# The action of a few random permutations on $r$-tuples and an application to cryptography


Joel Friedman[1], Antoine Joux[2], Yuval Roichman[3], Jacques Stern[4], and Jean-Pierre Tillich[5]

[1] Dept. of Mathematics, Univ. of British Columbia, Vancouver V6T 1Z2, Canada, e-mail jf@math.ubc.ca

[2] CELAR, France, e-mail joux@ssig.celar.fr

[3] Dept. of Applied Mathematics, Massachussetts Instit. of Tech., Cambridge MA 02138, USA, e-mail yuval@math.mit.edu, partially supported by Univ. of British Columbia

[4] Ecole Normale Supérieure, 45 rue d'ULM, 75005 Paris, France, e-mail stern@dmi.ens.fr

[5] GREYC, Université de Caen, 14000 Caen, France, e-mail jean-pierre.tillich@info.unicaen.fr, research supported in part by a CNET grant and a NSERC Postdoctoral Fellowship. This author enjoyed the hospitality of the University of British Colombia (Vancouver, Canada) while part of this research was carried out.



**Abstract.** We prove that for every $r$ and $d \geq 2$ there is a $C$ such that for most choices of $d$ permutations $\pi_1, \pi_2, \ldots, \pi_d$ of $S_n$, a product of less than $C \log n$ of these permutations is needed to map any $r$-tuple of distinct integers to another $r$-tuple. We came across this problem while studying a seemingly unrelated cryptographic problem, and use this result in order to show that certain cryptographic devices using permutation automata are highly insecure. The proof techniques we develop here give more general results, and constitute a first step towards the study of expansion properties of random Cayley graphs over the symmetric group, whose relevance to theoretical computer science is well-known (see [B&al90]).


## 1 Introduction

Consider the following random graph model: we independently choose $d$ permutations of the numbers from 1 to $n$, $\pi_1, \pi_2, \cdots, \pi_d$, each permutation equally likely. We construct a directed graph, $G = (V, E)$ with vertex set the set of $r$-tuples of distinct elements of $\{1, \ldots, n\}$ with a directed edge from $(u_1, u_2, \cdots, u_r)$ to $(v_1, v_2, \cdots, v_r)$ iff $(v_1, v_2, \cdots, v_r) = (\pi_i(u_1), \pi_i(u_2), \cdots, \pi_i(u_r))$ for one of those $\pi_i$'s. We denote this probability space of random directed graphs $\mathcal{G}_{n,d,r}$. Such graphs are $d$-regular (and may have multiple edges or self-loops). We will consider the associated space of undirected graphs $\mathcal{G}^*_{n,d,r}$ too. This space is simply obtained from the first one by replacing each directed edge of the former graph with an undirected edge. The latter space is therefore formed by undirected $2d$-regular graphs.

It should be noted that $r = 1$ corresponds to the common probabilistic model of $2d$-regular graphs (as studied in [BS87, Fri91] for example), and that $r = n$ is just the common probabilistic model of random $2d$-regular Cayley graphs over $S_n$.

We will show that for every fixed $r$ and for all $d \geq 2$ almost all graphs in $\mathcal{G}^*_{n,d,r}$ have a small second eigenvalue when the number of vertices becomes large, and that this implies that almost all graphs of $\mathcal{G}^*_{n,d,r}$ and of $\mathcal{G}_{n,d,r}$ are good expanders, and also have a small diameter.

This issue has been raised by the study of the security of some low cost cryptographic devices constructed from permutation automata (see section 2). We will exhibit in what follows a probabilistic algorithm which reconstructs the secrete device, and which can be shown to run in polynomial time by using the aforementioned result. This shows that such schemes are highly insecure.

Actually, the results obtained here are more general than that, and should be put in the broader context of whether or not Cayley graphs over $S_n$ are good expanders for a fixed number of random generators, and/or have a small diameter and mixing time. This is quite an important open problem, for a survey see [B&al90, Lub1, Lub2]. A solution of this problem is of great theoretical importance, while a positive solution would be useful for instance for generating random permutations quickly. Our results can be a first step towards a solution of the above open problem.

Besides being a step towards the study of random Cayley graphs over $S_n$ :

- Those results cover the case of graphs of very small degree, which was not really addressed by previous works of A. Broder, J. Friedman, and E. Shamir (especially the case $d = 2$ which is worth studying!). For instance our results show that as soon as $d \geq 2$, random $2d$-regular graphs have almost always a small second largest eigenvalue, and are therefore good certifiable expanders.
- We address here the issue of the expansion properties of directed graphs too. We provide here tools to achieve such results, provided that the directed graphs we are interested in have the same indegree and outdegree for each vertex. Although most of the theory on expanders has been developed for undirected graphs— we wish to lay emphasis on the fact that for some applications, expansion properties of directed graphs have to be estimated, this is the case for example in [TZ93, Zem94]— and in this article (see section 2).

## 2 A cryptographic device using permutation automata

With the development of memory card technology, and in particular the development of pre-paid cards, that give access to some service, the protection of the service issuer against fraud is becoming a crucial issue. However, for low-cost applications, the service provider might not afford to replace his memory cards by smart cards containing classical cryptographic protocols for identification of genuine cards. Still, it might be possible to devise (classical) identification protocols to improve the security offered by memory cards, while keeping their cost within reasonable bounds. In particular, permutation automata have been considered as offering a general design methodology for such purposes. We recall

here some definitions and describe an identification protocol which, albeit never published, has been circulating in the smart-card community.
**Definition** An automaton is a tuple $(Q, B, \delta, q_0)$ where:

- $Q$ is a finite nonempty set of states,
- $B$ is a finite nonempty set of input symbols or basic actions,
- $\delta$ is the next-state function which maps $Q \times B$ into $Q$,
- $q_0$ is a member of $Q$, the initial state.

**Definition** A *permutation* automaton is such that, for every action $b$, the function $\delta(., b)$ is a permutation of $Q$.

We let $A = B^*$ be the set of finite sequences of basic actions, and we extend the domain of the function $\delta$ to $A$ in the usual way.

We now consider the special case $B = \{0, 1\}$, $Q = [1..n]$ and $q_0 = 1$. Moreover, we fix a parameter $L$ and we let $B^L$ be the subset of all words of length $L$ in $A$.

The basic idea of the identification protocol is to install a secret permutation automaton in each memory card. All these automata are generated from a master key, and the card reader can reconstruct it before starting the identification. During the identification itself, the card reader sends a random query $w$ from $B^L$ to the memory card, which computes the image of $q_0$ by $w$, and outputs this result. The card reader checks this result, accepts the card and issues the service if it is correct, and rejects otherwise.

What is important in the above, is the fact that the length of the queries is fixed. In the designer's mind, this was presumably enough to prevent a statistical analysis of the outputs. This was even expected to remain true if the user was allowed to make repeated experiments with the automaton. Thus our cryptographic problem can be interpreted as a problem of learning theory for automata, for details see for example [Ang78, AS83, F&al93, Gol78, RS87, RS89]. However, for all the attacks based on learning theory that we are aware of, the number of experiments needed to construct an automata which simulates the identification, grows with the length of the query. Thus, these attacks can be defeated by limiting the number of identifications that a single card can perform. Yet, we show that even in this context, the above identification algorithm is insecure by describing an algorithm that allows to reconstruct the given automaton after a few queries (the number of queries does not depend on their size). The algorithm is quite direct but the main achievement of the paper is an actual proof that, with high probability, the algorithm succeeds with only a polynomial number of queries, when the permutation automaton is chosen at random. This is by no means obvious, and relies on the expansion properties of Schreier graphs.

Our algorithm is based on a new representation of finite automata, which plays in our setting the same role as the diversity-based representation of Rivest and Schapire ([RS87]). Using this representation, we describe an algorithm by which we can recover the description of the automaton. We give an account of our numerical experiments and, finally, we prove some results on random permutations from which we can estimate the complexity of the algorithm for randomly chosen permutation automata.

## 3 Representation of a finite automaton by characteristic relations.

Let $U = \{u_1, \cdots, u_r\}$ be a set of elements of $A$ of the same length $k < L$ (we use the notations of section 2). Given a state $q$, we let $E_q$ be the equivalence relation on $U$ defined by

$$\delta(q, u_i) = \delta(q, u_j)$$

We call $E_q$ the characteristic relation attached to $q$. Clearly, if $E_q$ and $E_{q'}$ are distinct, then $q \neq q'$. If the converse holds, we will say that $U$ *discriminates* the given automaton. Simple examples show that it is not always the case that discriminating families exist. Still, as will be discussed further, for randomly chosen automata, these are practically easy to obtain.

The following result shows that characteristic relations can be computed.

**Theorem 1** *Let $a$ be an element of $A$ of length $\leq l$ (where $l < L - k$). Then the characteristic relation attached to $\delta(q_0, a)$ can be computed by performing $r$ experiments with inputs of fixed size $L$.*

Set $q = \delta(q_0, a)$. The results follows from the fact that equality

$$\delta(q, u_i) = \delta(q, u_j)$$

can be tested by comparing the answers given by the automaton to $a^\frown u_i^\frown w$ and $a^\frown u_j^\frown w$, where $w$ is any fixed word of length $L - l - k$.

*Remark: extending to the yes/no case* If the output of the automaton only consists of a yes/no answer depending on the state that is reached after the input has been processed, then another equivalence relation can be defined as

$$\text{answer}(\delta(q, u_i)) = \text{answer}(\delta(q, u_j))$$

This equivalence relation can be used in place of $E_q$ in the algorithm that will be given in the next section.

## 4 The algorithm.

We fix a set $U = \{u_1, \cdots, u_r\}$ of elements of $A$ of the same length $k < L$. Our algorithm has three steps: a sampling step, a computing step and an identification step, each as follows.

**Sampling step :** Pick at random elements $a_i$ of $A$ of (small) length $\leq l$; set $q_i := \delta(q_0, a_i)$. Repeat until the set of equivalence relations $E_{q_i}$ has $n$ distinct members. Renumber the chosen elements so that $E_{q_1}, \cdots E_{q_n}$ are distinct and discard the other values.

**Computing step :** For $i := 1$ to $n$ and for each $b$ in $B$ compute $E_{\delta(q_i,b)}$. *Comment: this can be done by comparing the answers given by the automaton to* $a_i^\frown b^\frown u_j^\frown w$ *where* $w$ *is any fixed word of appropriate length and* $j$ *ranges over* $\{1, \cdots, r\}$.

**Identification step :** Choose random words $w_i$ of length $n$. Compute $E_{\delta(q_0,w_i)}$ using the table built in the computing step and identify this equivalence relation with the output of the automaton under $w_i$. Repeat until all output states have been identified.

## 5 Correctness and complexity of the algorithm

It is not difficult to show that the correctness of the algorithm described in section 4 follows from the truth of the following two statements

1. Any state can be reached from $q_0$ by applying a sequence of actions of length $\leq l$.
2. $U$ *discriminates* the given automaton.

Those properties both ensure that the sampling step terminates successfully and that the computation step is accurate. As for the identification step, it is fairly easy to check that it ends up very quickly with overwhelming probability. Still, it need not be the case that all states are output states: to ensure this property, it is enough to replace statement 1 by the analogous statement with actions of length exactly $l$. Mathematical results on this variant of statement 1 can actually be proven as well but will not be included in the present paper.

The sampling step of this algorithm is the most crucial one, if it can construct $n$ elements $a_i$ such that all $E_{q_i}$ are distinct then the whole algorithm will succeed, otherwise it won't.

We now claim that the two above properties are the consequence of expansion properties of directed random graphs $G_1$ and $G_3$ of $\mathcal{G}_{n,2,1}$ and $\mathcal{G}_{n,2,3}$ respectively, both defined by the two (random) permutations $\pi_1 = \delta(.,0)$ and $\pi_2 = \delta(.,1)$.

**Definition 1.** A directed graph $G(V,E)$ with $n$ vertices is a $c$-expander if, for any subset $X$ of vertices with size $\leq n/2$ the following inequalities hold

$$|N^+(X)| \geq c|X| \text{ and } |N^-(X)| \geq c|X|$$

where $N^+(X)$ (respectively $N^-(X)$) denotes the set of vertices not in $X$ which are endpoints (respectively initial points) of an edge with initial point (respectively endpoint) in $X$.

From Theorem 4 in section 6 we know that $G_1$ is almost always a $\alpha_1$-expander and $G_3$ an $\alpha_3$-expander. We need the following theorem, whose proof can be found, for example, in [Zem94].

**Theorem 2.** *If $G$ is an $\alpha$-expander with $v$ vertices then the diameter of $G$ is smaller than*

$$2(1 + \log_{1+\alpha} v).$$

In particular, $G_1$ has diameter smaller than $2(1 + \log_{1+\alpha_1}(n))$, thus the first property needed for the algorithm to succeed hold as soon as $l > 2(1 + \log_{1+\alpha_1}(n))$. Moreover, the number of elements of length smaller than $l$ is polynomial in $n$ if we choose $l = 2(2 + \log_{1+\alpha_1}(n))$. Thus the sampling step will take polynomial time, once $U$ is correctly chosen.

We now want to prove that the small diameter of $G_3$ implies the second property. Let us remark that in order to prove this property it suffices for any pair of states $(x, y)$ to produce a pair of words of the same length (smaller than $k$) $w_1$ and $w_2$ such that $\delta(x, w_1) = \delta(x, w_2)$ and $\delta(y, w_1) \neq \delta(y, w_2)$. $w_1$ and $w_2$ can be completed to length $k$ by appending any fixed word of appropriate length at their ends. We construct $U$ as the union of all words $w_1(x, y)$ and $w_2(x, y)$.

Let $d_3$ denotes the diameter of $G_3$, and given a pair $(x, y)$, choose $(z, r, s)$ such that $x, y, z, r, s$ are all distinct (we suppose that $n \geq 5$). Since $G_3$ has diameter $d_3$, there exists a word $m_1$ of length $\leq d_3$ that goes from edge $(x, y, z)$ to edge $(y, z, r)$ and likewise a word $m_2$ of length $\leq d_3$ that goes from edge $(x, y, z)$ to edge $(y, z, s)$. Let $w_1 = m_1 \frown m_2$ and $w_2 = m_2 \frown m_1$, then clearly, $w_1$ and $w_2$ have the same length and:

$$\delta(x, w_1) = z = \delta(x, w_2) \text{ and } \delta(y, w_1) = s \neq r = \delta(y, w_2).$$

Thus the second property holds, if $k \geq 2d_3 = 4(1 + \log_{1+\alpha_3}(n))$. Moreover, it suffices to choose $U = B^k$, the set of all words of length $k$, whose size is polynomial in $n$ if we choose $k = 4(1 + \log_{1+\alpha_3}(n))$.

Thus, we have constructed a polynomial time algorithm that reconstructs the secret automaton of a given card as soon as $n$ is large enough for the expansion properties of $G_1$ and $G_3$ to hold. Moreover, we have implemented this algorithm for a small value $n = 8$ and even in this case it succeeds with a good probability, and with less than 30 queries.

## 6 The main theorem

We state in this section our main theoretical results, full proofs are given in [FJRST95].

Before describing our main theorem, let us introduce some notations. Let us recall that the adjacency matrix $A = (a_{ij})$ of a graph with $N$ vertices is the $N \times N$ matrix indexed by the vertices of the graph, such that entry $a_{ij}$ is the number of edges between $i$ and $j$. In our case the adjacency matrix $A = (a_{i,j})$ of the graph $G^*$ of $\mathcal{G}^*_{n,d,r}$ obtained by choosing the permutations $\pi_1, \pi_2, \cdots, \pi_d$, is defined by :

$$a_{ij} = \#\{l \mid \pi_l(i) = j\} + \#\{l \mid \pi_l(j) = i\}$$

Let us note that each self-loop counts twice for the corresponding (diagonal) entry of the adjacency matrix. $G^*$ is a $2d$-regular graph, therefore its adjacency matrix has real eigenvalues $2d = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ with $N = n(n-1)\cdots(n-r+1)$; let $\rho = \max_{i\geq 2}|\lambda_i| = \max(\lambda_2, -\lambda_N)$. Our main result asserts that $\rho$ is almost always well separated from $2d$

**Theorem 3.** *If $k \leq 2\lfloor (r+1)\log_{\frac{d^2}{2d-1}} n\rfloor$, then*

$$E\{\rho^k\} \leq \left[2d\left(\frac{\sqrt{2d-1}}{d}\right)^{1/(r+1)}\Big(1+O(\log\log n/\log n)\Big)\right]^k,$$

*and for every $\epsilon > 0$ we have*

$$\text{Prob}\left\{\rho \leq (1+\epsilon)2d\left(\frac{\sqrt{2d-1}}{d}\right)^{1/(r+1)}\right\} = 1 - o_\epsilon(1)$$

As corollaries we obtain that almost all graphs of $\mathcal{G}^*_{n,d,r}$ are good expanders, and the same property holds for almost all graphs of $\mathcal{G}_{n,d,r}$. The result for directed graphs follows from an argument relating the (edge-)expansion properties of a directed graphs to the (edge-)expansion properties of its associated undirected graph, obtained by replacing each directed edge by an undirected edge, and then relating the (edge-)expansion properties of an undirected to the second largest eigenvalue in absolute value of the undirected graph.

More precisely :

**Theorem 4.** *For every fixed $d \geq 2$,$r \geq 1$ and real $\epsilon$, the probability that the graph $\mathcal{G}_{n,d,r}$ is a c-expander tends to 1, as $n$ tends to infinity, for*

$$c = \frac{1-\epsilon}{2}\left(1-\left(\frac{\sqrt{2d-1}}{d}\right)^{\frac{1}{1+r}}\right)$$

Note that this result is far from being optimal, especially for $r = 1$, where the standard counting argument as used for example in [Bol88, Fri91] for undirected graphs gives us sharper estimates on the expansion constant. It must be noted here that this argument applies to the case $d = 2$ to the directed graph model with $r = 1$, and shows that $c > 0.16$, whereas the bound of the theorem 4 gives only $c > 0.034$ (see [JST93]). Unfortunately, this counting trick seems to fail for $r > 1$. Nevertheless, for most practical applications, this theorem is actually sufficient (see for example section 3).

### 6.1 Sketch of the proof of the main theorem

We will sketch here the proof of Theorem 3.

**Remark** Throughout this section we view $r$ as fixed. And we use the notations Let $N = n(n-1)\cdots(n-r+1)$ be the number of vertices of the graphs $\mathcal{G}^*_{n,d,r}$

or $\mathcal{G}_{n,d,r}$ we consider here. We note $\Pi$ the alphabet $\{\pi_1, \pi_1^{-1}, \pi_2, \pi_2^{-1}, \cdots, \pi_d, \pi_d^{-1}\}$, where the $\pi_i$'s are permutations in $S_n$.

We begin by describing the general approach, which follows essentially the approach initiated by Broder and Shamir in [BS87].

The idea of the proof is to get a rather tight upper bound on $\mathrm{E}\left\{\rho^{2k}\right\}$ for rather large values of $k$. This is obtained by upper bounding this quantity by the expectation of the number of closed walks of length $2k$ minus $(2d)^{2k}$, when we choose a random graph from $\mathcal{G}^*_{n,d,r}$.

This is justified by what follows : if we call $A = (a_{ij})$ the adjacency matrix of a random graph of $\mathcal{G}^*_{n,d,r}$, then an entry $b_{ij}$ of $A^{2k}$ represents the number of walks on the graph of length $2k$ from $i$ to $j$. Therefore

$$\text{Number of closed walks of length } 2k = \mathrm{Tr}(A^{2k}) = \sum_{i=1}^{N} \lambda_i^{2k} \geq (2d)^{2k} + \rho^{2k}$$

Let us note that the expectation $\mathrm{E}\left\{i \overset{2k}{\rightarrow} i\right\}$ of the number of walks starting from a vertex $i$ and ending at the same vertex, can be seen as the probability of the following event. We first choose a random word $w = w_1 w_2 \cdots w_{2k}$ in $\Pi^{2k}$ (all the $(2d)^{2k}$ possible words are chosen with the same probability $\frac{1}{(2d)^{2k}}$), and then we assign the letters $\pi_i$ a permutation of $S_n$ chosen uniformly at random. We have $\mathrm{E}\left\{i \overset{2k}{\rightarrow} i\right\} = \mathrm{Prob}\left\{w_1 w_2 \cdots w_{2k}(i) = i\right\}$.

So for each word $w$ of length $2k$ over the alphabet $\Pi$, and for each vertex $v$ of the graph $\mathcal{G}^*_{n,d,r}$, let $P(w,v)$ denote the probability that when $\pi_1, \ldots, \pi_d$ are assigned permutations at random, the walk determined by $w$ starting in $v$ ends in $v$. Clearly $P(w,v) = P(w)$ is independent of $v$. Straightforwrd calculations show that

$$\begin{aligned}\mathrm{E}\left\{\rho^{2k}\right\} &\leq \left(\sum_{i=1}^{N} \mathrm{E}\left\{i \overset{2k}{\rightarrow} i\right\}\right) - (2d)^{2k} \\ &\leq N \sum_{w \in \Pi^{2k}} \Big(P(w) - 1/N\Big) \qquad (1)\end{aligned}$$

Our task is reduced now to estimate the quantity $P(w) - \frac{1}{N}$. By using techniques developped in [BS87, Fri91] it is enough to estimate this quantity when $w$ is a strongly irreducible word, that is $w$ contains no consecutive occurrence of $\pi$ and $\pi^{-1}$ for any $\pi \in \Pi$, and the same property holds for the concatenation $w^\frown w$. Indeed, it can be checked that, by repeatedly cancelling in a given word $w$ first and last letters if they are inverses, and occurences of consecutive inverse letters, we obtain a strongly irreducible word $w'$ such that $P(w) = P(w')$.

The key lemma which shows that the quantity $P(w) - 1/N$ is rather small is

**Lemma 5.** *Let $w$ be a strongly irreducible word of length $2s$ with $s > 1$ such that $w$ is not of the form $w = u^m$ for some $u \in \Pi^*$ and some $m \geq 2$. Then*

$$P(w) = \frac{1}{n^r} + O\left(\frac{s^{2r+2}}{n^{r+1}}\right).$$

The proof of this lemma can be found in [FJRST95]. Let $k$ be fixed. For $s > 1$ let $R_s$ be the set of words in $\Pi^{2k}$ which strongly reduce to a word of length $2s$ which is not periodic; let $R_0$ be the rest of the words, i.e. those that strongly reduce to the empty word or to a periodic word. Let $p_{2k,2s}$ be the probability that a random word in $\Pi^{2k}$ reduces to a word in $R_s$, and let $P_s$ denote the average of $P(w)$ over all words, $w$, in $R_s$. Clearly we have

$$\sum_{w\in\Pi^{2k}} \Big(P(w) - 1/N\Big)/(2d)^{2k} = \sum_{s=0}^{k} p_{2k,2s}(P_s - 1/N).$$

Our main lemma gives us a bound on $P_s$ for $s \geq 1$; we trivially have $P_0 \leq 1$. To estimate the right-hand-side of the above equation we only need to estimate $p_{2k,0}$

**Lemma 6.** *We have*

$$p_{2k,0} \leq (3k+1)\left(\frac{2d-1}{d^2}\right)^k$$

(for a proof see [FJRST95]).

Fromhere the main theorem easily follows. Straightforward calculations show that

$$\sum_{s=0}^{k} p_{2k,2s}(P_s - 1/N) \leq (3k+1)\left(\frac{2d-1}{d^2}\right)^k + O(k^{2r+2}/n^{r+1}).$$

Combining this with equation 1, and using $N \leq n^r$, yields

$$\mathrm{E}\left\{\rho^{2k}\right\} \leq \left(O(k^{2r+2}/n) + 4n^r(3k+1)\left(\frac{2d-1}{d^2}\right)^k\right)(2d)^{2k}.$$

Taking $k$ to be the greatest integer $K$ less than $(r+1)\log_{\frac{d^2}{2d-1}} n$, we have

$$\left(\mathrm{E}\left\{\rho^{2K}\right\}\right)^{1/(2K)} \leq 2d\left(\frac{\sqrt{2d-1}}{d}\right)^{1/(r+1)}\Big(1 + O(\log\log n/\log n)\Big).$$

Finally, Hölder's inequality implies that for any $k \leq 2K$

$$\mathrm{E}\left\{\rho\right\} \leq (\mathrm{E}\left\{\rho^k\right\})^{1/k} \leq \left(\mathrm{E}\left\{\rho^{2K}\right\}\right)^{1/(2K)} \leq 2d\left(\frac{\sqrt{2d-1}}{d}\right)^{1/(r+1)}\Big(1+O(\log\log n/\log n)\Big) \tag{2}$$

which completes the proof of the first statement of the main theorem. The second claim is just a consequence of Markov's inequality :

$$\mathrm{Prob}\left\{\rho > \alpha\right\} \leq \frac{\mathrm{E}\left\{\rho^{2l}\right\}}{\alpha^{2l}},$$

by putting $l = K$ and $\alpha = (1+\epsilon)2d\left(\frac{\sqrt{2d-1}}{d}\right)^{1/(r+1)}$, and using inequality 2.

□

### 6.2 The link between expansion properties of $\mathcal{G}_{n,d,r}$ and $\mathcal{G}^*_{n,d,r}$

To obtain this link, we will not compare directly the expansion constant of a directed graph $G$ and its associated undirected graph $G^*$ (which is the graph obtained from the directed one by replacing each directed edge by an undirected edge), but we will compare their isoperimetric number first. This number is defined as follows

**Definition**

The isoperimetric number $i$ of a directed graph $G(V, E)$ with $n$ vertices is the largest number such that for any subset $X$ of vertices with size $\leq n/2$ the following inequalities hold

$$|\partial^+(X)| \geq i|X|$$

$$|\partial^-(X)| \geq i|X|$$

where $\partial^+(X)$ (respectively $\partial^-(X)$) denotes the set of edges with initial point (respectively endpoint) in $X$ and endpoint (respectively initial point) in $V \setminus X$. The isoperimetric number $i^*$ of an undirected graph $G^*(V, E)$ with $n$ vertices is the largest number such that for any subset $X$ of vertices with size $\leq n/2$ one has

$$|\partial(X)| \geq i^*|X|$$

where $\partial(X)$ denotes the set of edges between $X$ and $V \setminus X$.

For a *regular* directed graph $G$ the following lemma can be obtained

**Lemma 7.** *Assume that $G$ is a regular directed graph, and $G^*$ its associated undirected graph. Let $i$ and $i^*$ be the isoperimetric numbers of $G$ and $G^*$ respectively. Then*

$$i = \frac{i^*}{2}$$

**Proof** For a directed regular graph, for any subset $X$ of vertices of the graph $|\partial^+(X)| = |\partial^-(X)|$. When we consider the associated undirected graph, we obtain $|\partial(X)| = |\partial^+(X)| + |\partial^-(X)|$. Therefore $|\partial(X)| = 2|\partial^+(X)| = 2|\partial^-(X)|$ and these equalities imply the lemma. □

It is readily checked that a directed $d$-regular graph whose isoperimetric number is $i$ is a $i/d$-expander (see [FJRST95]), and that the isoperimetric number $i^*$ of its associated undirected graph is greater than or equal to $d - \frac{\lambda}{2}$ by using theorem 4.1 given in [Moh89], where $\lambda$ is the second largest eigenvalue of this undirected graph in absolute value. Therefore

**Lemma 8.** *Let $G$ be a regular directed graph of degree $d$, $G^*$ its associated undirected graph. Let $\lambda$ be the second largest eigenvalue in absolute value of the adjacency matrix of $G^*$. Then $G$ is a $c$-expander, where $c = \frac{1}{2} - \frac{\lambda}{4d}$.*

Theorem 4 appears therefore as a consequence of lemma 8 and theorem 3.

## References


[AKS83] M. Ajtai, J. Komlòs, E. Szemerédi, "Sorting in $c \log n$ parallel steps", *Combinatorica* **3** (1983), 1-19.

[Ang78] D. Angluin. "On the complexity of minimum inference of regular sets", *Information and Control* **39** (1978), 302-320.

[AS83] D. Angluin and C.H. Smith. "Inductive inference, theory and methods", *Computing Surveys* **15(3)** (1983), 237-269.

[AM85] N. Alon and V.D. Milman. "$\lambda_1$, isoperimetric inequalities for graphs and superconcentrators", *J. Comb. Theory*, Ser. B, **38**, (1985), 73-88.

[Bab94] L. Babai. "Transparent proofs and limits to approximation", *preprint*, (1994).

[B&al90] L. Babai, G. Hetyei, W.M. Kantor, A. Lubotzky, A. Seres. "On the diameter of finite groups", *31st annual Symposium on Foundations of Computer Science*, (1990), 857-865.

[BGG90] M. Bellare, O. Goldreich, S. Goldwasser. "Randomness in interactive proofs", *31st Annual Symposium on Foundations of Computer Science*, IEEE Computer Society Press, (1990), 563-572.

[Bol85] B. Bollobas. *Random Graphs*, Academic Press, London (1985).

[Bol88] B. Bollobas. "The isoperimetric number of random regular graphs", *Europ. J. Combinatorics* **9** (1988), 241-244.

[BV82] B. Bollobas and W. F. de la Vega. "The diameter of random-regular graphs", Combinatorica, **2**, (1982), 125-134.

[BS87] A. Broder, E. Shamir. "On the second eigenvalue of random regular graphs", *28th annual Symposium on Foundations of Computer Science*, (1987), 286-284.

[Del89] C. Delorme. "Counting closed paths in trees", *Technical Report* n.516, University of Paris-Sud, Laboratoire de recherche en informatique Orsay, September 1989 (in French).

[Fil91] J. Fill. "Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains with an application to the exclusion processes" Ann. Appl. Prob. 1, (1991), 62-87.

[F&al93] Y. Freund, M. Kearns, D. Ron, R. Rubinfeld, R.E. Schapire and L. Sellie. "Efficient learning of typical finite automata from random walks", *25th ACM Symposium on the Theory of Computing* (1993), 315-324.

[FJRST95] J. Friedman,A. Joux,Y. Roichman,J. Stern,J.P. Tillich. "The action of a few permutations on $r$-tuples is quickly transitive", *submitted.*

[Fri91] J. Friedman. "On the second eigenvalue and random walks in random $d$-regular graphs", *Combinatorica* **11** (4) (1991), 331-362.

[FKS89] J. Friedman, J. Kahn, E. Szemeredi. "On the second eigenvalue in random regular graphs", *21st annual Symposium on Theory of Computing*, ACM press, (1989), 587-598.

[Gol78] E.M. Gold. "Complexity of automaton identification from given data", *Information and Control* **37** (1978), 302-320.

[G&al90] O. Goldreich, R. Impagliazzo, L. Levin, R. Venkatesen, D. Zuckerman. "Security preserving amplification of randomness", *31st Annual Symposium on Foundations of Computer Science*, IEEE Computer Society Press, (1990), 318-326.

[IZ89] R. Impagliazzo, D. Zuckerman. "How to recycle random bits", *30th Annual Symposium on Foundations of Computer Science*, IEEE Computer Society Press, (1989), 248-253.

[JST93] A. Joux, J. Stern, J.P. Tillich. "Inferring finite automata by queries of fixed length", *Preprint.*

[Kah91] N. Kahale. "Better expansions for Ramanujan graphs", *32nd Annual Symposium on Foundations of Computer Science* (1991), 398-404.

[Kah92] N. Kahale. "On the second eigenvalue and linear expansion of regular graphs", *33rd Annual Symposium on Foundations of Computer Science* (1992), 296-303.

[LR92] J. Lafferty, D. Rockmore. "Fast Fourier analysis for $SL_2$ over a finite field, and related numerical experiments", *Experimental Mathematics* **1**, (1992), 115-139.

[Lub1] A. Lubotzky. *Discrete groups, expanding graphs and invariant measures*, Progress in Mathematics, Vol. 125, Birkhäuser 1994.

[Lub2] A. Lubotzky. "Cayley graphs: eigenvalues, expanders and random walks", to appear in Survey in Combinatorics, 1995.

[McK81] B. McKay. "The expected eigenvalue distribution of a large regular graph", *Linear Algebra and its Applications*, **40**, (1981), 203-216.

[Mih89] M. Mihail. "Conductance and convergence of Markov chains—a combinatorial treatment of expanders", *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, 1989.

[Moh89] B. Mohar. "Isoperimetric number of graphs", *Journal of Comb. Theory* **(B)** (1989), 274-291.

[Pip77] N. Pippenger. "Superconcentrators", *SIAM J. Comput.*, **6**, (1977), 298-304.

[RS87] R.L. Rivest and R.E. Schapire. "Diversity based inference of finite automata" *Proceedings of the 28th Annual Symposium on the Foundations of Computer Science* (1987), 78-87.

[RS89] R.L. Rivest and R.E. Schapire. "Inference of finite automata using homing sequences" *Proceedings of the 21st ACM Symposium on the Theory of Computing* (1989), 411-420.

[Tan84] R.M. Tanner. "Explicit constructions of concentrators from generalized $N$-gons", *SIAM J. Alg. Disc. Meth.*, **5**, (1984), 287-293.

[TZ93] J.P. Tillich, G. Zémor. "Group-theoretic hash functions", *Proceedings of the 1st French-Israeli Workshop in algebraic coding 1993*, Springer Verlag, Lecture Notes **781**, 90-110.

[TZ94] J.P. Tillich, G. Zémor. "Hashing with $SL_2$", *Advances in Cryptology, Proceedings of CRYPTO94*, Springer Verlag, Lecture Notes **839**, 40-49.

[Vaz91] U. Vazirani. "Rapidly mixing markov chains", *Proceedings of Symposia in Applied Mathematics*, Volume **44**, (1991), 99-121.

[Zem94] G. Zémor. "Hash Functions and Cayley graphs", to appear in *Design, Codes and Cryptography*, of October 1994.

This article was processed using the LaTeX macro package with LLNCS style