Random walks on generating sets for finite groups

F. R. K. Chung [1]
University of Pennsylvania
Philadelphia, PA 19104

R. L. Graham
AT&T Research
Murray Hill, NJ 07974

Submitted: August 31, 1996; Accepted: November 12, 1996

Dedicated to Herb Wilf on the occasion of his sixty-fifth birthday

Abstract

We analyze a certain random walk on the cartesian product G^n of a finite group G which is often used for generating random elements from G. In particular, we show that the mixing time of the walk is at most $c_r n^2 \log n$ where the constant c_r depends only on the order r of G.

1. Introduction

One method often used in computational group theory for generating random elements from a given (non-trivial) finite group G proceeds as follows (e.g., see [2]). A fixed integer $n \geq 2$ is initially specified. Denote by G^n the set $\{(x_1, \ldots, x_n) : x_i \in G, 1 \leq i \leq n\}$. If $\bar{x} = (x_1, \ldots, x_n) \in G^n$, we denote by $\langle \bar{x} \rangle$ the subgroup of G generated by $\{x_i : 1 \leq i \leq n\}$. Let $G^* \subseteq G^n$ denote the set of all $\bar{x} \in G^n$ such that $\langle \bar{x} \rangle = G$. We execute a random walk on G^* by taking the following general step. Suppose we are at a point $\bar{p} = (p_1, \ldots, p_n) \in G^*$. Choose a random pair of indices (i, j) with $i \neq j$. (Thus, each such pair is chosen with probability $\frac{1}{n(n-1)}$.) We then move to one of $\bar{p}' = (p_1', \ldots, p_n')$ where

$$p_k' = \begin{cases} p_i p_j \quad or \quad p_i p_j^{-1} & \text{if } k = i, \text{ each with probability } 1/2 \\ p_k & \text{if } k \neq i \ . \end{cases}$$

This rule determines the corresponding transition matrix Q of the walk. We note that with this rule, we always have $\bar{p}' \in G^*$. It is also easy to check that for $n \geq n_0(G)$, this walk is irreducible and aperiodic (see Section 5 for more quantitative remarks), and has a stationary distribution π which is uniform (since G^* is a multigraph in which every vertex has degree $2n(n-1)$).

[1] Research supported in part by NSF Grant No. DMS 95-04834

Starting from some fixed initial distribution f_0 on G^*, we apply this procedure some number of times, say t, to reach a distribution f_0Q^t on G^* which we hope will be close to "random" when t is large. A crucial question which must be faced in this situation is just how rapidly this process mixes, i.e., how large must t be so that f_0Q^t is close to uniform. In this note, we apply several rather general comparison theorems to give reasonably good bounds on the mixing time for Q. In particular, we show (see Theorem 1) that when $t \geq c(G)n^2 \log n$, where $c(G)$ is a constant depending only on G, then Q^t is already quite close to uniform (where we usually will suppress f_0).

This problem belongs to a general class of random walk problems suggested recently by David Aldous [1]. In fact, he considers a more general walk in which only certain pairs of indices (i,j) are allowed in forming $p'_k = p_ip_j$ or $p_ip_j^{-1}$. These pairs can be described by a graph H on the vertex set $\{1,2,\cdot,n\}$. The case studied in this note corresponds to taking H to be a complete graph.

We first learned of this problem from a preprint of Diaconis and Saloff-Coste [6], part of which has subsequently appeared [7]. In it, they wrote "$\cdots$ for $G = \mathbb{Z}_p$ with $p = 2,3,4,5,7,8,9$ we know that $n^2 \log n$ steps are enough whereas for $G = \mathbb{Z}_6$ or $\mathbb{Z}_{10}$ we only know that $n^4 \log n$ are enough. Even in the case of $\mathbb{Z}_6$ it does not seem easy to improve this." Our main contribution in this note is to show that by direct combinatorial constructions, a mixing time of $c(G)n^2 \log n$ can be obtained for *all* groups G where $c(G)$ is a constant depending just on G. Subsequently, they have now [8] also obtained bounds of the form $c(G)n^2 \log n$ for all groups G by including a more sophisticated path construction argument than they had previously used in [6].

2. Background

A weighted graph $\Gamma = (V,E)$ consists of a vertex set V, and a weight function $w : V \times V \to \mathbb{R}$ satisfying $w(u,v) = w(v,u) \geq 0$ for all $u, v \in V$. The edge set E of Γ is defined to be the set of all pairs uv with $w(u,v) > 0$. A simple (unweighted) graph is just the special case in which all weights are 0 or 1. The *degree* d_v of a vertex v is defined by

$$d_v := \sum_u w(u,v)\ .$$

Further, we define the $|V| \times |V|$ matrix L by

$$L(u,v) = \begin{cases} d_v - w(v,v) & \text{if } u = v, \\ -w(u,v) & \text{if } uv \in E,\ u \neq v, \\ 0 & \textit{otherwise} . \end{cases}$$

In particular, for a function $f : V \to \mathbb{R}$, we have

$$Lf(x) = \sum_{\substack{y \\ xy \in E}} (f(x) - f(y)) w(x,y) .$$

Let T denote the diagonal matrix with the (v,v) entry having the value d_v. The *Laplacian* $\mathcal{L}_\Gamma$ of Γ is defined to be

$$\mathcal{L} = \mathcal{L}_\Gamma = T^{-1/2} L T^{-1/2} .$$

In other words,

$$\mathcal{L}(u,v) = \begin{cases} 1 - \frac{w(v,v)}{d_v} & \text{if } u = v, \\ -\frac{w(u,v)}{\sqrt{d_u d_v}} & \text{if } uv \in E,\ u \neq v, \\ 0 & \textit{otherwise} . \end{cases}$$

Since $\mathcal{L}$ is symmetric and non-negative definite, its eigenvalues are real and non-negative. We denote them by

$$0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$$

where $n = |V|$.

It follows from standard variational characterizations of eigenvalues that

$$\lambda_1 = \inf_f \sup_c \frac{\sum\limits_{u,v \in E} (f(u) - f(v))^2 w(u,v)}{\sum\limits_x d_x (f(x) - c)^2} . \tag{1}$$

For a connected graph Γ, the eigenvalues satisfy

$$0 < \lambda_i \leq 2$$

for $i \geq 1$. Various properties of the eigenvalues can be found in [3].

Now, the usual random walk on an unweighted graph has transition probability $1/d_v$ of moving from a vertex v to any one of its neighbors. The transition matrix P then satisfies

$$P(v,u) = \begin{cases} 1/d_v & \text{if } uv \in E, \\ 0 & \textit{otherwise} . \end{cases}$$

That is,

$$fP(u) = \sum_{\substack{v \\ uv \in E}} \frac{1}{d_v} f(v)$$

for any $f : V \to \mathbb{R}$. It is easy to check that

$$P = T^{-1/2}(I - R)T^{1/2} = T^{-1}A$$

where A is the adjacency matrix of the graph.

In a random walk on a connected weighted graph Γ, the transition matrix P satisfies

$$1TP = 1T \ .$$

Thus, the stationary distribution is just $1T/vol(\Gamma)$, where $vol(\Gamma) = \sum_x d_x$ and $\mathbf{1}$ is the all ones vector. Our problem is to estimate how rapidly fP^k converges to its stationary distribution, as $k \to \infty$, starting from some initial distribution $f : V \to \mathbb{R}$. First, consider convergence in the L^2 (or Euclidean) norm. Suppose we write

$$fT^{-1/2} = \sum_i a_i \phi_i$$

where ϕ_i denotes the eigenfunction associated with λ_i and $\|\phi_i\| = 1$. Since $\phi_0 = 1 \cdot T^{1/2}/\sqrt{vol(\Gamma)}$ then

$$a_0 = \frac{\langle fT^{-1/2}, 1\, T^{1/2}\rangle}{\|1\, T^{1/2}\|} = \frac{1}{\sqrt{vol(\Gamma)}}$$

since $\langle f, 1\rangle = 1$. We then have

$$\begin{aligned}
\|fP^s - 1T/vol(\Gamma)\| &= \|fT^{-1/2}(I-\mathcal{L})^s T^{1/2} - a_0\phi_0 T^{1/2}\| \\
&= \left\| \sum_{i\neq 0} (1-\lambda_i)^s a_i \phi_i T^{1/2} \right\| \\
&\leq (1-\lambda)^s \|f\| \\
&\leq e^{-s\lambda}\|f\|
\end{aligned}$$

where

$$\lambda = \begin{cases} \lambda_1 & \text{if } 1-\lambda_1 \geq \lambda_{n-1} - 1, \\ 2 - \lambda_{n-1} & otherwise\ . \end{cases}$$

So, after $s \geq (1/\lambda)\log(1/\epsilon)$ steps, the L_2 distance between fP^s and its stationary distribution is at most $\epsilon\|f\|$.

Although λ occurs in the above bound, in fact only λ_1 is crucial, in the following sense. If it happens that $1 - \lambda_1 < \lambda_{n-1} - 1$, then we can consider a random walk on the modified graph Γ' formed by adding a loop of weight cd_v to each vertex v where $c = (\lambda_1 + \lambda_{n-1})/2 - 1$. The new graph has (Laplacian) eigenvalues $\lambda'_k = \frac{1}{1+c}\lambda_k \leq 1$, $0 \leq k \leq n-1$, so that $1 - \lambda'_1 \geq \lambda'_{n-1} - 1$.

Consequently (see [3]), we only need to increase the number of steps of this "lazy" walk on Γ to $s \geq (1/(\lambda')\log(1/\epsilon)$ to achieve that same L_2 bound on $\epsilon\|f\|$ where λ' is

$$\lambda' = \begin{cases} \lambda_1 & \text{if } 1-\lambda_1 \geq \lambda_{n-1}-1, \\ \frac{2\lambda_1}{\lambda_1+\lambda_{n-1}} & otherwise\ . \end{cases}$$

We note that we have $\lambda' \geq 2\lambda_1/(2+\lambda_1) \geq 2\lambda_1/3$.

A stronger notion of convergence is measured by the L_∞, or relative pointwise distance, which is defined as follows. After s steps, the relative pointwise distance of P to its stationary distribution π is given by

$$\Delta(s) := \max_{x,y} \frac{|P^s(y,x)-\pi(x)|}{\pi(x)}\ .$$

Let δ_z denote the indicator function defined by

$$\delta_z(x) = \begin{cases} 1 & \text{if } x = z, \\ 0 & otherwise\ . \end{cases}$$

Set

$$T^{1/2}\delta_x = \sum_i a_i \phi_i$$

and

$$T^{-1/2}\delta_y = \sum_i \beta_i \phi_i\ .$$

In particular,

$$\alpha_0 = \frac{d_x}{\sqrt{vol(\Gamma)}}, \quad \beta_0 = \frac{1}{\sqrt{vol(\Gamma)}}\ .$$

Hence,

$$\begin{aligned} \Delta(t) &= \max_{x,y} \frac{|\delta_y(P^t)\delta_y - \pi(x)|}{\pi(x)} \\ &= \max_{x,y} \frac{|\delta_y T^{-1/2}(I-\mathcal{L})^t T^{1/2}\delta_x - \pi(x)|}{\pi(x)} \\ &\leq \max_{x,y} \sum_{i\neq 0} \frac{|(1-\lambda_i)^t \alpha_i \beta_i|}{d_x/vol(\Gamma)} \qquad (2) \\ &\leq (1-\lambda)^t \max_{x,y} \frac{\|T^{1/2}\delta_x\|\|T^{-1/2}\delta_y\|}{d_x/vol(\Gamma)} \\ &\leq (1-\lambda)^t \frac{vol(\Gamma)}{\min_{x,y}\sqrt{d_x d_y}} \\ &\leq e^{-t\lambda}\frac{vol(\Gamma)}{\min_x d_x}\ . \end{aligned}$$

Thus, if we choose t so that

$$t \geq \frac{1}{\lambda}\log\frac{vol(\Gamma)}{e\min_x d_x}$$

then after t steps, we have $\Delta(t) \leq \epsilon$. We also remark that requiring $\Delta(t) \to 0$ is a rather strong condition. In particular, it implies that another common measure, the total variation distance $\Delta_{TV}(t)$ goes to zero just as rapidly, since

$$\begin{aligned} \Delta_{TV}(t) &= \max_{A \subset V} \max_{y \in V} \left| \sum_{x \in A} P^t(y,x) - \pi(x) \right| \\ &\leq \max_{\substack{A \subset V \\ vol(A) \leq \frac{1}{2} vol(\Gamma)}} \sum_{x \in A} \pi(x)\Delta(t) \\ &\leq \frac{1}{2}\Delta(t) \ . \end{aligned}$$

We point out here that the factor $\frac{vol(\Gamma)}{\min_x d_x}$ can often be further reduced by the use of so-called logarithmic Sobolev eigenvalue bounds (see [9] and [3] for surveys). In particular, Diaconis and Saloffe-Coste have used these methods in their work on rapidly mixing Markov chains. We will follow their lead and apply some of these ideas in Section 4.

3. An eigenvalue comparison theorem

To estimate the rate at which $\Delta(t) \to 0$ as $t \to \infty$, we will need to lower bound $\lambda_1(\Gamma^*)$, the smallest non-zero Laplacian eigenvalue of the graph Γ^* on G^*, defined by taking as edges all pairs $\bar{x}\bar{y} \in E^*$ where $\bar{x} \in G^*$ and $\bar{y}$ can be reached from $\bar{x}$ by taking one step of the process Q. Our comparison graph Γ^n on G^n will have all edges $\bar{x}\bar{y} \in E$ where $\bar{x}$ and $\bar{y}$ are any two elements of G^n which differ in a single coordinate (so that Γ^n is just the usual Cartesian product of G with itself n times).

Lemma 1. *Suppose $\Gamma = (V, E)$ is a connected (simple) graph and $\Gamma' = (V', E')$ is a connected multigraph with Laplacian eigenvalues $\lambda_1 = \lambda_1(\Gamma)$ and $\lambda'_1 = \lambda_1(\Gamma')$, respectively. Suppose $\phi : V \to V'$ is a surjective map such that:*

(i) If d_x and $d'_{x'}$ denote the degrees of $v \in V$ and $x' \in V'$, respectively, then for all $x' \in V'$ we have

$$\sum_{x \in \phi^{-1}(x')} d_x \geq a d'_{x'} \ .$$

(ii) For each edge $e = xy \in E$ there is a path $P(e)$ between $\phi(x)$ and $\phi(y)$ in E' such that:

(a) The number of edges of $P(e)$ is at most ℓ;

(b) For each edge $e' \in E'$, we have

$$|\{xy \in E : e' \in P(e)| \leq m \ .$$

Then we have

$$\lambda_1' \geq \frac{a}{\ell m}\lambda_1 \tag{3}$$

Proof. For $h : V \to \mathbb{C}$, define $h^2 : E \to \mathbb{C}$ by setting $h^2(e) = (h(x) - h(y))^2$ for $e = xy \in E$ (with a similar definition for $h : V' \to \mathbb{C}$ and $h^2 : E' \to \mathbb{C}$).

We start by letting $g : V' \to \mathbb{C}$ be a function achieving equality in (1) (or rather, the version of (1) for λ_1'). Define $f : V \to \mathbb{C}$ by setting

$$f(x) = g(\phi(x)) \quad \text{for} \quad x \in V \ .$$

Thus,

$$\begin{aligned}
\lambda_1' &= \sup_c \frac{\sum_{e' \in E'} g^2(e')}{\sum_{v' \in V'} (g(v') - c)^2 d'_{v'}} \\
&\geq \frac{\sum_{e' \in E'} g^2(e')}{\sum_{v' \in V'} (g(v') - c)^2 d'_{v'}} \quad \text{for all } c \\
&= \frac{\sum_{e' \in E'} g^2(e')}{\sum_{e \in E} f^2(e)} \cdot \frac{\sum_{e \in E} f^2(e)}{\sum_{v \in V} (f(v) - c)^2 d_v} \cdot \frac{\sum_{v \in V} (f(v) - c)^2 d_v}{\sum_{v' \in V'} (g(v') - c)^2 d'_{v'}} \\
&= I \times II \times III \ .
\end{aligned} \tag{4}$$

First, we treat factor I. Using Cauchy-Schwarz, we have for all $e \in E$,

$$f^2(e) \leq \ell \sum_{e' \in P(e)} g^2(e')$$

by (a). Hence by (b),

$$m \sum_{e' \in E} g^2(e') \geq \sum_{e \in E} \sum_{e' \in E'} g^2(e') \geq \frac{1}{\ell} \sum_{e \in E} f^2(e)$$

i.e.,

$$\frac{\sum_{e' \in E'} g^2(e')}{\sum_{e \in E} f^2(e)} \geq \frac{1}{\ell m} \tag{5}$$

which gives a bound for factor I. To bound factor III, we have

$$\begin{aligned}
\sum_{x \in V} (f(x) - c)^2 d_x &= \sum_{x' \in V'} \sum_{x \in \phi^{-1}(x')} (f(x) - c)^2 d_x \\
&= \sum_{x' \in V'} (g(x') - c)^2 \sum_{x \in \phi^{-1}(x')} d_x \\
&\geq a \sum_{x' \in V'} (g(x') - c)^2 d'_{x'} \qquad \text{by (i)} \ .
\end{aligned} \tag{6}$$

Finally, for factor II we choose c_0 so that

$$\sup_c \frac{\sum_{e \in E} f^2(e)}{\sum_{v \in V} (f(v)-c)^2 d_v} = \frac{\sum_{e \in E} f^2(e)}{\sum_{v \in V} (f(v)-c_0)^2 d_v} \geq \lambda_1 \tag{7}$$

by (1).

Hence, by (4), (5), (6) and (7) we have

$$\lambda_1' \geq \frac{a}{\ell m} \lambda_1$$

which is just (3). ∎

Note that in the case that Γ and Γ' are regular with degrees k and k', respectively, then (i) holds with $a = k/k'$, and (3) becomes

$$\lambda_1' \geq \frac{k}{k' \ell m} \lambda_1 \ . \tag{3'}$$

4. A comparison theorem for the log-Sobolev constant

Given a connected weighted graph $\Gamma = (V, E)$, the log-Sobolev constant $\alpha = \alpha(\Gamma)$ is defined by

$$\alpha = \inf_{f \neq constant} \frac{\sum_{e \in E} f^2(e)}{\sum_x f^2(x) d_x \log \frac{f^2(x)}{\sum_y f^2(y)\pi(y)}} \tag{8}$$

where f ranges over all non-constant functions $f : V \to \mathbb{R}$ and π is the stationary distribution of the nearest neighbor random walk on Γ. In a recent paper [9], Diaconis and Saloffe-Coste show that

$$\Delta_{TV}(t) \leq e^{1-c} \quad if \quad t \geq \frac{1}{2\alpha} \log\log \frac{vol(\Gamma)}{\min_x d_x} + \frac{c}{\lambda_1} \ . \tag{9}$$

This is strengthened in [3], where the slightly stronger inequality is proved

$$\Delta(t) \leq e^{2-c} \quad if \quad t \geq \frac{1}{2\alpha} \log\log \frac{vol(\Gamma)}{\min_x d_x} + \frac{c}{\lambda_1} \tag{10}$$

and

$$\Delta_{TV}(t) \leq e^{1-c} \quad if \quad t \geq \frac{1}{4\alpha} \log\log \frac{vol(\Gamma)}{\min_x d_x} + \frac{c}{\lambda_1} \tag{11}$$

using the alternate (equivalent) definition:

$$\alpha = \inf_{f \neq constant} \frac{\sum_{e \in E} f^2(e)}{S(f)} \tag{12}$$

where

$$S(f) := \inf_{c>0} \sum_{x \in V} (f^2(x) \log f^2(x) - f^2(x) - f^2(x) \log c + c) d_x \ . \tag{13}$$

While (10) is typically stronger than (2), it depends on knowing (or estimating) the value of α, which if anything is harder to estimate than λ_1 for general graphs. We can bypass this difficulty to some extent by the following (companion) comparison theorem for α. Its statement (and proof) is in fact quite close to that of Lemma 1.

Lemma 2. *Suppose $\Gamma = (V, E)$ is a connected (simple) graph and $\Gamma' = (V', E')$ is a connected multigraph, with logarithmic Sobolev constants $\alpha = \alpha(\Gamma)$ and $\alpha' = \alpha(\Gamma')$, respectively. Suppose $\phi : V \to V'$ is a surjective map such that (i), (ii) and (iii) of Lemma 1 hold. Then*

$$\alpha' \geq \frac{a}{\ell m} \alpha \ . \tag{14}$$

Proof: Consider a function $g : V' \to \mathbb{R}$ achieving equality in (14). Define $f : V \to \mathbb{R}$ as in the proof of Lemma 1. Then we have

$$\begin{aligned} \alpha' &= \frac{\sum_{e' \in E'} g^2(e')}{S(g)} \\ &= \frac{\sum_{e' \in E'} g^2(e')}{\sum_{e \in E} f^2(e)} \cdot \frac{\sum_{e \in E} f^2(e)}{S(f)} \cdot \frac{S(f)}{S(g)} \\ &= I' \times II' \times III' \ . \end{aligned} \tag{15}$$

Exactly as in the proof of Lemma 1, we obtain

$$I' \geq \frac{1}{\ell m}, \quad II' \geq \alpha \ .$$

It remains to show $III' \geq a$ (which we do using a nice idea of Holley and Stroock; cf. [9]). First, define

$$F(\xi, \zeta) := \xi \log \xi - \xi \log \zeta - \xi + \zeta$$

for all ξ, $\zeta > 0$. Note that $F(\xi, \zeta) \geq 0$ and for $\zeta > 0$, $F(\xi, \zeta)$ is convex in ξ. Thus, for some $c_0 > 0$,

$$\begin{aligned} S(f) &= \sum_{x \in V} F(f^2(x), c_0) d_x \\ &= \sum_{x' \in V'} \left(\sum_{x \in \phi^{-1}(x')} d_x \right) F(g(x')^2) \end{aligned}$$

$$\begin{aligned} &\geq \sum_{x' \in V'} a d'_{x'} F(g(x')^2) \quad \textit{since } F \geq 0 \\ &\geq a \sum_{x' \in V'} F(g(x')^2 d'_{x'}) \quad \text{by convexity} \\ &= aS(g)\ . \end{aligned}$$

This implies $III' \geq a$ and (14) is proved. ∎

As in ($3'$), if Γ and Γ' are regular with degrees k and k', respectively, then

$$(13') \qquad \alpha' \geq \frac{k}{k'\ell m}\alpha\ .$$

5. Defining the paths

In this section we describe the key path constructions for our proof. For our finite group G, we say that $B \subseteq G$ is a *minimal basis* for G if $\langle B \rangle = G$ but for any proper subset $B' \subset B$, we have $\langle B' \rangle \neq G$. Define

$$b(G) := \max\{|B| :\ B \text{ is a minimal basis for } G\}\ .$$

Further, define $w(G)$ to be the least integer such that for any minimal basis B, and any $g \in G$, we can write g as a product of at most w terms of the form $x^{\pm 1}$, $x \in B$. Finally, define $s(G)$ to be the cardinality of a *minimum* basis for G. We abbreviate $b(G)$, $w(G)$ and $s(G)$ by b, w and s, respectively, and, as usual, we set $r := |G|$. In particular, the following crude bounds always hold:

$$(16) \qquad s \leq b \leq \frac{\log r}{\log 2} = \log_2 r, \quad w < r\ .$$

Let R denote $\lfloor \log_2 r \rfloor$. We will assume $n > 2(s+R)$. To apply Lemmas 1 and 2, we must define the map $\phi : \Gamma^n \to \Gamma^*$ and the paths $P(e)$, $e \in E^n$. Let $\{g_1, \ldots, g_s\}$ be a fixed minimum basis for G.

For $\bar{x} = (x_1, \ldots, x_n) \in \Gamma^n$, define

$$\phi(\bar{x}) = \begin{cases} \bar{x} & \text{if } \langle \bar{x} \rangle = G, \\ (g_1, \ldots, g_s, x_{s+1}, \ldots, x_n) & \text{if } \langle \bar{x} \rangle \neq G\ . \end{cases}$$

Next, for each edge $e = \bar{x}\bar{y} \in E^n$, we must define a path $P(e)$ between $\phi(\bar{x})$ and $\phi(\bar{y})$ in Γ^*. Suppose $\bar{x}$ and $\bar{y}$ just differ in the i^{th} component so that

$$\bar{x} = (x_1, \ldots, x_i, \ldots, x_n), \quad \bar{y} = (y_1, \ldots, y_i, \ldots, y_n)$$

where $x_j = y_j$ for $j \neq i$, and $x_i \neq y_i$. There are three cases:

(I) $\bar{\boldsymbol{x}} \in \boldsymbol{G}^*$, $\bar{\boldsymbol{y}} \in \boldsymbol{G}^*$. Let I denote interval

$$\begin{cases} \{i+1,\dots,i+s+2R\} & \text{if } i \le n-s-2R, \\ \{n-s-2R,\dots,n\} \setminus \{i\} & \text{if } i > n-s-2R \,. \end{cases}$$

Choose a subset $J \subset I$ so that:

$$\begin{array}{rl} (i) & |J| = s \\ (ii) & \langle \{x_k : k \in [n] \setminus |J|\} \rangle = G \\ (iii) & \langle \{y_k : k \in [n] \setminus |J|\} \rangle = G \,. \end{array}$$

That is, the values $x_j = y_j$, $j \in J$, are not needed in generating G using the x_k or the y_k. Write J as $\{j_1, j_2, \dots, j_s\}$. In this case $\phi(\bar{x}) = \bar{x}$, $\phi(\bar{y}) = \bar{y}$. To form $P(e)$:

(i) Use a basis from the elements x_k, $k \notin J$, to change x_{j_1} to g_1, x_{j_2} to $g_2, \dots, x_{j_s}$ to g_s. This takes at most ws steps;

(ii) Next, use $g_1, \dots, g_s$ to change x_i to y_i. This takes at most w steps;

(iii) Finally, use a basis from the elements y_k, $k \notin J$, to change g_1 back to $x_{j_1} = y_{j_1}, \dots, g_s$ back to $x_{j_s} = y_{j_s}$. This takes at most ws steps. Hence, for case (I), $P(e)$ has length at most $w(2s+1)$.

(II) $\bar{\boldsymbol{x}} \notin \boldsymbol{G}^*, \bar{\boldsymbol{y}} \in \boldsymbol{G}^*$. In this case, $\phi(\bar{x}) = (g_1, \dots, g_s, x_{s+1}, \dots, x_n)$, $\phi(\bar{y}) = (y_1, \dots, y_n)$ where $x_j = y_j$ for $j \neq i$, and $x_i \neq y_i$. This time we locate a set J of s indices $j_1, \dots, j_s$, with $i < j_1 < \cdots < j_s \le i+s+R$ so that $\langle \{y_k : k \in [n] \setminus J\} \rangle = G$. If there is not enough room, i.e., $i > n-s-R$, then we locate J to lie in $\{n-s-R, \dots, n\} \setminus \{i\}$. In addition, if it happens that $i \le s$, then we take $J \subseteq \{s+1, \dots, 2s+R\}$. Now, to form $P(e)$:

(i) Use $g_1, \dots, g_s$ in $\phi(\bar{x})$ to change x_{j_1} to g_1, x_{j_2} to $g_2, \dots, x_{j_s}$ to g_s.

(ii) Use the newly formed $g_1, \dots, g_s$ (with indices in J) to change coordinate 1 from g_1 to y_1, coordinate 2 from g_2 to y_2,..., coordinate s from g_s to y_s. Then change x_i to y_i.

(iii) Finally use a basis in $\{y_k : k \notin [n] \setminus J\}$ to change coordinates $j_1, \dots, j_s$ to $y_{j_1}, \dots, y_{j_s}$, respectively. In this case, the length of $P(e)$ is at most $w(3s+1)$.

III) $\bar{\boldsymbol{x}} \notin \boldsymbol{G}^*, \bar{\boldsymbol{y}} \notin \boldsymbol{G}^*$. Thus, $\phi(\bar{x}) = (g_1, \ldots, g_s, x_{s+1}, \ldots, x_n)$, $\phi(\bar{y}) = (g_1, \ldots, g_s, y_{s+1}, \ldots, y_n)$ and of course, $x_j = y_j$, $j \neq i$, and $x_i \neq y_i$. If $i > s$ then we can change x_i to y_i in at most w steps (and this forms $P(e)$). If $i \leq s$ then $\phi(\bar{x}) = \phi(\bar{y})$ and $P(e)$ does not have to be defined.

The main point in the preceding slightly complicated construction is that it guarantees a rather small value of m. The reason is that the only coordinates u which change in edges of $P(e)$ are either in $\{1, 2, \ldots, s\}$ or fairly close to i, e.g. $|i - u| \leq 2(s + R)$. Furthermore, if a changing coordinate $u \in \{1, \ldots, s\}$ and $i > 2(s + R)$, so that we have some edge $e' = (z_1, \ldots, z_s, \ldots, z_u, \ldots, z_n), (z_1, \ldots, z_s, \ldots, z'_u, \ldots, z_n)$ in Γ^*, then we search $(z_1, \ldots, z_n)$ for the first interval of length $2(s + R)$ which contains $g_1, \ldots, g_s$, say $\{w + 1, \ldots, w + 2(s + R)\}$. By our construction, such an interval must exist.

Furthermore, it is not hard to see that in this case $|i - w| < 4(s + R)$ (and this is somewhat generous). Consequently, the original point $\bar{x}$ in e must agree with $(z_1, \ldots, z_n)$ in all but at most $10(s + R)$ coordinates. If follows that for these choices of ϕ and $P(e)$, $e \in E^n$, we have

$$m \leq 10(s + R) r^{10(s+R)} . \tag{17}$$

Also by previous remarks, we have

$$\ell \leq w(3s + 1) . \tag{18}$$

Observe that $deg\, \Gamma^n = (r - 1)n$ and $deg\, \Gamma^* = 2n(n - 1)$.

Consequently, by (3′), (13′), (17) and (18) (after some simplifications)

$$\lambda'_1 \geq \frac{\lambda_1}{1600 R^2 r^{20R} n}, \quad \alpha' \geq \frac{\alpha}{1600 R^2 r^{20R} n} . \tag{19}$$

6. Putting it all together

The final pieces we need to bound $\Delta(t)$ in (10) are the values of $\lambda_1(\Gamma^n)$ and $\alpha(\Gamma^n)$. Fortunately, these are easy to derive since λ_1 and α behave very nicely under Cartesian products. In particular, we have

$$\lambda_1(\Gamma^n) = \alpha(\Gamma^n) = \frac{(r - 1)}{r} \cdot \frac{1}{n} . \tag{20}$$

Note that

$$vol(\Gamma^*) \leq vol(\Gamma^n) = |\Gamma^n| \cdot 2n(n - 1) .$$

Thus, by (10) we have

$$\Delta(t) \leq e^{2-c} \quad if \quad t \geq \frac{1}{2\alpha'} \log\log |\Gamma^n| + \frac{c}{\lambda'_1}$$

and, by (11) we have

$$\Delta_{TV}(t) \le e^{1-c} \quad if \quad t \ge \frac{1}{4\alpha'} \log\log|\Gamma^n| + \frac{c}{\lambda_1'} \ .$$

This implies

Theorem 1.

$$\Delta(t) \le e^{2-c} \quad if \quad t \ge 800R^2 r^{20R+1} n^2 (\log n + \log\log r + c) \ . \tag{21}$$

In other words, $c_r n^2 \log n$ steps are enough to force the distribution to be close to uniform, which is what we claimed in the Introduction. Also, we have

Theorem 2.

$$\Delta_{TV}(t) \le e^{1-c} \quad if \quad t \ge 400R^2 r^{20R+1} n^2 (\log n + \log\log r + c) \ . \tag{22}$$

7. Concluding remarks

Of course, the preceding techniques using comparison theorems can be applied to many other random walk problems of this general type. For example, one could restrict the preceding moves so that $p_i \to p_i p_j^{\pm 1}$ is only allowed if (i,j) belongs to some specified set (this determines an underlying digraph).

It is probably true that the correct answer in (21) is actually $c_r n \log n$ (this is conjectured in [6]). Some evidence in favor of this is our recent result in [4] that $O(n \log n)$ steps do suffice when $G = \mathbb{Z}_2$.

References

[1] David Aldous, talk at 1994 National IMS meeting.

[2] F. Celler, C. R. Leedham-Green, S. Murray, A. Niemeyer and E. A. Obrien, Generating random elements of a finite group, Comm. Alg., 23 (1995), 4931–4948.

[3] F. R. K. Chung, Spectral Graph Theory, Amer. Math. Soc., Providence, RI (to appear 1996).

[4] F. R. K. Chung and R. L. Graham, Stratified random walk on an n-cube, preprint.

[5] P. Diaconis, R. L. Graham and J. A. Morrison, Asymptotic analysis of a random walk on a hypercube with many dimensions, Random Structures and Algorithms, 1 (1990), 51–72.

[6] P. Diaconis and L. Saloffe-Coste, Walks on generating sets of group, Stanford Technical Report No. 481, July 1995, 36 pages.

[7] P. Diaconis and L. Saloffe-Coste, Walks on generating sets of abelian groups, Probab. Theory Relat. Fields, 105 (1996), 393-421.

[8] P. Diaconis and L. Saloffe-Coste, Walks on generating sets of groups, Stanford Technical Report No. 497, July 1996, 40 pages. Probab. Theory Relat. Fields, 105 (1996), 393-421.

[9] P. Diaconis and L. Saloffe-Coste, Logarithmic Sobolev inequalities for finite Markov chains (preprint 1995).

[10] P. Diaconis and M. Shashahani, Time to reach stationarity in the Bernoulli-Laplace diffusion model, SIAM J. Math. Anal., 18 (1987), 208–218.