

Walks on generating sets of Abelian groups^{*}

P. Diaconis¹, L. Saloff-Coste²

¹ Harvard University, Department of Mathematics, Cambridge, MA 02138, USA

² CNRS, Université Paul Sabatier, Statistique et Probabilités, F-31062 Toulouse Cedex, France
(e-mail: lsc@cict.fr)

Received: 28 July 1995 / In revised form: 18 January 1996

Summary. This paper studies a challenge problem posed by D. Aldous which also arises in algorithms for manipulating finite groups. The main tools used are comparison of two Markov chains on different but related state spaces and logarithmic Sobolev inequalities. As usual, the comparison argument involves some combinatorics of paths.

Mathematics Subject Classification (1991): 60J10, 60K35

1. Introduction

In a survey talk at the 1994 national IMS meeting, David Aldous suggested a natural class of problems which seemed beyond available theory. The same problems arise in the analysis of widely used algorithms for calculations with large finite groups [3, 13]. We give below reasonable (if not perfect) analyses of special cases of Aldous' problem.

Let \mathcal{G} be a connected graph with vertex set \mathcal{V} and oriented edge set \mathcal{A} . We assume throughout that \mathcal{A} is symmetric (i.e., $(u, v) \in \mathcal{A}$ implies $(v, u) \in \mathcal{A}$) and that there are no self loops. We write $u \sim v$ if $(u, v) \in \mathcal{A}$, i.e., if u and v are neighbors in \mathcal{G} . Let G be a finite group. We consider a process in $G^{\mathcal{V}}$. Each vertex $v \in \mathcal{V}$ is labeled by an element of the finite group G . A typical step picks an edge $a = (u, v)$ uniformly at random in \mathcal{A} . The group element at v is then multiplied, say on the right, by the group element at u or its inverse each with equal probability $1/2$. All the other vertices remain unchanged. Let us describe two special cases of interest.

* Research partially supported by NATO grant CRG 950686

Example 1a: A particle system

Let G be the two element group $\mathbb{Z}_2 = \{0, 1\}$. Then the sites can be viewed as ON/OFF. The process can be described as follows: At each time, a vertex v is chosen at random (with probability proportional to the number $N(v)$ of neighbors). If v is OFF, the chain stays in its current state. If v is ON, a randomly chosen neighbor of v is changed to its opposite. An alternative description is as follows: pick an oriented edge (v, u) uniformly at random. If v is OFF, the chain stays. If v is ON, u is changed to its opposite.

In Sect. 2, we show that this chain started at $x_0 \in \mathcal{X} = \{x : x(v) = 1 \text{ for some } v \in \mathcal{V}\}$ is an aperiodic irreducible chain with uniform stationary distribution. We also give a bound of order $|\mathcal{V}|^5$ for the relaxation time. This is the original problem of Aldous. Better bounds are derived for specific graphs: on the complete graph on n vertices the mixing time is at most order $n^2 \log n$. We conjecture that the right answer is about $n \log n$ for the complete graph on n vertices.

Example 1b: Group algorithms

Let \mathcal{G} be the complete graph on n vertices. Let G be a finite group. Let S be a set of generators of G . Assume $n > |S|$. Start the walk by labeling $|S|$ vertices with the elements of S and label all the remaining vertices with the identity. It is easy to see that the walk always contains a generating set at the vertices. Variants of this walk are used to generate random elements of G in a variety of implementations. See [3]. The idea is that one may begin with a simple generating set (e.g., transpositions in the symmetric group). After a while, the vertices are labeled with an essentially random generating set and the algorithm outputs the label at site v each time. A good deal of experimentation [3] indicates that this "shuffling of generators" gives crucial speed ups. Only very special cases will be considered here. Besides the case $G = \mathbb{Z}_2$ studied in Sect. 2, we will treat the case $G = \mathbb{Z}_{p^a}$ of a cyclic p -group, p prime in Sect. 3, and the case $G = \mathbb{Z}_p^k$ in Sect. 4. More general cases, including results for the symmetric group, will be presented in the companion paper [11].

Let us now give a formal description of the walk associated with a finite graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ and a finite group G . The state space is a subset \mathcal{X} of the set of all functions $x : \mathcal{V} \rightarrow G$, and the walk is given by

$$P(x, y) = \frac{N(x, y)}{2|\mathcal{A}|} \quad (1.1)$$

where $N(x, y)$ is defined as follows: If the labelings x and y differ at more than one vertex then $N(x, y) = 0$. If x and y differ exactly at one $v \in \mathcal{V}$ and if $x(v) = g$, $y(v) = h$, then $N(x, y)$ is the number of vertices u such that $u \sim v$ and $x(u) = [g^{-1}h]^{\pm 1}$ ($N(x, y)$ is twice this number if $g^{-1}h = [g^{-1}h]^{-1}$). Finally, $N(x, x)$ is equal to twice the number of edges (u, v) such that $x(u) = \text{id}$. Observe that $N(x, y) = N(y, x)$ so that this process is always reversible with the uniform measure as reversing measure. However, it is never irreducible if we take \mathcal{X}

equal to the set of all maps from \mathcal{V} to G . For instance, the map $x \equiv \text{id}$ has no neighbors.

Here is a brief description of our main results. In Sect. 2 we treat the binary case ($G = \mathbb{Z}_2$). For the complete graph, we show that order $n^2 \log n$ steps suffice for convergence to stationarity. This and a comparison argument are used to treat general graphs. The heart of the arguments involve log-Sobolev techniques and a simple approach to comparing two chains on different spaces using a kind of “interpolation”.

Section 3 shows how the techniques of Sect. 2 extend when $G = \mathbb{Z}_{p^a}$, p prime. In this case, we show that order $n^2 p^{2a}$ steps suffice (up to logarithmic factors). One reason for including this extension is to expose the dramatic combinatorial explosion in the interpolation argument.

Section 4 uses a different set of techniques to bound the walk on the complete graph when $G = \mathbb{Z}_p^k$ for prime p . We show that $n^4 (\log p)^3$ steps suffice. For $k = 1$, fixed n and large p , this is much sharper than what was proved in Sect. 3. It begins to show the kind of speed ups claimed in practice. The results lean on hard character estimates of Hildebrand and the deeper parts of expander theory.

In [11], we treat the walk (1.1) on the complete graph for more general finite groups. In particular, we show that for any fixed G and large n , order $n^2 \log n$ steps are enough. We also consider cases where both n and G grow. Sections 2,3 of this paper and [11] use the same comparison techniques. The combinatorics of paths is however much more complicated in [11] where non trivial results from finite group theory are also needed.

Background and notation

This paper uses the geometric tools of Markov chain theory developed in [8–10, 12]. Some notation used below concludes this introduction. Given any reversible Markov chain on a finite state space \mathcal{X} , with kernel $P(x, y)$ and stationary measure π , we set

$$\text{Var}_\pi(f) = \frac{1}{2} \sum_{x, y \in \mathcal{X}} |f(x) - f(y)|^2 \pi(x)\pi(y), \tag{1.2}$$

$$\mathcal{E}_P(f, f) = \frac{1}{2} \sum_{x, y \in \mathcal{X}} |f(x) - f(y)|^2 P(x, y)\pi(x) \tag{1.3}$$

and

$$\mathcal{L}_\pi(f) = \sum_x |f(x)|^2 \log \left(\frac{|f(x)|}{\|f\|_2} \right)^2 \pi(x) \tag{1.4}$$

where $\|f\|_2 = (\sum_x |f(x)|^2 \pi(x))^{1/2}$. The subscripts will be dropped whenever no confusion could possibly arise.

For the iterated kernel of P , we use the notation

$$P_x^\ell(y) = P^\ell(x, y) = \sum_z P^{\ell-1}(x, z)P(z, y).$$

To measure distances between probability distributions, we use the total variation distance

$$\|\pi - \mu\|_{TV} = \max_{A \subset \mathcal{X}} |\pi(A) - \mu(A)| = \frac{1}{2} \sum_{x \in \mathcal{X}} |\pi(x) - \mu(x)|.$$

However, the techniques used in this paper yield bounds on the ℓ^2 distance

$$\|(P_x^\ell / \pi) - 1\|_{2, \pi} = \left(\sum_{y \in \mathcal{X}} \left| \frac{P_x^\ell(y)}{\pi(y)} - 1 \right|^2 \pi(y) \right)^{1/2}$$

which dominates $\|P_x^\ell - \pi\|_{TV}$. As explained in [10], ℓ^2 bounds are equivalent to bounds on

$$\sup_y \left| \frac{P_x^\ell(y)}{\pi(y)} - 1 \right|.$$

Denote by

$$\beta_0(P) = 1 \geq \beta_1(P) \geq \dots \geq \beta_{|\mathcal{X}|-1}(P) = \beta_{\min}(P) \geq -1$$

the eigenvalues of the chain P and let

$$\beta(P) = \max \{ \beta_1(P), -\beta_{\min}(P) \}.$$

Using (1.2) (1.3), the second largest eigenvalue $\beta_1(P)$ can be expressed as

$$1 - \beta_1(P) = \min \left\{ \frac{\mathcal{E}_P(f, f)}{\text{Var}_\pi(f)} : f \neq 0 \right\}.$$

A classical and easy bound (e.g., [12, 20]) on variation distance is given by

$$\|P_x^\ell - \pi\|_{TV} \leq \frac{1}{2\sqrt{\pi(x)}} \beta^\ell. \tag{1.5}$$

Recall that the log-Sobolev constant $\alpha(P)$ of a reversible Markov chain (P, π) is defined as the largest non-negative number α such that

$$\alpha \mathcal{L}_\pi(f) \leq \mathcal{E}_P(f, f) \tag{1.6}$$

for any function f . We will use $\alpha(P)$ to prove mixing rates that improve upon those obtained through (1.5). More precisely, we will use the following

Theorem 1.1 *Let (P, π) be a (finite) reversible Markov chain. Then*

$$\|P_x^\ell - \pi\|_{TV} \leq 2e^{-c} \text{ for } \ell \geq 1 + \frac{c}{1 - \beta} + \frac{1}{4\alpha} \log \log \frac{1}{\pi(x)}, \quad c > 0.$$

This theorem shows that, starting at x , the chain is close to its stationary distribution after order $\frac{1}{4\alpha} \log \log \frac{1}{\pi(x)}$ steps whereas (1.5) shows that $\frac{1}{2\sqrt{\pi(x)}} \log \frac{1}{\pi(x)}$ is enough. Thus, whenever $\frac{1}{\alpha}$ is roughly of the same order as $\frac{1}{-\log(\beta)}$, Theorem 1.1 improves upon (1.5) (recall that when β is close to one, $-\log \beta \sim 1 - \beta$ and it is this quantity that is often used in practice). We refer the reader to [10] for the proof of Theorem 1.1 and a discussion of the use of log-Sobolev inequalities for finite Markov chains. The present paper illustrates this technique with non-trivial examples.

2. The binary case

This section treats the case where $G = \mathbb{Z}_2$. We start with a simple lemma which determines the state space in this case.

Lemma 2.1 *Let \mathcal{G} be a connected graph with vertex set \mathcal{V} and edge set \mathcal{A} . The random walk on $\mathcal{X} = \mathbb{Z}_2^{\mathcal{V}} \setminus \{0\}$ defined at (1.1) is symmetric, connected, aperiodic and has the uniform distribution $\pi(x) = (2^{|\mathcal{V}|} - 1)^{-1}$ as reversible measure.*

Proof: We only have to show that the walk is irreducible and aperiodic since we already mentioned that it is symmetric. The fact that it is irreducible can be seen by observing that any state $x \in \mathcal{X}$ can be transformed, in at most $|\mathcal{V}|$ elementary steps, into the state $x_* \equiv 1 \in \mathcal{X}$ (i.e., all the labels are 1). Since there is always some holding, the chain is aperiodic. Lemma 3.2 gives extensions of Lemma 2.1 to general groups.

The special case where \mathcal{G} is the complete graph on n vertices will be used as the basis for a comparison argument. This special case itself will be studied via comparison with the familiar walk on \mathbb{Z}_2^n where $n = |\mathcal{V}|$.

2.A The complete graph

Let \mathcal{G} be the complete graph on n vertices, $n > 1$. Thus, $\mathcal{A} = \{\mathcal{V} \times \mathcal{V} \setminus \{(v, v) : v \in \mathcal{V}\}\}$. For $x \in \mathcal{X} = \mathbb{Z}_2^n \setminus \{0\}$, let $|x|$ be the number of 1's in x . The chain defined at (1.1) is given by

$$K(x, y) = \begin{cases} 0 & \text{if } \sum |x(v) - y(v)| > 1 \\ \frac{|x|}{n(n-1)} & \text{if } \exists u_0 \in \mathcal{V}, \forall u \neq u_0, x(u) = y(u); x(u_0) = 0, y(u_0) = 1 \\ \frac{|x|-1}{n(n-1)} & \text{if } \exists u_0 \in \mathcal{V}, \forall u \neq u_0, x(u) = y(u); x(u_0) = 1, y(u_0) = 0 \\ \frac{n-|x|}{n} & \text{if } x = y. \end{cases} \tag{2.1}$$

Theorem 2.2 *The chain K at (2.1) satisfies*

$$\beta_1(K) \leq 1 - \frac{2}{n(n+1)}, \quad \beta_{\min}(K) \geq -1 + \frac{2}{n+3}, \quad \alpha(K) \geq \frac{1}{n(n+1)}.$$

Further,

$$\|K_x^\ell - \pi\|_{TV} \leq 2e^{-c} \text{ for } \ell \geq 1 + \frac{n(n+1)}{2} \left(\frac{\log n}{2} + c \right), \quad c > 0.$$

The last statement in this theorem shows that the chain K is close to its stationary distribution after order $n^2 \log n$ steps. The best lower bound we know is that order $n \log n$ steps are required. We conjecture that this is the right answer.

Theorem 2.2 is proved following the introduction of a comparison chain. Let $\mathcal{L} = \mathbb{Z}_2^n$ be the hypercube of dimension $n = |\mathcal{V}|$. Let Q denote the familiar chain on \mathcal{L} which proceeds by adding 1 mod (2) to a coordinate chosen uniformly at random. Thus

$$Q(x, y) = \begin{cases} \frac{1}{n} & \text{if } |x - y| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

This has the uniform distribution $\mu(x) = 2^{-n}$ as reversible measure.

The following notation is crucial for the next proposition which is the heart of our argument. To any function f defined on $\mathcal{X} = \mathbb{Z}_2^{\mathcal{Z}} \setminus \{0\}$, we associate the function \tilde{f} defined on $\mathcal{L} = \mathbb{Z}_2^{\mathcal{Z}}$ by

$$\tilde{f}(v) = \begin{cases} f(v) & \text{if } v \neq 0 \\ \frac{1}{n} \sum_{|y|=1} f(y) & \text{if } v = 0. \end{cases} \tag{2.2}$$

The function \tilde{f} extends to \mathcal{L} the function f initially defined on \mathcal{X} . This extension will be used to compare a Dirichlet form on \mathcal{X} with a Dirichlet form on \mathcal{L} .

Proposition 2.3 *The chain (K, π) on $\mathcal{X} = \mathbb{Z}_2^{\mathcal{Z}} \setminus \{0\}$ defined at (2.1) and the chain (Q, μ) on $\mathcal{L} = \mathbb{Z}_2^{\mathcal{Z}}$ satisfy*

$$\text{Var}_{\pi}(f) \leq \frac{2^n}{2^n - 1} \text{Var}_{\mu}(\tilde{f}), \tag{2.3}$$

$$\mathcal{L}_{\pi}(f) \leq \frac{2^n}{2^n - 1} \mathcal{L}_{\mu}(\tilde{f}) \tag{2.4}$$

and

$$\mathcal{E}_Q(\tilde{f}, \tilde{f}) \leq \frac{(2^n - 1)(n + 1)}{2^n} \mathcal{E}_K(f, f). \tag{2.5}$$

Proof: The first stated inequality is easy if one recalls that

$$\text{Var}_{\pi}(f) = \inf_{c \in \mathbb{R}} \sum_{x \in \mathcal{X}} |f(x) - c|^2 \pi(x).$$

The second is obtained by using a similar trick due to Holley and Stroock [17]: observe that $\xi \log \xi - \xi \log \zeta - \xi + \zeta \geq 0$ for all $\xi, \zeta > 0$ and that

$$\mathcal{L}_{\pi}(f) = \inf_{c > 0} \sum_x (|f(x)|^2 \log |f(x)|^2 - |f(x)|^2 \log c - |f(x)|^2 + c) \pi(x).$$

We now proceed with the proof of (2.5). Let f be a function on \mathcal{X} and \tilde{f} be the function on \mathcal{L} defined at (2.2). Then, write

$$\begin{aligned} \mathcal{E}_Q(\tilde{f}, \tilde{f}) &= \frac{1}{2^n n} \sum_{\substack{(x,y) \in \mathcal{L} \times \mathcal{L} \\ |x-y|=1, |x| < |y|}} |\tilde{f}(x) - \tilde{f}(y)|^2 \\ &= \frac{1}{2^n n} \sum_{\substack{(x,y) \in \mathcal{X} \times \mathcal{X} \\ |x-y|=1, |x| < |y|}} |f(x) - f(y)|^2 + \frac{1}{2^n n} \sum_{|y|=1} |f(y) - \tilde{f}(0)|^2. \end{aligned} \tag{2.6}$$

For the second term, call it R , use

$$\frac{1}{n} \sum_{|y|=1} |f(y) - \tilde{f}(0)|^2 = \frac{1}{2n^2} \sum_{\substack{|x|=|y|=1 \\ x \neq y}} |f(x) - f(y)|^2$$

to write

$$\begin{aligned} R &= \frac{1}{2^n n} \sum_{|x|=1} |f(x) - \tilde{f}(0)|^2 \\ &= \frac{1}{2^{n+1} n^2} \sum_{\substack{|x|=1, |y|=1 \\ x \neq y}} |f(x) - f(y)|^2. \end{aligned}$$

Now, use $\frac{1}{2}(a - b)^2 \leq (a - c)^2 + (c - b)^2$ to get

$$\begin{aligned} R &\leq \frac{1}{2^n n^2} \sum_{\substack{|x|=1, |y|=1 \\ x \neq y}} (|f(x) - f(x+y)|^2 + |f(x+y) - f(y)|^2) \\ &= \frac{2}{2^n n^2} \sum_{\substack{|x|=1, |z|=2 \\ |x-z|=1}} |f(x) - f(z)|^2. \end{aligned} \tag{2.7}$$

Putting together (2.6) and (2.7) gives

$$\begin{aligned} \mathcal{E}_Q(\tilde{f}, \tilde{f}) &\leq \frac{1}{2^n n} \sum_{\substack{(x,y) \in \mathcal{X} \times \mathcal{X} \\ |x| < |y|, |x-y|=1}} |f(x) - f(y)|^2 + \frac{2}{2^n n^2} \sum_{\substack{|x|=1, |z|=2 \\ |x-z|=1}} |f(x) - f(z)|^2 \\ &= \frac{1}{2^n n} \left(\sum_{\substack{(x,y) \in \mathcal{X} \times \mathcal{X}, \\ 2 \leq |x| < |y|, |x-y|=1}} |f(x) - f(y)|^2 + \frac{n+2}{n} \sum_{\substack{|x|=1, |y|=2 \\ |x-y|=1}} |f(x) - f(y)|^2 \right) \\ &\leq \frac{n+2}{2^n n^2} \sum_{\substack{(x,y) \in \mathcal{X} \times \mathcal{X} \\ |x| < |y|, |x-y|=1}} |f(x) - f(y)|^2. \end{aligned}$$

From (2.1) $\pi(x)K(x, y) \geq [(2^n - 1)n(n - 1)]^{-1}$ and thus

$$\begin{aligned} \mathcal{E}_Q(\tilde{f}, \tilde{f}) &\leq \frac{(2^n - 1)(n + 2)(n - 1)}{2^{n+1} n} \sum_{(x,y) \in \mathcal{X} \times \mathcal{X}} |f(x) - f(y)|^2 K(x, y) \pi(x) \\ &\leq \frac{(2^n - 1)(n + 1)}{2^n} \mathcal{E}_K(f, f). \end{aligned}$$

Proof of Theorem 2.2: The upper bound on β_1 . The first eigenvalue of the random walk Q on \mathcal{X} is easily computed to be $1 - 2/n$ (see [6]). This translates into the Poincaré inequality

$$\text{Var}_\mu(h) \leq \frac{n}{2} \mathcal{E}_Q(h, h)$$

for any function h on \mathcal{X} . Thus, (2.3) and (2.5) in Proposition 2.3 yield

$$\text{Var}_\pi(f) \leq \frac{n(n + 1)}{2} \mathcal{E}_K(f, f).$$

This gives the desired bound on $\beta_1(K)$ by the variational characterization of eigenvalues.

Remark: For $1 \leq j \leq n$, set $m_j = \sum_{i=1}^j \binom{n}{i}$. Then the eigenvalues $\beta_i(Q)$ of Q , in decreasing order, are given by

$$\beta_0(Q) = 1, \quad \beta_i(Q) = 1 - \frac{2j}{n} \quad \text{for } m_j \leq i < m_{j+1},$$

where j varies from 1 to n . See, e.g., [5, 6]. It follows from the minimax characterization of eigenvalues and from Proposition 2.3 that the eigenvalues $\beta_i(K)$ of K satisfy

$$\beta_i(K) \leq 1 - \frac{2j}{n(n+1)} \quad \text{for } m_j \leq i < m_{j+1},$$

with $1 \leq j \leq n$.

Proof of Theorem 2.2: The lower bound on β_{\min} . We will use a slight variation on Proposition 2 of [12], page 40.

Suppose (K, π) is a reversible Markov chain on the finite state space \mathcal{X} . For each $x \in \mathcal{X}$, let Σ_x be some fixed set of cycles at x of odd length. Let $\Sigma = \cup_x \Sigma_x$. For each cycle $\sigma \in \Sigma$, let $|\sigma|$ be its length. Finally, let θ be a non-negative function defined on Σ and such that, for each $x \in \mathcal{X}$,

$$\sum_{\sigma \in \Sigma_x} \theta(\sigma) = \pi(x).$$

Such a function θ is called a flow on odd cycles (we will later encounter other kinds of flows for comparison between two chains). Then, the argument in [12], page 40, easily gives

Lemma 2.4 *With the above notation, for any finite reversible Markov chain and any flow θ on odd cycles,*

$$\beta_{\min}(K) \geq -1 + \frac{2}{I(\theta)}$$

where

$$I(\theta) = \max_{\substack{(x,y) \\ K(x,y) > 0}} \left(\frac{1}{K(x,y)\pi(x)} \sum_{\substack{\sigma \in \Sigma \\ \sigma \ni (x,y)}} r(\sigma, (x,y)) |\sigma| \theta(\sigma) \right).$$

Here, $r(\sigma, (x,y))$ is the number of times the edge (x,y) is used in σ (one can always assume that $r(\sigma, (x,y)) \leq 2$ and, in our applications, it will always be at most 1).

To apply this to the case at hand, consider the following cycles. If $x \in \mathcal{X}$ has $|x| \neq n$, let σ_x be the trivial loop (x,x) at x , $\Sigma_x = \{\sigma_x\}$ and set $\theta(\sigma_x) = \pi(x)$. If $x = x_* \equiv 1$, let x_v be the labeling with a single zero at v and let σ_v be the loop x_*, x_v, x_v, x_* of length 3. There are n distinct such cycles which form Σ_{x_*} . For any $\sigma \in \Sigma_{x_*}$, set $\theta(\sigma) = \pi(x_*)/n$. To estimate $I(\theta)$, we look at the different possible cases where

$$I(\theta, x, y) = \frac{1}{K(x, y)\pi(x)} \sum_{\sigma \ni (x, y)} |\sigma| \theta(\sigma)$$

is non-zero. First, pick an edge (z, z) with $|z| \leq n - 2$. Then $K(z, z) = (|z| - n)/n \geq 2/n$ and $\sum_{\sigma \ni (z, z)} |\sigma| \theta(\sigma) = \pi(z)$ (the sum contains only one term!). Thus $I(\theta, z, z)$ is bounded by $n/2$ in this case. Second, let $|z| = n - 1$. Then $K(z, z) \geq 1/n$ and $\sum_{\sigma \ni (z, z)} |\sigma| \theta(\sigma) = (1 + 3/n)\pi(z)$. Thus, $I(\theta, z, z)$ is bounded by $n + 3$ in this case. Third, pick $v \in \mathcal{V}$ and (x_*, x_v) [or (x_v, x_*)]. Then, $K(x_*, x_v) = 1/n$, $\sum_{\sigma \ni (x_*, x_v)} |\sigma| \theta(\sigma) = 3\pi(x_*)/n$ and $I(\theta, x_*, x_v) = 3$. This gives the desired bound on $\beta_{\min}(K)$.

The bounds on $\beta_1(K)$ and $\beta_{\min}(K)$ together with (1.5) show that the total variation distance between K_x^ℓ and π is small after order n^3 steps. The bounds on higher eigenvalues do not provide further improvement on this estimate because of the lack of invariance of the chain (see [8], Sect. 6). Instead, we will use the log-Sobolev constant $\alpha(K)$.

End of the proof of Theorem 2.2: The lower bound on $\alpha(K)$. The log-Sobolev constant of the chain (Q, μ) is $\alpha(Q) = 1/n$, see [15]. Thus, (2.4) and (2.5) in Proposition 2.3 imply that

$$\alpha(K) \geq 1/[n(n + 1)].$$

The last and main statement in Theorem 2.2 now follows if we use the bounds on $\beta_1(K)$, $\beta_{\min}(K)$ and $\alpha(K)$ in Theorem 1.1.

2.B The general binary case

We now consider the case of the binary labeling of a general graph \mathcal{G} . This will be studied by comparison with the special case of the complete graph. Let us fix notation. We let \mathcal{V} be the vertex set of the finite graph \mathcal{G} with oriented edge set \mathcal{A} . We assume that there are no self-loops, that \mathcal{A} is symmetric, and that \mathcal{G} is connected. For each $(u, v) \in \mathcal{V} \times \mathcal{V}$, we fix a path $\gamma_{u, v}$ joining u to v in \mathcal{G} and let $|\gamma_{u, v}|$ be the length of this path. We set

$$\Delta = \max_{a \in \mathcal{A}} \left(\sum_{\substack{u, v \\ \gamma_{u, v} \ni a}} |\gamma_{u, v}| \right). \tag{2.8}$$

In the binary case, the chain defined at (1.1) on $\mathcal{B} = \mathbb{Z}_2^{\mathcal{V}} \setminus \{0\}$ is given by

$$P(x, y) = \begin{cases} \frac{N_{\mathcal{G}}(x, y)}{|\mathcal{B}|} & \text{if } x \text{ and } y \text{ differ at exactly one vertex} \\ \frac{\sum_{v: x(v)=0} \#\{u \sim v\}}{|\mathcal{B}|} & \text{if } x = y \\ 0 & \text{otherwise .} \end{cases} \tag{2.9}$$

Here, for x and y that differ at exactly one vertex, say u , $N_{\mathcal{G}}(x, y)$ is defined by

$$N_{\mathcal{G}}(x, y) = \text{the number of neighbors } v \text{ of } u \text{ such that } x(v) = 1. \quad (2.10)$$

The stationary distribution is $\pi(x) = (2^n - 1)^{-1}$.

This chain will be studied by comparison with the chain K defined at (2.1) which corresponds to the complete graph.

Theorem 2.5 *The chain (P, π) at (2.9) has second largest eigenvalue bounded by*

$$\beta_1(P) \leq 1 - \frac{n - 1}{4(n + 1)|\mathcal{A}|\Delta}.$$

Its log-Sobolev constant satisfies

$$\alpha(P) \geq \frac{n - 1}{8(n + 1)|\mathcal{A}|\Delta}.$$

Further the smallest eigenvalue satisfies

$$\beta_{\min} \geq -1 + \frac{2nd_0}{|\mathcal{A}|(n + 3)}$$

where d_0 is a lower bound on the degree of any vertex in \mathcal{G} . Finally, it follows from the above estimates that

$$\|P_x^\ell - \pi\|_{TV} \leq 2e^{-c} \text{ for } \ell \geq 1 + \frac{4|\mathcal{A}|\Delta(n + 1)}{n - 1} \left(\frac{\log n}{2} + c \right), \quad c > 0.$$

Before embarking with the proof, let us describe a few specific examples.

Example 2a: Let \mathcal{G} be the n point path $1 \cdots n$. In this case, $|\mathcal{A}| = 2(n - 1)$ and Δ can be bounded by $\Delta \leq n^3/8$. This gives

Corollary 2.6 *When \mathcal{G} is the n point path, the chain P at (2.9) satisfies*

$$\beta_1(P) \leq 1 - \frac{1}{(n + 1)n^3}, \quad \alpha(P) \geq \frac{1}{2(n + 1)n^3},$$

and

$$\|P_x^\ell - \pi\|_{TV} \leq 2e^{-c} \text{ for } \ell \geq 1 + (n + 1)n^3 \left(\frac{\log n}{2} + c \right), \quad c > 0.$$

Example 2b: Let \mathcal{G} be a square grid with sides of length about \sqrt{n} . Then, $|\mathcal{A}|$ is of order n , Δ of order n^2 (for some reasonable choice of paths, see, e.g., [8]). In this case, the upper bound on the relaxation time $\tau = (1 - \beta_1(P))^{-1}$ is of order n^3 . The chain is close to equilibrium after order $n^3 \log n$ steps.

Example 2c: For any graph on n vertices, we have the universal bounds $|\mathcal{A}| \leq n(n - 1)$, $\Delta \leq n^3$. This gives an upper bound of order n^5 for the relaxation time and shows that the chain is close to equilibrium after order $n^5 \log n$ steps. We do not know any examples where this is matched by a lower bound.

For the comparison argument, we will use Theorem 2.3 of [8], which we now recall. Suppose we have two reversible irreducible chains, say (K, μ) and (P, π) on the same state space \mathcal{B} . For each pair $(x, y) \in \mathcal{B} \times \mathcal{B}$ with $x \neq y$ and $K(x, y) > 0$, let $\mathcal{S}_{x,y}$ be some fixed set of paths $x_0 = x, x_1, \dots, x_k = y$ joining x to y and such that $P(x_i, x_{i+1}) > 0$. Set

$$\mathcal{S} = \bigcup_{\substack{(x,y): x \neq y \\ K(x,y) > 0}} \mathcal{S}_{x,y}.$$

By definition, a (P, K) -flow is a non-negative function η defined on \mathcal{S} and such that

$$\sum_{\gamma \in \mathcal{S}_{x,y}} \eta(\gamma) = K(x, y)\mu(x),$$

for all $(x, y) \in \mathcal{B} \times \mathcal{B}$ such that $x \neq y$ and $K(x, y) > 0$.

Theorem 2.7 ([8]) *Given a (P, K) -flow η , the Dirichlet forms \mathcal{E}_K and \mathcal{E}_P satisfy*

$$\mathcal{E}_K \leq A(\eta)\mathcal{E}_P \tag{2.11}$$

with

$$A(\eta) = \max_{\substack{(x,y) \\ P(x,y) > 0}} \left(\frac{1}{P(x, y)\pi(x)} \sum_{\gamma \in \mathcal{S}(x,y)} |\gamma| \eta(\gamma) \right). \tag{2.12}$$

Here $|\gamma|$ denote the length of the path γ and $\mathcal{S}(x, y)$ is the subset of those paths in \mathcal{S} that contain the edge (x, y) .

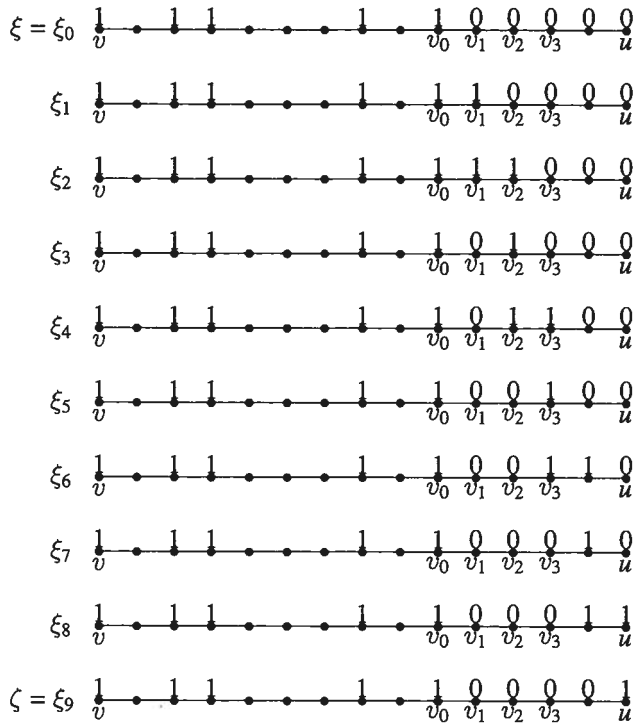
Warning: $\mathcal{S}_{x,y}$ and $\mathcal{S}(x, y)$ are two different subsets of \mathcal{S} . The former is defined for each pair (x, y) such that $K(x, y) > 0$, whereas the latter is defined for each pair (x, y) such that $P(x, y) > 0$.

Proof of Theorem 2.5: We now proceed to define a (P, K) -flow for the chain K at (2.1) and P at (2.9). First, we describe a set $\mathcal{S}_{\xi, \zeta}$ of admissible paths joining ξ to ζ when $K(\xi, \zeta) > 0$ (and $\xi \neq \zeta$).

If $\xi \neq \zeta$ and $K(\xi, \zeta) > 0$, there exists $u \in \mathcal{V}$ such that $\xi(u) \neq \zeta(u)$ and $\xi(v) = \zeta(v)$ for all $v \neq u$. Moreover, there exists at least one $v \neq u$ such that $\xi(v) = 1$. We first treat the case when

$$\xi(u) = 0, \quad \zeta(u) = 1. \tag{2.13}$$

Let $I(\xi, \zeta)$ be the set of all vertices $v \in \mathcal{V}$, $v \neq u$ such that $\xi(v) = \zeta(v) = 1$. For each $v \in I(\xi, \zeta)$, we are given a path $\gamma_{v,u}$ joining v to u in the underlying graph \mathcal{G} . Let v_0 be the vertex on $\gamma_{v,u}$ which is closest to u and has $\xi(v_0) = 1$. Now define a path from ξ to ζ as follows: let $v_0, v_1, v_2, \dots, v_k = u$ be the vertices on the path toward u , starting at v_0 . First create a 1 at v_1 , then a 1 at v_2 , erase the 1 at v_1 , create a 1 at v_3 , erase the 1 at v_2 , etc, until a 1 is created at $v_k = u$ and the 1 at v_{k-1} is erased. The resulting labeling is ζ . Note that the last step requires (2.13). The following picture illustrates this construction. All unlabeled vertices are 0. The picture shows only the vertices that are on the path $\gamma_{v,u}$.



For each $v \in I(\xi, \zeta)$, and under the extra hypothesis that (2.13) is satisfied, this describes a path $\xi_0 = \xi, \xi_1, \dots, \xi_k = \zeta$ from ξ to ζ . Call this path $\gamma_{\xi, \zeta}(v)$. We will say that $\gamma_{\xi, \zeta}(v)$ is **built on** the underlying path $\gamma_{v, u}$. Observe that $|\gamma_{\xi, \zeta}(v)| \leq 2|\gamma_{v, u}|$.

If (2.13) is not satisfied, i.e. if $\xi(u) = 1, \zeta(u) = 0$, then the above description give us a path $\gamma_{\zeta, \xi}(v)$ from ζ to ξ for each $v \in I(\zeta, \xi) = I(\xi, \zeta)$. Now, we get a path from ξ to ζ that we call $\gamma_{\xi, \zeta}(v)$ by moving along $\gamma_{\zeta, \xi}(v)$ backwards.

Let us pause here to look at what has been achieved so far: for each pair (ξ, ζ) with $\xi \neq \zeta$ and $K(\xi, \zeta) > 0$, we have built a set of paths $\mathcal{S}_{\xi, \zeta}$ indexed, with repetitions, by the set $I(\xi, \zeta)$ of those vertices v such that $\xi(v) = \zeta(v) = 1$. Observe indeed that two different $v_1, v_2 \in I(\xi, \zeta)$ can give rise to the same path $\gamma = \gamma_{\xi, \zeta}(v_1) = \gamma_{\xi, \zeta}(v_2)$.

We define a (P, K) -flow η on the set of all these paths by setting, for any path $\gamma \in \mathcal{S}_{\xi, \zeta}$,

$$\eta(\gamma) = \frac{\#\{v \in I(\xi, \zeta) : \gamma = \gamma_{\xi, \zeta}(v)\}}{\min(|\xi|, |\zeta|)} K(\xi, \zeta) \pi(\xi). \tag{2.14}$$

Now, we are left with the task of bounding the constant $A(\eta)$ at (2.12).

Proposition 2.8 *For the (P, K) -flow η defined at (2.14), the comparison constant $A(\eta)$ satisfies*

$$A(\eta) \leq \frac{8|\mathcal{A}|\Delta}{n(n-1)}.$$

Here $n = |\mathcal{V}|$ is the number of vertices in the underlying graph \mathcal{G} , $|\mathcal{A}|$ is the number of oriented edges in \mathcal{G} , and Δ is defined at (2.8).

Proof: For the flow at (2.14) and with the parametrization of the paths in \mathcal{S} by ξ, ζ such that $K(\xi, \zeta) > 0$ and $v \in I(\xi, \zeta)$, the constant $A(\eta)$ of (2.12) becomes

$$A(\eta) = \frac{|\mathcal{A}|}{n(n-1)} \max_{\substack{(x,y) \\ P(x,y) > 0}} \left(\frac{1}{N_{\mathcal{G}}(x,y)} \sum_{\substack{\xi, \zeta, v \\ \gamma_{\xi, \zeta}(v) \ni (x,y)}} |\gamma_{\xi, \zeta}(v)| \right).$$

Here the sum runs over all ξ, ζ such that $K(\xi, \zeta) > 0$ and all $v \in I(\xi, \zeta)$.

Let us fix the ordered pair (x, y) such that $P(x, y) > 0$ and set

$$F = F(x, y) = \sum_{\substack{\xi, \zeta, v \\ \gamma_{\xi, \zeta}(v) \ni (x,y)}} |\gamma_{\xi, \zeta}(v)|.$$

We want to estimate F . Let w be the one and only vertex where x and y differ. Now, any path $\gamma_{\xi, \zeta}(v)$ that appears in S must be supported by an underlying path $\gamma_{v,u}$ in the underlying graph \mathcal{G} . This path $\gamma_{v,u}$ must go through w . Moreover, there must exist at least one neighbor w' of w in \mathcal{G} such that $x(w') = 1$ and (w, w') or (w', w) is an edge in $\gamma_{v,u}$. Finally, the path $\gamma_{\xi, \zeta}(v)$ has length at most $2|\gamma_{v,u}|$.

Conversely, we claim that, given an ordered pair (v, u) such that $\gamma_{v,u}$ contains w and a neighbor w' of w satisfying $x(w') = 1$, there are **at most two** (ξ, ζ) such that $\gamma_{\xi, \zeta}(v) \ni (x, y)$ with underlying path $\gamma_{v,u}$. Taking this claim for granted, we obtain

$$\begin{aligned} F &\leq 4 \sum_{\substack{w' \sim w \\ x(w')=1}} \left(\sum_{\substack{v,u \\ \gamma_{v,u} \ni (w,w')}} |\gamma_{v,u}| + \sum_{\substack{v,u \\ \gamma_{v,u} \ni (w',w)}} |\gamma_{v,u}| \right) \\ &\leq 8N_{\mathcal{G}}(x, y)\Delta, \end{aligned}$$

where $N_{\mathcal{G}}(x, y)$ is defined at (2.10). Proposition 2.8 follows.

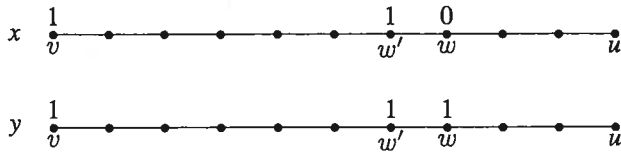
We must still prove the claim. Roughly speaking, the claim follows because, given (x, y) and $\gamma_{v,u}$, the two unknown configurations ξ and ζ can differ from x only at u , at w , and at one of the two neighbors of w along $\gamma_{v,u}$. This certainly offers only a finite number of possibilities. Observe that this is true only because of the way we constructed $\gamma_{\xi, \zeta}(v)$ out of $\gamma_{v,u}$. To show that this finite number is 2, we have to be more careful.

Proof of the claim: We are given x and y that differ only at some $w \in \mathcal{V}$. We are also given (v, u) such that $\gamma_{v,u} \ni w$ and a neighbor w' of w such that $x(w) = 1$ and (w, w') or (w', w) belongs to $\gamma_{v,u}$. We are looking for all the (ξ, ζ) with $K(\xi, \zeta) > 0$ such that $\gamma_{\xi, \zeta}(v)$ is built on $\gamma_{v,u}$.

First, observe that there are obvious necessary conditions that any such ordered pair (ξ, ζ) must satisfy. These are: “ $x(v) = \xi(v) = 1$ ” as well as “ ξ and ζ differ at u and only at u ”.

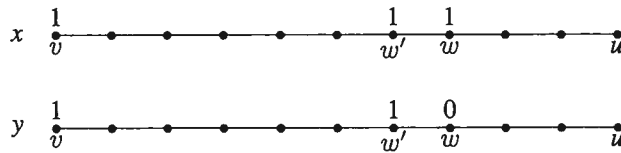
Case 1 Consider the case where $(w', w) \in \gamma_{v,u}$.

Case 1.1 Assume that $x(w) = 0$.



Then, we are moving 1 towards u . This can only happen if $\xi(u) = 0, \zeta(u) = 1$ (Observe also that this implies that $x(t) = 0$ for any vertex t following w along the path $\gamma_{v,u}$). We are left with only two possibilities. Either $\xi(w') = 1$, and this was the very first step on the path $\gamma_{\xi,\zeta}(v)$ that we can now reconstruct completely from the known data. Or $\xi(w') = 0$. Then, the first vertex s towards v such that $x(s) = 1$ indicates where we started moving 1 towards u and we can reconstruct the path from what we know.

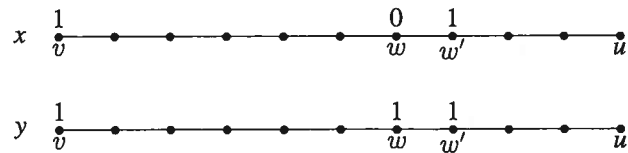
Case 1.2 Assume that $x(w) = 1$.



Then we are moving 1 towards v . This implies that $\xi(u) = 1, \zeta(u) = 0$. (Observe again that this implies $x(t) = 0$ for any vertex t following w along the path $\gamma_{v,u}$). We are left again with only two possibilities. Either $\xi(w') = 1$, and this was the very last step on the path $\gamma_{\xi,\zeta}(v)$. Or $\xi(w') = 0$, and we must keep moving the label 1 towards v until we reach the first 1 to the left of w' . In both cases, we can easily reconstruct the path $\gamma_{\xi,\zeta}(v)$ from what we know.

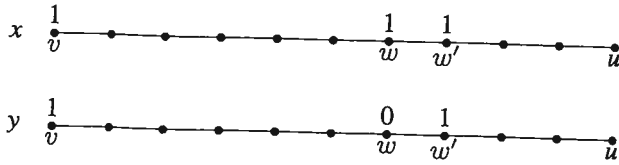
Case 2 Consider now the case where $(w, w') \in \gamma_{v,u}$.

Case 2.1 Assume that $x(w) = 0$.



Then, we are moving 1 towards v . Thus, $\xi(u) = 1, \zeta(u) = 0$. Either $\xi(w) = 1$ and there is only one last step to be done in order to reach ζ , namely, erasing the 1 at w' . Or $\xi(w) = 0$, and we must keep moving 1 towards v until we reach the first label 1 on our way. Again this lets us find ξ and ζ .

Case 2.2 Assume that $x(w) = 1$.



Then, we are moving 1 towards u . There is only one thing to do: continue to move the 1 along the path to find ζ and ξ !

This case by case study proves the claim and finishes the proof of Proposition 2.8.

The first two statements in Theorem 2.5 follow from Proposition 2.8 and the results previously obtained for the complete graph. The lower bound on the smallest eigenvalue is given by the following lemma.

Lemma 2.9 *The smallest eigenvalue of the chain P at (2.9) satisfies*

$$\beta_{\min} \geq -1 + \frac{2nd_0}{|\mathcal{A}|(n+3)}.$$

Here, d_0 is the smallest degree of a vertex in the underlying graph \mathcal{G}

Proof: We use Lemma 2.4. We also use the same cycles σ and the same “flow” θ as in the proof of the lower bound on β_{\min} in Theorem 2.2 (the case of the complete graph). We refer the reader to the lines following Lemma 2.4 for unexplained notation. We have to bound

$$\frac{1}{P(x,y)\pi(x)} \sum_{\sigma \ni (x,y)} |\sigma|\theta(\sigma)$$

for $x = y \neq x_*$, and for (x_*, x_v) and (x_v, x_*) . Let d_0 be the smallest degree of a vertex in the underlying graph. First, pick z with $|z| \leq n - 2$. Then, $P(x, y) \geq 2d_0/|\mathcal{A}|$ and $\sum_{\sigma \ni (z,z)} |\sigma|\theta(\sigma) = \pi(z)$. Next, let $|z| = n - 1$. Then $P(x, y) \geq d_0/|\mathcal{A}|$ and $\sum_{\sigma \ni (z,z)} |\sigma|\theta(\sigma) = (1 + 3/n)\pi(z)$. Finally, for (x_*, x_v) [or (x_v, x_*)], $P(x_*, x_v) \geq d_0/|\mathcal{A}|$ and $\sum_{\sigma \ni (x_*, x_v)} |\sigma|\theta(\sigma) \leq \pi(x_*)/n$.

To conclude the proof of Theorem 2.5 and obtain the convergence in total variation, we just have to invoke the estimates on β_1, β_{\min} and α and use these in Theorem 1.1.

3. Cyclic p -groups

This section explores the comparison argument of Sect. 2 for groups other than \mathbb{Z}_2 . We mainly focus on the case where \mathcal{G} is the complete graph on n vertices and $G = \mathbb{Z}_q$ with $q = p^a$, p prime. For large n and fixed q , we extend the results obtained in the binary case. More precisely, we prove

Theorem 3.1 *Let \mathcal{G} be the complete graph on n vertices. If $G = \mathbb{Z}_q$ with $q = p^a$, p prime, the walk K defined at (1.1) satisfies*

$$\|K_x^\ell - \pi\|_{TV} \leq 2e^{-c} \text{ for } \ell \geq 1 + 6n^2q^2(\log(q-1))(\log \log q^n) + 8n^2q^2c, \quad c > 0.$$

This shows that the chain reaches stationarity after order $n^2 \log n$ steps for any fixed $q = p^a$. Thus, for $G = \mathbb{Z}_p$ with $p = 2, 3, 4, 5, 7, 8, 9$, $n^2 \log n$ steps are enough. For $G = \mathbb{Z}_6$ or $G = \mathbb{Z}_{10}$, we also know that $n^2 \log n$ steps are enough but the argument is much more involved. See [11].

Let us introduce some notation. Fix a finite group G and set $\mathcal{L} = G^n$. For x, y in $\mathcal{L} \times \mathcal{L}$, write

$$x \sim y \text{ if } x \text{ and } y \text{ differ exactly in one coordinate,}$$

and write

$$x \approx y \text{ if } \begin{cases} x \text{ and } y \text{ differ exactly in one coordinate, say } x_i \neq y_i, \\ \text{and there exists } j \neq i \text{ such that } y_i = x_i x_j^{\pm 1}. \end{cases}$$

If $x \approx y$ with $x_i \neq y_i$, let

$$N(x, y) = \begin{cases} \text{the number of } j \text{ such that } x_i^{-1}y_i = x_j^{\pm 1} \text{ if } x_i^{-1}y_i \neq (x_i^{-1}y_j)^{-1} \\ \text{twice this number if } x_i^{-1}y_i = (x_i^{-1}y_i)^{-1}. \end{cases}$$

Finally, let $N(x)$ be the number of coordinates equal to the identity in x . With this notation the chain (1.1) on the complete graph on n vertices is given by

$$K(x, y) = \begin{cases} 0 & \text{if } x \not\approx y \text{ and } x \neq y \\ \frac{N(x, y)}{2n(n-1)} & \text{if } x \approx y \\ \frac{N(x)}{n} & \text{if } x = y. \end{cases} \tag{3.1}$$

In what follows, \mathcal{K} denotes the set of all generating n -tuples. Lemma 3.2 below shows that K is irreducible on \mathcal{K} with stationary measure $\pi(x) = |\mathcal{K}|^{-1}$ when n is large enough.

We now discuss the irreducibility of K . Let G be a finite group. Let $S \subset G$ be a set of generators. Say S is minimal if no smaller subset of S generates G . Define $\bar{m}(G)$ as the maximal size of a minimal generating set. If S is a generating set with $|S| = \bar{m}(G)$, then deleting successive elements of S results in a strictly decreasing sequence of subgroups. If $|G| = \prod p^{a_p}$ is a factorization of the size of G into distinct prime powers, we see

$$\bar{m}(G) \leq \sum_{p \mid |G|} a_p.$$

For example, let G be the symmetric group S_d . Then, for primes $p \leq d$, $a_p = [d/p] + [d/p^2] + \dots \leq d/[p(1 - 1/p)] \leq 2d/p$. Thus,

$$\bar{m}(S_d) \leq 2d \sum_{p \leq d} \frac{1}{p} \sim 2d \log \log d.$$

In fact, using results of Babai [1] and Cameron et al. [2], $\overline{m}(S_d) \leq 2d$. The classical generating set $S = (1, 2), (2, 3), \dots, (d - 1, d)$ shows $\overline{m}(S_d) \geq d - 1$.

Let $m(G)$ be the smallest size of a generating set of G . Thus, for the symmetric group S_d , $m(S_d) = 2$. For p -groups $m(G) = \overline{m}(G)$ and for cyclic p groups, $m(G) = \overline{m}(G) = 1$. The numbers $m(G), \overline{m}(G)$, appear in the following slight generalization of a result of Celler et al. [3] which gives a useful condition for the walk at (1.1) to be irreducible on the set of generating sequences.

Lemma 3.2 *Let G be a finite group. Then, the random walk (1.1) on a complete graph of $n \geq m(G) + \overline{m}(G)$ vertices gives an irreducible symmetric Markov chain on the set of n -tuples (x_1, \dots, x_n) which generate G . The stationary distribution is uniform.*

Proof: Let \mathcal{X} be the set of n -tuples of elements of G which generate G . Fix a generating sequence (y_1, \dots, y_m) of length $m = m(G)$. Any n -tuple (x_1, \dots, x_n) which generates G can be brought to $(y_1, y_2, \dots, y_m, \text{id}, \dots, \text{id})$. Indeed, a subsequence of length at most $\overline{m}(G)$ in (x_1, \dots, x_n) generates and so one can produce y_1, \dots, y_m in the complementary positions to this generating sequence. Using y_1, \dots, y_m , we can set all the remaining positions to id . Then, it is easy to order the y_i as we wish. This shows that the Markov chain (1.1) is irreducible on \mathcal{X} . Since it is symmetric and has some holding, it is ergodic with uniform stationary distribution.

Remarks: For some classes of groups, the conclusion of Lemma 3.2 holds for all $n \geq m(G) + 1$. Diaconis and Graham [7] show this for p -groups and for Abelian group. Note that often $m(G) \leq \overline{m}(G)$. For example, for \mathbb{Z}_{pq} , $m(G) = 1$, $\overline{m}(G) = 2$. Note that determining the size of the state space \mathcal{X} is a complicated problem (see [11]).

For the comparison argument we will use the chain Q on $\mathcal{X} = G^n$ defined by

$$Q(x, y) = \begin{cases} 0 & \text{if } x \not\sim y \text{ and } x \neq y \\ \frac{1}{n|G|} & \text{if } x \sim y \\ \frac{1}{|G|} & \text{if } x = y. \end{cases} \tag{3.2}$$

This chain picks a coordinate uniformly at random and multiplies this coordinate by a uniformly chosen element of G . Its stationary measure is $\mu(x) = |\mathcal{X}|^{-1} = |G|^{-n}$. It is a product chain with second largest eigenvalue

$$\beta_1(Q) = 1 - \frac{1}{n}.$$

Its log-Sobolev constant can be computed exactly using Lemma 3.2 and Corollary 5.5 in [10]. It is given by

$$\alpha(Q) = \frac{(|G| - 2)}{n|G| \log(|G| - 1)}.$$

We need a bound on the least eigenvalue of K .

Proposition 3.3 *The chain K at (3.1) has its least eigenvalue bounded by*

$$\beta_{\min}(K) \geq -1 + \frac{n - b}{n(n - 1)(|G|^2 - 1)} \geq -1 + \frac{1}{2n|G|^2}.$$

Here $n \geq 2\bar{m}(G)$.

Proof: To obtain the stated inequality we need a simple universal estimate. Let G be any finite group with generating set S . Write $|g|_S$ for the word length of $g \in G$ with respect to S ($|\text{id}|_S = 0$). We claim that

$$\sum_{g \in G} |g|_S \leq \frac{|G|(|G| - 1)}{2}. \tag{3.3}$$

Indeed, for each i , the existence of an element of length i implies the existence of at least one element of length $i - 1$. Thus, the worst case is when there is exactly one element of length i for each $0 \leq i \leq |G| - 1$ and $\sum_{i=0}^{|G|-1} i = |G|(|G| - 1)/2$.

Now, to prove Proposition 3.3, use Lemma 2.4 with the following flow θ on cycles: If one of the coordinates of x is the identity, set

$$\Sigma_x = \{\sigma_x\} \text{ with } \sigma_x = (x, x).$$

If none of the coordinates of x is the identity, fix a generating subset S occupying b coordinates $\{i_1, \dots, i_b\}$ of x and pick a coordinate, say x_i , not in this subset. Write x_i as a word using elements in S . This describes a path $\gamma_{x,i}$ from x to x^i where x^i is the n -tuple with i^{th} coordinate the identity and all other coordinates equal to those of x . Set

$$\Sigma_x = \{\sigma_{x,i} : i \notin \{i_1, \dots, i_b\}\}$$

where $\sigma_{x,i}$ is the cycle that goes from x to x^i along $\gamma_{x,i}$, holds at x^i for one step and goes back to x . Now, for any cycle σ , set

$$\theta(\sigma) = \begin{cases} \frac{1}{|\Sigma_x|} \pi(x) & \text{if } \sigma \in \Sigma_x \\ 0 & \text{otherwise.} \end{cases}$$

Observe that $|\Sigma_x| = n - b$ when none of the coordinates of x is the identity. Then, we have to bound

$$I(\theta) = \max_{\substack{(x,y) \\ K(x,y) > 0}} \frac{1}{K(x,y)\pi(x)} \sum_{\sigma \ni (x,y)} |\sigma| \theta(\sigma).$$

First, examine the case where $x = y$ contains more than one coordinate equal to the identity. Then, the quantity we have to bound becomes $n/N(x) \leq n/2$.

Second, if $x = y$ contains exactly one coordinate, say x_i , equal to the identity. Then, we have to bound

$$n \left(1 + \sum_{g \in G} \frac{2|g|_x + 1}{n - b} \right) \leq \frac{n(n - b + |G|^2 - 1)}{n - b}.$$

Here $|g|_x$ denotes the length of g in some generating set which depends on x and we have used (3.3) to obtain the right hand side.

Finally, if x, y differ at exactly one coordinate, say $x_i \neq y_i$, then we have to bound

$$\frac{2n(n-1)}{N(x,y)} \sum_{g \in G} \frac{2|g|_x + 1}{n-b} \leq \frac{2n(n-1)(|G|^2 - 1)}{n-b}.$$

Using the fact that $n \geq 2b$, we get

$$I(\theta) \leq \frac{2n(n-1)(|G|^2 - 1)}{n-b} \leq 4n|G|^2.$$

This proves Proposition 3.3.

We now define an extension of functions $f : \mathcal{X} \rightarrow \mathbb{R}$ to \mathcal{Z} similar to (2.2). For the rest of this section we restrict our attention to the case $G = \mathbb{Z}_q$, $q = p^a$, p a prime. Set

$$\mathcal{M} = \{t : 1 \leq t \leq q - 1; p \text{ does not divide } t\}.$$

Thus $|\mathcal{M}| = p^{a-1}(p-1)$. Further, set

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in \mathcal{X} \\ \frac{1}{|\mathcal{M}|n} \sum_{\substack{m \in \mathcal{M} \\ i \leq i \leq n}} f(x_m^i) & \text{if } x \in \mathcal{Z} \setminus \mathcal{X} \end{cases}$$

Here x_m^i is the n -tuple obtained from x by replacing the i^{th} entry of x by m . The following lemma is easy.

Lemma 3.4 *The chains K and Q defined by (3.1), (3.2) satisfy*

$$\text{Var}_\pi(f) \leq \frac{|\mathcal{Z}|}{|\mathcal{X}|} \text{Var}_\mu(\tilde{f}), \tag{3.4}$$

$$\mathcal{L}_\pi(f) \leq \frac{|\mathcal{Z}|}{|\mathcal{X}|} \mathcal{L}_\mu(\tilde{f}). \tag{3.5}$$

Proof: Use the argument given for the similar statements (2.3) (2.4) in Proposition 2.3. Actually, any extension of f would do the job here.

Remark: In the present case where $G = \mathbb{Z}_{p^a}$, p a prime, the size of the state space \mathcal{X} is easy to compute. Namely, $|\mathcal{X}| = p^{na} - p^{n(a-1)}$. Thus, $\mathcal{X} \sim \mathcal{Z}$ as p and n tend to infinity. Our comparison argument does not require us to determine the size of \mathcal{X} because the ratio $|\mathcal{Z}|/|\mathcal{X}|$ in the above comparison of the variances cancel with $|\mathcal{X}|/|\mathcal{Z}|$ in the comparison of the Dirichlet forms of Proposition 3.5 below.

We now reach the crucial part of the comparison argument. The Dirichlet forms $\mathcal{E}_Q(\tilde{f}, \tilde{f})$ and $\mathcal{E}_K(f, f)$ must be compared.

Proposition 3.5 For $G = \mathbb{Z}_q$, $q = p^a$, $p \geq 3$ prime, and $n \geq 2$, the chain K and Q defined at (3.1), (3.2) satisfy

$$\mathcal{E}_Q(\tilde{f}, \tilde{f}) \leq \frac{8(n-1)q^2|\mathcal{K}|}{|\mathcal{E}|} \mathcal{E}_K(f, f)$$

for all $f : \mathcal{K} \rightarrow \mathbb{R}$.

Proof: Write

$$\begin{aligned} \mathcal{E}_Q(\tilde{f}, \tilde{f}) &= \frac{1}{2n|G|^{n+1}} \\ &\times \left(\sum_{\substack{x, y \in \mathcal{K} \\ x \sim y}} |f(x) - f(y)|^2 + 2 \sum_{\substack{x \in \mathcal{E} \setminus \mathcal{K}, y \in \mathcal{K} \\ x \sim y}} |\tilde{f}(x) - f(y)|^2 + \sum_{\substack{x, y \in \mathcal{E} \setminus \mathcal{K} \\ x \sim y}} |\tilde{f}(x) - \tilde{f}(y)|^2 \right) \\ &= \frac{1}{2n|G|^{n+1}} (R_1 + 2R_2 + R_3). \end{aligned} \quad (3.6)$$

We first leave R_1 as it is and bound R_2 and R_3 in terms of R_1 . For R_3 , write

$$\begin{aligned} R_3 &= \sum_{\substack{x, y \in \mathcal{E} \setminus \mathcal{K} \\ x \sim y}} |\tilde{f}(x) - \tilde{f}(y)|^2 \\ &= \frac{1}{n^2|\mathcal{M}|^2} \sum_{\substack{x, y \in \mathcal{E} \setminus \mathcal{K} \\ x \sim y}} \left| \left(\sum_{i, m} f(x_m^i) \right) - \left(\sum_{i, m} f(y_m^i) \right) \right|^2 \\ &\leq \frac{1}{n|\mathcal{M}|} \sum_{\substack{x, y \in \mathcal{E} \setminus \mathcal{K} \\ x \sim y}} \sum_{i, m} |f(x_m^i) - f(y_m^i)|^2 \\ &= \frac{1}{n|\mathcal{M}|} \sum_{\substack{z, w \in \mathcal{E} \\ z \sim w}} |f(z) - f(w)|^2 \sum_{\substack{x, y \in \mathcal{E} \setminus \mathcal{K} \\ x \sim y}} \sum_{i, m} 1_{x_m^i = z} 1_{y_m^i = w} \\ &\leq \frac{(|G| - |\mathcal{M}|)}{n|\mathcal{M}|} R_1 = \frac{1}{n(p-1)} R_1. \end{aligned} \quad (3.7)$$

To see the last inequality observe that $x_m^i = z$ determines i as the only coordinate in z such that $z_i \in \mathcal{M}$ (because $x \in \mathcal{E} \setminus \mathcal{K}$). Thus, z and w determine i, m and a place j where they differ. Now x and y must differ only at j . We are left with the choice of the coordinate $x_i = y_i$ in $G \setminus \mathcal{M}$. The factor $|G| - |\mathcal{M}|$ accounts for this choice.

For R_2 , recall that

$$R_2 = \sum_{\substack{x \in \mathcal{E} \setminus \mathcal{K}, y \in \mathcal{K} \\ x \sim y}} \left| \left(\frac{1}{n|\mathcal{M}|} \sum_{i, m} f(x_m^i) \right) - f(y) \right|^2.$$

In the above sum we have $y \sim x$ so that there must be an $\ell \in \mathcal{M}$ and a $j \in \{1, \dots, n\}$ such that $y = x_\ell^j$. Thus,

$$R_2 = \sum_{x \in \mathcal{X} \setminus \mathcal{X}} \sum_{j, \ell} \left| \left(\frac{1}{n|\mathcal{M}|} \sum_{i, m} f(x_m^i) \right) - f(x_\ell^j) \right|^2$$

$$= \frac{1}{2n|\mathcal{M}|} \sum_{x \in \mathcal{X} \setminus \mathcal{X}} \sum_{j, \ell} \sum_{i, m} |f(x_m^i) - f(x_\ell^j)|^2.$$

If $i = j$, $x_m^i \sim x_\ell^j$. For $i \neq j$ write

$$|f(x_m^i) - f(x_\ell^j)|^2 \leq 2 \left(|f(x_m^i) - f([x_m^i]_\ell^j)|^2 + |f([x_m^i]_\ell^j) - f(x_\ell^j)|^2 \right).$$

Observe that these differences are all taken over edges (z, w) with $z, w \in \mathcal{X}$ and $z \sim w$. Moreover

$$\sum_{x \in \mathcal{X} \setminus \mathcal{X}} \sum_{j, \ell} \sum_{i \neq j, m} |f(x_m^i) - f(x_\ell^j)|^2$$

$$\leq 2 \sum_{\substack{z, w \in \mathcal{X} \\ z \sim w}} |f(z) - f(w)|^2 \sum_{x \in \mathcal{X} \setminus \mathcal{X}} \sum_{j, \ell} \sum_{i \neq j, m} \left(1_{z=x_m^i} 1_{w=[x_m^i]_\ell^j} + 1_{z=[x_m^i]_\ell^j} 1_{w=x_\ell^j} \right)$$

$$\leq 4(|G| - |\mathcal{M}|)R_1 = 4p^{a-1}R_1.$$

To obtain the last inequality, observe that the condition $z = x_m^i$, $w = [x_m^i]_\ell^j$ determines i as the only place with coordinate in \mathcal{M} and j as the only place where z and w differ. It also determines m and ℓ and all but the i^{th} coordinates of x . There are $|G| - |\mathcal{M}|$ distinct possible choices for x_i . Similarly, the condition $z = [x_m^i]_\ell^j$, $w = x_\ell^j$ determines i, j, ℓ, m and leaves only $|G| - |\mathcal{M}|$ distinct choices for x .

In conclusion we have proved that

$$\mathcal{E}_Q(\tilde{f}, \tilde{f}) \leq \frac{1}{2n|G|^{n+1}} \left(1 + \frac{3}{n(p-1)} \right) R_1. \tag{3.8}$$

The next step is to bound

$$R_1 = \sum_{\substack{x, y \in \mathcal{X} \\ x \sim y}} |f(x) - f(y)|^2$$

in terms of

$$\sum_{\substack{z, w \in \mathcal{X} \\ z \approx w}} |f(z) - f(w)|^2.$$

Let e_i denote the n -tuple with all coordinates zero except the i^{th} which is one. Given $x, y \in \mathcal{X}$ such that $x \sim y$, consider three cases:

- 1) If $x \approx y$, do nothing. This will contribute a factor 1 to the edge $z = x$, $w = y$.
- 2) If $x \not\approx y$ differ exactly at the i^{th} coordinate and there exists $j \neq i$ such that $x_j = y_j \in \mathcal{M}$, write

$$|f(x) - f(y)|^2 \leq \frac{q}{2} \sum_{\ell=1}^{q-1} |f(x + \ell x_j e_i) - f(x + (\ell + 1)x_j e_i)|^2.$$

When summing over all x, y this will contribute at most a factor of $q|G|^2/2 = q^3/2$ to each edge (z, w) with $z \approx w$. Indeed, the condition $z = x + \ell x_j e_i$, $w = x + (\ell + 1)x_j e_i$ determines i and all but the i^{th} coordinates of x and y . Pick the i^{th} coordinate of x and that of y , each among the $|G|$ possible choices. Finally, this also determines ℓ .

3) If $x \not\approx y$ differ exactly at the i^{th} coordinate and all the other coordinates are divisible by p , write

$$\begin{aligned} |f(x) - f(y)|^2 &\leq \frac{q+1}{n-1} \sum_{j \neq i} \left(|f(x + (y_i - x_i - x_j)e_j) - f(y + (y_i - x_i - y_j)e_j)|^2 \right. \\ &+ \sum_{\ell=1}^{q-1} \left(|f(x + \ell x_i e_j) - f(x + (\ell + 1)x_i e_j)|^2 + |f(y + [\ell y_i - x_i]e_j) \right. \\ &\quad \left. \left. - f(y + [(\ell + 1)y_i - x_i]e_j)|^2 \right) \right) \end{aligned}$$

When summing over all x, y this will contribute at most a factor

$$\frac{q+1}{n-1} (2|G|(|G| - |\mathcal{M}|) + 1) \leq 1 + q^3.$$

Thus, this case by case study gives

$$R_1 \leq 2q^3 \sum_{\substack{z, w \in \mathcal{X} \\ z \approx w}} |f(z) - f(w)|^2.$$

This inequality, together with (3.6), (3.7), (3.8), yields

$$\begin{aligned} \mathcal{E}_Q(\tilde{f}, \tilde{f}) &\leq \frac{1}{2n|G|^{n+1}} \left(1 + \frac{3}{n(p-1)} \right) R_1 \\ &\leq \frac{1}{2n|G|^{n+1}} \left(1 + \frac{3}{n(p-1)} \right) 2q^3 \sum_{\substack{z, w \in \mathcal{X} \\ z \approx w}} |f(z) - f(w)|^2 \\ &\leq \frac{8(n-1)q^2 |\mathcal{X}|}{|\mathcal{E}|} \mathcal{E}_K(f, f) \end{aligned}$$

which proves Proposition 3.5.

Proof of Theorem 3.1: It follows from Lemma 3.4, Proposition 3.5, and the results given above for Q that the second largest eigenvalue of K is bounded by

$$\beta_1(K) \leq 1 - \frac{1}{8n^2q^2}.$$

Similarly, for the log-Sobolev constant,

$$\alpha(K) \geq \frac{q - 2}{8n^2q^3 \log(q - 1)}.$$

Now, apply Proposition 3.3 and Theorem 1.1.

Remark: Assume that $a = 1$, i.e., $q = p$ is a prime. Then there is no difficulty in extending the comparison argument of Sect. 2.B to study the walk (1.1) with $G = \mathbb{Z}_p$ on a connected graph \mathcal{G} . This produces just an extra factor of p :

Theorem 3.6 For $n, p \geq 3$, p prime, let $G = \mathbb{Z}_p$ and $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ be a connected graph. Then the chain K at (3.1) on the complete graph and the chain P on $\mathcal{K} = \mathbb{Z}_p \setminus \{0\}$ defined at (1.1) with underlying graph \mathcal{G} satisfy

$$\mathcal{E}_K \leq A \mathcal{E}_P$$

with

$$A = \frac{4p|\mathcal{A}|\Delta}{n(n-1)}.$$

For any graph \mathcal{G} , this, together with an easy bound on the least eigenvalue, shows that the chain P reaches stationarity after order $n^5 p^3 (\log p) (\log \log p^n)$.

In principle, the argument can also be used if $q = p^a$, $a > 1$, but it gets more complicated and we have not worked out the details.

4. A matrix walk argument

This section connects the walk (1.1) to certain matrix walks. Namely, consider the walk (1.1) based on the complete graph on n vertices with $G = \mathbb{Z}_p^k$ with p prime and $k \leq n$. At each stage, this walk proceeds by picking a pair of coordinates at random and adding ± 1 times one coordinate to the other. This can be represented as the walk generated by the last k columns of the walk on $SL_n(\mathbb{Z}_p)$ generated by elementary transvections. To define this formally, for $i \neq j$, let $E_{i,j}$ be the $n \times n$ matrix with ones down the diagonal and a 1 in position (i, j) , zeros elsewhere. This operates on n copies of \mathbb{Z}_p^k arranged as an $n \times k$ array by adding the j^{th} row to the i^{th} row. The following proposition shows that the walk generated by $E_{i,j}^{\pm 1}$ has mixing time of order $n^4 (\log p)^2$. We relate this to the walks of the present paper following the proof.

Theorem 4.1 Let Q be the uniform distribution on the set $\{E_{i,j}^{\pm 1}\}$ of elementary transvections in $SL_n(\mathbb{Z}_p)$. Then,

$$\|Q^{(\ell)} - U\|_{TV} \leq Ae^{-at} \text{ for } \ell \geq (n \log p)^2 (n^2 \log p + t) \text{ with } t > 0.$$

with $A, a > 0$ universal computable constants, independent of ℓ, n, p . Here $Q^{(\ell)}$ denotes the ℓ^{th} convolution power of the probability measure Q .

The proof uses comparison techniques from [8, 9] to allow work of Hildebrand to be brought to bear. We need a corollary of Theorem 2.7 which is not explicitly stated in either [8] or [9] and which has some independent interest. It gives a comparison between two Markov chains on the same state space in presence of symmetry.

Let $(K, \mu), (P, \pi)$ be two reversible irreducible chains on the same state space \mathcal{X} . Consider the two edge sets $\mathcal{A}(K), \mathcal{A}(P)$ where

$$\mathcal{A}(P) = \{(x, y) \in \mathcal{X} \times \mathcal{X} : P(x, y) > 0, x \neq y\} \tag{4.1}$$

and similarly for $\mathcal{A}(K)$. For each (x, y) in $\mathcal{A}(K)$ let $\Gamma_{x,y}$ be the set of all geodesic paths joining x to y in the graph $(\mathcal{X}, \mathcal{A}(P))$. Define a K, P -flow (see the definition before Theorem 2.7) η by setting

$$\eta(\gamma) = \begin{cases} \frac{K(x,y)\mu(x)}{|\Gamma_{x,y}|} & \text{if } \gamma \in \Gamma_{x,y} \\ 0 & \text{otherwise.} \end{cases}$$

Then, Theorem 2.7 shows that $\mathcal{E}_K \leq A(\eta)\mathcal{E}_P$ with

$$A(\eta) = \max_{(z,w) \in \mathcal{A}(P)} \frac{1}{P(z,w)\pi(z)} \sum_{(x,y) \in \mathcal{A}(K)} \sum_{\substack{\gamma \in \Gamma_{x,y} \\ \gamma \ni (z,w)}} \frac{|\gamma|K(x,y)\mu(x)}{|\Gamma_{x,y}|}$$

where $|\gamma|$ is the length of the path γ which depends only on x, y if $\gamma \in \Gamma_{x,y}$.

Our main assumption will be that there is a group H acting on the state space \mathcal{X} which preserves both chains and acts transitively on $\mathcal{A}(P)$.

Theorem 4.2 *Assume that K and P are two reversible Markov chains on the same finite state space \mathcal{X} with uniform stationary measure $\pi \equiv 1/|\mathcal{X}|$. Assume further that there is a group H acting on \mathcal{X} an such that*

- i) $\pi(hx) = \pi(x), \mu(hx) = \mu(x)$ for all $h \in H, x \in \mathcal{X}$
- ii) $P(hx, hy) = P(x, y), K(hx, hy) = K(x, y)$ for all $h \in H, x, y \in \mathcal{X}$
- iii) H acts transitively on $\mathcal{A}(P)$

with $\mathcal{A}(P)$ defined at (4.1). Then $\mathcal{E}_K \leq A \mathcal{E}_P$ with

$$A = \frac{1}{1 - \varepsilon} \sum_{\substack{(x,y) \\ K(x,y) > 0}} d_P^2(x, y)K(x, y)\pi(x) \leq \frac{1}{1 - \varepsilon} \max_{\substack{(x,y) \\ K(x,y) > 0}} d_P^2(x, y).$$

Here, d_P is the distance function of the graph $(\mathcal{X}, \mathcal{A}(P))$ and $\varepsilon = P(x, x)$ which does not depend on x .

Proof: Observe that under these circumstances the group H acts transitively on \mathcal{X} which forces $\pi = \mu = 1/|\mathcal{X}|$. Moreover, the graph $(\mathcal{X}, \mathcal{A}(P))$ must be regular of degree $|\mathcal{A}(P)|/|\mathcal{X}|$, and P must be of the form

$$P(x, y) = \begin{cases} \varepsilon & \text{if } x = y \\ \frac{|\mathcal{B}|}{|\mathcal{A}(P)|}(1 - \varepsilon) & \text{if } (x, y) \in \mathcal{A}(P) \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the quantity

$$\frac{1}{P(z, w)\pi(z)} \sum_{(x,y) \in \mathcal{A}(K)} \sum_{\substack{\gamma \in \Gamma_{x,y} \\ \gamma \ni (z,w)}} \frac{|\gamma|K(x, y)\mu(x)}{|\Gamma_{x,y}|}$$

simplifies to

$$\frac{|\mathcal{A}(P)|}{1 - \varepsilon} \sum_{(x,y) \in \mathcal{A}(K)} \sum_{\substack{\gamma \in \Gamma_{x,y} \\ \gamma \ni (z,w)}} \frac{|\gamma|K(x, y)\mu(x)}{|\Gamma_{x,y}|}$$

and does not depend on $(z, w) \in \mathcal{A}(P)$. It follows that the comparison constant $A(\eta)$ satisfies

$$\begin{aligned} A(\eta) &= \frac{1}{|\mathcal{A}(P)|} \sum_{(z,w) \in \mathcal{A}(P)} \frac{|\mathcal{A}(P)|}{1 - \varepsilon} \sum_{(x,y) \in \mathcal{A}(K)} \sum_{\substack{\gamma \in \Gamma_{x,y} \\ \gamma \ni (z,w)}} \frac{|\gamma|K(x, y)\mu(x)}{|\Gamma_{x,y}|} \\ &= \frac{1}{1 - \varepsilon} \sum_{(x,y) \in \mathcal{A}(K)} \sum_{\gamma \in \Gamma_{x,y}} \frac{|\gamma|^2 K(x, y)\mu(x)}{|\Gamma_{x,y}|} \\ &= \frac{1}{1 - \varepsilon} \sum_{(x,y) \in \mathcal{A}(K)} d_P^2(x, y) K(x, y)\mu(x) \end{aligned}$$

where, in the last equality, $d_P(x, y)$ denotes the distance between x and y in the graph $(\mathcal{B}, \mathcal{A}(P))$. This proves the desired result.

Let us apply this result to the chain associated with the probability measure Q on $SL_n(\mathbb{Z}_p)$ where

$$Q(x) = \begin{cases} \frac{1}{2n(n-1)} & \text{if } x = E_{i,j} \text{ for some } (i, j), i \neq j \\ 0 & \text{otherwise.} \end{cases}$$

For comparison, introduce the chain associated with the probability measure \tilde{Q} which is uniform over the set of all transvections. Thus,

$$\tilde{Q}(x) = \begin{cases} \frac{1}{|\mathcal{T}|} & \text{if } x \text{ is a transvection} \\ 0 & \text{otherwise.} \end{cases}$$

Here \mathcal{T} denote the set of all transvections and $|\mathcal{T}|$ is its cardinality. See Suzuki [21], page 73, for background on transvections. It turns out we will not need to know the value of $|\mathcal{T}|$.

First, observe that the automorphism group of the Cayley graph of $SL_n(\mathbb{Z}_p)$ with generating set $E_{i,j}^{\pm 1}$ acts transitively on oriented edges. To change $E_{i,j}$ to $E_{k,\ell}$, conjugate by the permutation matrix that changes i into k and j into ℓ . To change $E_{i,j}$ to $E_{i,j}^{-1}$, conjugate by the matrix I_i which has zeros off the diagonal, -1 in position (i, i) and ones at all other diagonal entries. To obtain a transitive

action on the set of oriented edges, we only need to compose the natural action of $SL_n(\mathbb{Z}_p)$ on its Cayley graph with the above transformations. Observe also that all these transformations preserve the set \mathcal{T} of all transvections. Hence, we are in position to use Theorem 4.2 to compare the Dirichlet form $\mathcal{E}_{\tilde{Q}}$ of all transvections with \mathcal{E}_Q . We still need to write any transvection $T \in \mathcal{T}$ as a word using the elementary transvections $E_{i,j}^{\pm 1}$. Using results of Lubotzky, Phillips and Sarnack [18], the diameter of $SL_2(\mathbb{Z}_p)$ in the four generators

$$\begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ \pm 1 & 1 \end{pmatrix}$$

is $c_1 \log p$ for a universal constant c_1 . It follows that one can write the transvection $E_{i,j}(x)$, $x \in \mathbb{Z}_p$, which has ones on the diagonal, an x in position (i, j) , and zeros elsewhere in $c_1 \log p$ steps using the $E_{\alpha,\beta}$. These $E_{i,j}(x)$ are the set of all “elementary” transvections corresponding to the row and column operations of linear algebra. The argument in Suzuki [21], pg. 93, shows that the diameter of all of $SL_n(\mathbb{Z}_p)$ in the $E_{i,j}(x)$ is cn^2 . If one only needs to write a transvection, Magaard [19] shows that at most $4n - 5$ of the $E_{i,j}(x)$ are needed. We conclude that

$$\begin{aligned} &\text{the length of any transvection as a shortest word} \\ &\text{using the } E_{i,j}^{\pm 1} \text{ is at most } c_2 n \log p \end{aligned} \tag{4.2}$$

for an explicit constant $c_2 = 4c_1$. Using (4.2) and Theorem 4.2 yields

$$\mathcal{E}_{\tilde{Q}} \leq [c_2 n \log p]^2 \mathcal{E}_Q. \tag{4.3}$$

Now, Hildebrand [16] shows that there are universal constants $A, a > 0$ such that

$$\|\tilde{Q}^{(\ell)} - U\|_{TV} \leq Ae^{-a\ell}$$

provided $\ell \geq n + t$, $c > 0$. This implies in particular that

$$\beta_1(\tilde{Q}) \leq e^{-a}$$

since β_1 governs the asymptotic rate of convergence. Hildebrand’s bound and (4.3) imply the following result.

Proposition 4.3 *For Q as in Theorem 4.1, the second largest eigenvalue $\beta_1(Q)$ satisfies*

$$\beta_1(Q) \leq 1 - \frac{c}{(n \log p)^2}$$

for a universal constant c .

To prove Theorem 4.1 we still need one more ingredient which is a lower bound on the least eigenvalue.

Proposition 4.4 *For $n \geq 3$, the least eigenvalue of the chain associated with Q is bounded by*

$$\beta_{\min}(Q) \geq -1 + \frac{1}{25}.$$

This will be deduced from the following general result which will be proved as a corollary of Lemma 2.7.

Lemma 4.5 *Let P, π be a reversible Markov chain on a finite state space \mathcal{X} with uniform stationary measure $\pi \equiv 1/|\mathcal{X}|$. Assume that there is a group H acting on \mathcal{X} and such that $P(hx, hy) = P(x, y)$ for all $h \in H, x, y \in \mathcal{X}$. Assume that H acts transitively on the oriented edge set $\mathcal{A} = \{(x, y) \in \mathcal{X} \times \mathcal{X} : P(x, y) > 0\}$ (this implies $P(x, x) = 0$). Assume further that there exists at least one cycle of odd length D with at most R edges repeated twice. Then, the least eigenvalue of P satisfies*

$$\beta_{\min}(P) \geq -1 + \frac{2}{D^2}.$$

Proof: Use Lemma 2.7 with Σ_x the set of all cycles of odd length D based at x . We can clearly assume that there are no repeated edges. The hypotheses imply that this set is not empty. Let θ be the flow on loops of odd length given by

$$\theta = \begin{cases} \frac{\pi(x)}{|\Sigma_x|} & \text{if } \sigma \in \Sigma_x \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 2.7 yields

$$\beta_{\min}(P) \geq -1 + \frac{2}{I(\theta)}$$

with

$$I(\theta) = \max_{\substack{(x,y) \\ P(x,y) > 0}} \left(|\mathcal{A}| \sum_{z \in \mathcal{X}} \sum_{\substack{\sigma \in \Sigma_z \\ \sigma \ni (x,y)}} \frac{|\sigma| \pi(z)}{|\Sigma_z|} \right)$$

where $r(\sigma, (x, y))$ is the number of time the edge (x, y) is used in σ . We assume (without loss of generality) that $r(\sigma, (x, y)) \leq 2$. Using the action of H , we see that the quantity we take the maximum of is constant. Thus, the maximum can be replaced by the mean over edges in \mathcal{A} . This gives

$$\begin{aligned} I(\theta) &= \sum_{(x,y) \in \mathcal{A}} \sum_{z \in \mathcal{X}} \sum_{\substack{\sigma \in \Sigma_z \\ \sigma \ni (x,y)}} \frac{|\sigma|}{|\Sigma_z|} \\ &= \sum_{z \in \mathcal{X}} \sum_{\sigma \in \Sigma_z} \frac{|\sigma| \pi(z)}{|\Sigma_z|^2} \\ &= D^2. \end{aligned}$$

This proves Lemma 4.5.

Proof of Proposition 4.4: In order to apply Lemma 4.5 we only need to find one cycle of odd length because we already know that the automorphism group of our Cayley graph acts transitively on oriented edge. But, for $n \geq 3$ and for any distinct i, j, k

$$\text{Id} = E_{j,k}^{-1} E_{i,j}^{-1} E_{j,k} E_{i,k} E_{i,j}.$$

Thus, we have at least one loop of odd length 5 in our Cayley graph. Moreover, no edge is used twice. Thus, we conclude

$$\beta_{\min}(Q) \geq -1 + \frac{1}{25}.$$

Proof of Theorem 4.1: For $n \geq 3$, Propositions 4.3, 4.4 shows that

$$\beta(Q) = \max\{\beta_1(Q), -\beta_{\min}(Q)\} \leq 1 - \frac{c}{(n \log p)^2}.$$

Since $|SL_n(\mathbb{Z}_p)| \leq p^{n^2}$ the bound (1.5) gives

$$\|Q^{(\ell)} - U\|_{TV} \leq p^{n^2/2} e^{-c\ell/(n \log p)^2}$$

which proves Theorem 4.1.

Let us return to the subject of this paper. As mentioned above, for $k \leq n$ and under the random walk associated with Q , the last k columns of the matrix evolves like the random walk (1.1) on the complete graph with n vertices with $G = \mathbb{Z}_p^k$. Thus, Theorem 4.1 implies that order $n^4(\log p)^3$ steps suffice to reach stationarity, uniformly in n, p, k . For fixed p and k (e.g., $k = 1, p = 2$), these results are not as sharp as those of Sects. 2 and 3. On the other hand, for fixed k and large p , the results of this section are much sharper and seem unobtainable by elementary path arguments. They lean on the deep number theoretic results of Lubotsky Phillips and Sarnack [18]. Davidof and Sarnack [4] give remarkable simplifications but the results are still quite deep. Theorem 4.1 also handles the large k case (e.g., $p = 2, k = n$).

However, note that the above argument depends very much on working over \mathbb{Z}_p with p prime. It breaks down for composite p but can be extended without change to any finite field.

Finally, we remark that Gluck [14] has given a different proof of Hildebrand's results which generalizes to random walks on essentially arbitrary conjugacy classes of general Chevalley groups.

Acknowledgement: We thank David Aldous, Rosemary Bailey, Jordan Ellenberg, David Gluck, Ron Graham, Charles Leedham-Green, Kai Magaard, Dan Rockmore, Chris Rowley and Bálint Virág for their help with this paper.

References

1. Babai L. (1986): *On the length of subgroup chains in the symmetric group*. Comm. Alg. 14, 1729–1736
2. Cameron P., Solomon R., Turull A. (1989): *Chains of subgroups in symmetric groups*. Jour. Alg. 127, 340–352

3. Celler F., Leedham-Green C., Murray S., Niemeyer A., O'Brien E. (1995): *Generating random elements of a finite group*. Commun. Algebra 23, 4931–4948
4. Davidov G., Sarnack P.: Forthcoming book
5. Diaconis P. (1982): *Applications of non-commutative Fourier analysis to probability problems*. Springer L.N.M. 1362, 51–100
6. Diaconis P. (1986): *Group representations in probability and statistics*. IMS, Hayward
7. Diaconis P., Graham R.L. (1995): *On graphs and groups*. Preprint, Dept. of Math. Harvard University
8. Diaconis P., Saloff-Coste L. (1993): *Comparison theorems for reversible Markov chains*. Ann. Appl. Prob. 3, 696–730
9. Diaconis P., Saloff-Coste L. (1993): *Comparison techniques for random walk on finite groups*. Ann. Prob. 21, 2131–2156
10. Diaconis P., Saloff-Coste L. (1992): *Logarithmic Sobolev inequalities and finite Markov chains*. Preprint
11. Diaconis P., Saloff-Coste L. (1995): *Walks on generating sets of groups*
12. Diaconis P. and Stroock D. (1991): *Geometric bounds for eigenvalues for Markov chains*. Ann. Appl. Prob. 1, 36–61
13. Finkelstein L., Kantor W. (1993): *Groups and computation*. Amer. Math. Soc. Providence
14. Gluck D. (1994): *Random walk and character ratios on finite groups of Lie type*. Adv. Math. To appear
15. Gross L. (1976): *Logarithmic Sobolev inequalities*. Am. J. Math. 1061–1083
16. Hildebrand M. (1992): *Generating random elements in $SL_n(F_q)$ by random transvections*. J. Alg. Combinatorics 1, 133–150
17. Holley R., Stroock D. (1987): *Logarithmic Sobolev inequalities and stochastic Ising models*. J. Stat. Phys. 46, 1159–1194
18. Lubotzky A., Phillips R., Sarnack P. (1988): *Ramanujan graphs*. Combinatorica, 8, 261–277
19. Magaard K. (1995): Personal communication
20. Sinclair A. (1993): *Algorithms for random generation and counting: a Markov chain approach*. Birkhäuser, Boston
21. Suzuki M. (1982): *Group theory I*. Springer, New York

