right\}$ in Theorem 1. Then $G_q(F_q')$ is the Payley graph, and Theorem 1 can be applied since F_q is symmetric, i.e., $\left(\frac{i}{q}\right) = \left(\frac{-i}{q}\right)$ which follows from $q = 4k + 1$, and

$$C_2(F_q) = C_2\left(-1, \left(\frac{1}{q}\right), \left(\frac{2}{q}\right), \dots, \left(\frac{q-1}{q}\right)\right) \ll$$

$$\ll C_2\left(\left(\frac{1}{q}\right), \left(\frac{2}{q}\right), \dots, \left(\frac{q-1}{q}\right)\right) = o(q)$$

which can be proved by Weil's theorem (see, e.g., our '97 paper with Mauduit on pseudorandomness of binary sequences in which we also proved that $C_2\left(\left(\frac{1}{q}\right), \left(\frac{2}{q}\right), \dots, \left(\frac{q-1}{q}\right)\right) \ll q^{1/2}\log q$).

Finally, another (more complicated) application of Theorem 1:

As suggested earlier, we will start out from a family of ± 1 sequences with strong PR properties:

THEOREM B (L. Goubin, C. Mauduit, A. Sárközy, 2004). *Assume: q is a prime, $f(x) \in \mathbb{F}_q[x]$, $\deg f(x) = t\ (> 0)$, $f(x)$ has no multiple zero in $\overline{\mathbb{F}}_q$, and the sequence $E_q = (e_1, e_2, \dots, e_q)$ is defined by*

$$e_n = \begin{cases} \left(\frac{f(n)}{q}\right) & \text{for } (f(n), q) = 1, \\ +1 & \text{for } q \mid f(n) \end{cases}$$

(for $n = 1, 2, \dots, q$) where $\left(\frac{\cdots}{q}\right)$ is the Legendre symbol, $k \in \mathbb{N}$ satisfies one of the following assumptions:

(i) $k = 2$;

(ii) $k < q$, *and* 2 *is a primitive root modulo* q;

(iii) $(4t)^k < q$.

Then we have

$$C_k(E_q) < 10tkq^{1/2} \log q. \tag{5}$$

As we suggested earlier, if we have a family of ± 1 sequences with strong PR properties (small C_2) at hand, then the next step is to find a large subfamily of it consisting of sequences possessing the symmetry property, and then to adjust these sequences to be able to use Theorem 1: we transform them into *bit* sequences and modify them to ensure that their first element is 0. In this case we end up with the following theorem:

THEOREM 2. *Assume that q is prime with $q \to \infty$, $t \in \mathbb{N}$ with*

$$t = o\left(\frac{q^{1/2}}{\log q}\right), \tag{6}$$

and let $a_1, a_2, \dots, a_t \in \mathbb{F}_q$ be such that

$$a_i \neq 0 \quad \text{for} \quad i = 1, 2, \dots, t \tag{7}$$

and

$$a_i^2 \neq a_j^2 \quad \text{for} \quad i, j = 1, 2, \dots, t, \ i \neq j. \tag{8}$$

Define $f(x) \in \mathbb{F}_q[x]$ by

$$f(x) = \prod_{i=1}^{t}(x^2 - a_i^2) = \prod_{i=1}^{t}(x - a_i)(x + a_i) \tag{9}$$

and $F_q = (f_0, f_1, \dots, f_{q-1})$ by

$$f_i = \begin{cases} \left(\frac{f(i)}{q}\right) & \text{for } (f(i), q) = 1, \\ +1 & \text{for } q \mid f(i) \end{cases} \tag{10}$$

(for $i = 0, 1, \dots, q-1$), and write
$\overline{F}_q = (\overline{f}_0, \overline{f}_1, \dots, \overline{f}_{q-1}) = -f_0(f_0, f_1, \dots, f_{q-1}) = (-1, -f_0 f_1, \dots, -f_0 f_{q-1})$.
Then the circulant graphs $G_q = G_q(\overline{F}'_q)$ are quasi-random.

PROOF. We will apply Theorem 1 with $\overline{F}_q$ in place of F_n. Indeed, the first element of $\overline{F}_q$ is -1 so that (i) in the Theorem holds. F_q is symmetric, i.e. $f_i = f_{q-i}$ since the polynomial $f(x)$ in (9) and (10) is even, thus $\overline{F}_q$ is also symmetric so that (ii) also holds. Finally, it follows from (7), (8) and (9) that all the zeros of $f(x)$ are in $\mathbb{F}_q$ and they are distinct, thus we may apply Theorem B with F_q and 2 in place of E_q and k, so that we may apply it to estimate $C_2(F_q)$. Then by (6) (our assumption on t), we obtain from (5) in Theorem B that

$$\begin{aligned} C_2(\overline{F}_q) = C_2(F_q) &< 10tkq^{1/2}\log q = 20tq^{1/2}\log q \\ &= o\left(\frac{q^{1/2}}{\log q}\right) q^{1/2}\log q = o(q) \end{aligned}$$

so that $\overline{F}_q$ also satisfies (iii) in Theorem 1. Thus, indeed, we may apply Theorem 1 with $\overline{F}_q$ in place of F_n, and then we get that the circulant graphs $G_q = G_q(\overline{F}'_q)$ are quasi-random, which was to be proved.

There are 4 more constructions presented in our paper. (Two of them are "good" in a certain quantitative sense, while two others will serve to illustrate certain phenomena in our next papers.)

THANK YOU FOR YOUR ATTENTION!