Available online at www.sciencedirect.com

SCIENCE DIRECT®

Stochastic Processes and their Applications 114 (2004) 51–79

stochastic processes and their applications

www.elsevier.com/locate/spa

Modified logarithmic Sobolev inequalities for some models of random walk☆

Sharad Goel*

Center for Applied Mathematics, Cornell University, 657 Rhodes Hall, Ithaca, NY 14853, USA

Received 23 July 2003; received in revised form 16 March 2004; accepted 8 June 2004

Available online 1 July 2004

Abstract

Logarithmic Sobolev inequalities are a well-studied technique for estimating rates of convergence of Markov chains to their stationary distributions. In contrast to continuous state spaces, discrete settings admit several distinct log Sobolev inequalities, one of which is the subject of this paper. Here we derive modified log Sobolev inequalities for some models of random walk, including the random transposition shuffle and the top-random transposition shuffle on S_n, and the walk generated by 3-cycles on A_n. As an application, we derive concentration inequalities for these models.

Keywords: Markov chains; Logarithmic Sobolev inequalities

1. Introduction

Introduced in 1975 [18], logarithmic Sobolev inequalities can be used to estimate rates of convergence of Markov chains to their stationary distributions. While in $\mathbb{R}^n$ there are several equivalent formulations of the log Sobolev inequality, in discrete settings these formulations lead to distinct inequalities (see e.g. [7]). One such modification, considered in [17, 24, 30], is the topic of this paper.

In Section 2, we introduce the notation and review preliminary results relating logarithmic Sobolev inequalities to rates of convergence. As a first example, we discuss previous estimates for modified log Sobolev inequalities on the 2-point space (see e.g. [7]).

☆ This research was supported in part by NSF Grant DMS-0306194.

* Tel.: +1-6072554195; fax: +1-6072559860.

E-mail address: scg23@cornell.edu (S. Goel).

0304-4149/$ - see front matter

doi:10.1016/j.spa.2004.06.001

Section 3 presents the main results of this paper: modified logarithmic Sobolev inequalities for some models of random walk, including the random transposition shuffle and the top-random transposition shuffle on the symmetric group S_n, and the shuffle generated by 3-cycles on the alternating group A_n. As an application of these results we derive sharp bounds on rates of convergence. Previously, convergence results for these models had been obtained by Fourier analysis [10,13,26]. In this section we also show that a generic r-regular graph has modified log Sobolev constant much smaller than its spectral gap. After completing this work, it came to our attention that Gao and Quastel [17] had derived the modified log Sobolev inequality for the random transposition model.

Like log Sobolev inequalities, modified log Sobolev inequalities can be obtained via comparison chains. Section 4 outlines this method, and analyzes a perturbation of the top-random transposition shuffle that cannot be realized as a walk on a group, making it difficult to study by other methods.

It is well known that the Herbst argument shows that log Sobolev inequalities imply concentration inequalities (see e.g. [7,22]). As an application of our results, in Section 5 we present concentration inequalities for the models of random walk considered earlier.

The recent work on modified log Sobolev inequalities [7, 17] illustrates the fact that for non-diffusion Dirichlet forms, modified log Sobolev inequalities can give better results than the classical log Sobolev inequality. It is worth pointing out that the reason behind this does not seem well understood. There are, however, some drawbacks to the modified versions: First, they seem inadequate to control convergence in l^2; and second, the comparison techniques seem to be much more restricted.

2. Background

This section introduces the notation used throughout the paper, and reviews results relating Sobolev inequalities to convergence rates. After introducing the notation for Markov chains, we define mixing time, which intuitively is the time necessary for a chain to approach equilibrium. We then define Dirichlet forms and indicate how they are in turn used to define the spectral gap and the log Sobolev inequalities, two well-studied techniques for bounding mixing time. For more detailed coverage of this material, see [12, 28, 29].

Next we present the modified log Sobolev inequality, a discrete state-space variant of the log Sobolev inequality, which was previously considered in [17, 24, 30]. We recall that the modified log Sobolev constant controls entropy, which in turn controls mixing time. We also present two properties that the log Sobolev and the modified log Sobolev inequalities share, namely that they both behave well under products and that they satisfy similar difference equations.

Finally, we discuss modified log Sobolev inequalities on the 2-point space and related spaces, further considered in [7]. The asymmetric 2-point space is one of the simplest examples in which we can distinguish between the log Sobolev and modified log Sobolev inequalities.

2.1. Preliminaries

A Markov chain on a finite state space $\mathscr{X}$ can be identified with a kernel K satisfying

$$K(x,y) \geqslant 0, \qquad \sum_{y\in\mathscr{X}} K(x,y) = 1.$$

The associated Markov operator is defined by

$$Kf(x) = \sum_{y\in\mathscr{X}} f(y)K(x,y).$$

The iterated kernel K^n is defined by

$$K^n(x,y) = \sum_{z\in\mathscr{X}} K^{n-1}(x,z)K(z,y)$$

and can be interpreted as the probability of moving from x to y in exactly n steps. We say that a probability measure π on $\mathscr{X}$ is invariant with respect to K if

$$\sum_{x\in\mathscr{X}} \pi(x)K(x,y) = \pi(y).$$

That is, starting with distribution π and moving according to the kernel K leaves the distribution of the chain unchanged. Throughout, we will assume that K is irreducible: For each $x, y \in \mathscr{X}$ there is an n such that $K^n(x,y) > 0$. Under this assumption K has a unique invariant measure π and $\pi(x) > 0$ for $x \in \mathscr{X}$. It will be useful to further restrict ourselves to the case where (K,π) is reversible, that is $K = K^*$ is a self-adjoint operator on the Hilbert space $L^2(\pi)$. This is equivalent to requiring that the detailed balance condition holds

$$\frac{K(x,y)}{\pi(y)} = \frac{K(y,x)}{\pi(x)}.$$

The kernel K describes a discrete-time chain, which at each time step moves with distribution according to K. Alternatively, we can consider the continuous-time chain H_t, which waits an exponential time before moving. More precisely $H_t = EK_{N_t}$, where N_t has independent Poisson distribution with parameter t. The law of H_t is then given by

$$H_t(x,y) = \mathrm{e}^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} K^n(x,y).$$

In terms of Markov operators, this continuous-time process is associated with Markov semigroup

$$H_t = \mathrm{e}^{tL} \quad L = K - I,$$

where I is the identity operator.

In order to quantify a chain's distance from equilibrium we need to introduce a metric. Arguably the most natural and oft used choice is the total variation distance.

Definition 2.1. Let μ and π be two probability measures on the set $\mathscr{X}$. The total variation distance is

$$\|\mu - \pi\|_{TV} = \max_{A \subset \mathscr{X}} |\mu(A) - \pi(A)|.$$

Next we define the mixing time, a measure of how long it takes the chain to be close (in total variation) to equilibrium.

Definition 2.2. Define the mixing time τ for H_t by

$$\tau = \inf\left\{ t > 0 : \sup_x \|H_t(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \leqslant \frac{1}{\mathrm{e}} \right\}.$$

The constant e^{-1} is chosen for convenience but is essentially arbitrary since for $d(t) = \sup_x \|H_t(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}}$ we have $d(s+t) \leqslant 2d(s)d(t)$. Also note that $d(t)$ is a decreasing function of t. See [1] for details.

The models of random walk we examine in this paper are known to exhibit the cutoff phenomenon: The total variation distance of the chain from equilibrium stays close to 1 for a long time, and then rapidly drops toward 0. Consequently, we can compare the mixing time bounds we derive to known cutoff times, and will show that in many cases these agree well.

Definition 2.3. Let $\mathscr{F} = (\mathscr{X}_n, K_n, \pi_n)$ be an infinite family of finite chains. Let $H_{n,t} = \mathrm{e}^{t(K_n - I)}$ be the corresponding continuous time chain. Then $\mathscr{F}$ presents a cutoff in total variation with critical time $\{t_n\}_1^\infty$ if $t_n \to \infty$, and for $\varepsilon > 0$

$$\lim_{n\to\infty} \max_{\mathscr{X}_n} \|H^x_{n,(1-\varepsilon)t_n} - \pi_n\|_{\mathrm{TV}} = 1$$

and

$$\lim_{n\to\infty} \max_{\mathscr{X}_n} \|H^x_{n,(1+\varepsilon)t_n} - \pi_n\|_{\mathrm{TV}} = 0.$$

2.2. Dirichlet forms and Sobolev inequalities

Our primary tool to investigate mixing times will be the Dirichlet form, defined for a finite Markov chain (K,π) as

$$\mathscr{E}(f,g) = E_\pi\,[f(x)\cdot(I-K)g(x)].$$

In the case of reversible (K,π), we have the important equivalent definition

$$\mathscr{E}(f,g) = \frac{1}{2}\,E_\pi\left[\sum_{y\in X} (f(x)-f(y))(g(x)-g(y))K(x,y)\right].$$

When $g = f$, this equivalent definition allows us to write the Dirichlet form as a sum of non-negative terms.

Inequalities involving Dirichlet forms have provided useful quantitative results in finite Markov chain theory, including spectral gap and logarithmic Sobolev bounds.

Definition 2.4. For (K,π) a Markov chain with Dirichlet form $\mathscr{E}$, the spectral gap λ is defined by

$$\lambda = \min\left\{\frac{\mathscr{E}(f,f)}{\mathrm{Var}_\pi(f)}; \mathrm{Var}_\pi(f) \neq 0\right\},$$

where $\mathrm{Var}_\pi(f)$ denotes the variation of $f : E(f-Ef)^2$.

While in the reversible case λ is the smallest non-zero eigenvalue of $I-K$, in general λ is the smallest non-zero eigenvalue of $I-\frac{1}{2}(K+K^*)$. The following lemma shows that the spectral gap controls the mixing time; details can be found in [28].

Lemma 2.1. *Let $H_t^x(y)=H_t(x,y)$ where $H_t(x,y)$ is the kernel of the continuous time chain H_t associated with (K,π). If λ is the spectral gap for (K,π) then*

$$\|H_t^x - \pi\|_{\mathrm{TV}}^2 \leqslant \frac{1}{4}\frac{1}{\pi(x)}\cdot \mathrm{e}^{-2\lambda t}.$$

In particular, if $\pi_* = \min_x \pi(x)$ then the mixing time satisfies

$$\tau \leqslant \frac{1}{2\lambda}\left(\log\frac{1}{\pi_*}+1\right).$$

Introduced in [18] to study Markov semigroups in infinite dimensional settings, log Sobolev inequalities also play a role in the theory of finite Markov chains. A comprehensive overview of log Sobolev inequalities can be found in [19], while [12] develops the theory for finite chains. Below we recall some results that motivate the definition of modified logarithmic Sobolev constant that is the subject of this paper.

Definition 2.5. The entropy of a non-negative function f on $\mathscr{X}$ with respect to π is

$$\mathrm{Ent}_\pi(f) = E\left[f\log\frac{f}{Ef}\right].$$

For an arbitrary function f, we will use the notation

$$\mathscr{L}_\pi(f) = \mathrm{Ent}_\pi(f^2).$$

Observe that by Jensen's inequality applied to the function $\phi(t)=t\log t$, $\mathscr{L}(f)\geqslant 0$ and $\mathscr{L}=0$ if and only if f is constant. $\mathscr{L}(f)$ can be seen as variation of $\mathrm{Var}(f)$, and accordingly the logarithmic Sobolev constant is defined analogously to the spectral gap, where the $\mathrm{Var}(f)$ is replaced by $\mathscr{L}(f)$.

Definition 2.6. For a Markov chain (K,π) with Dirichlet form $\mathscr{E}$, the logarithmic Sobolev constant β is defined by

$$\beta = \min\left\{\frac{\mathscr{E}(f,f)}{\mathscr{L}_\pi(f)}; \mathscr{L}_\pi(f) \neq 0\right\}.$$

From the definition, it follows that β is the largest constant c such that the logarithmic Sobolev inequality

$$c\mathscr{L}(f) \leqslant \mathscr{E}(f,f)$$

holds for all functions f. It is well known that $2\beta \leqslant \lambda$ (see e.g. [12]).

The following results show that the log Sobolev constant bounds entropy, which in turn bounds total variation distance. We include the proof of Lemma 2.2 since it motivates our use of the modified logarithmic Sobolev constant; proofs for Lemma 2.3 and Corollary 2.1 can be found in [12, 28].

Lemma 2.2. *Let β be the log Sobolev constant for the reversible chain (K,π). Then for $f \geqslant 0$*

$$\mathrm{Ent}(H_t f) \leqslant \mathrm{e}^{-4\beta t}\mathrm{Ent}(f).$$

Proof. W.l.o.g. we may assume $\pi(f) = 1$. Then $\pi(H_t f) = 1$ and since $H_t^* = H_t$,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Ent}(H_t f) &= \frac{\mathrm{d}}{\mathrm{d}t}\sum_x H_t f(x)\log H_t f(x)\pi(x)\\
&= \sum_x [LH_t f(x)\cdot \log H_t f(x) + LH_t f(x)]\pi(x)\\
&= \sum_x [LH_t f(x)\cdot \log H_t f(x)]\pi(x)\\
&= -\mathscr{E}(H_t f, \log H_t f)\\
&\leqslant -4\mathscr{E}((H_t f)^{1/2}, (H_t f)^{1/2})\\
&\leqslant -4\beta\,\mathrm{Ent}(H_t f). \qquad (2.1)
\end{aligned}$$

Inequality (2.1) follows from the fact that for reversible chains

$$\forall f \geqslant 0, \quad \mathscr{E}(\log f, f) \geqslant 4\mathscr{E}(\sqrt{f}, \sqrt{f}) \qquad (2.2)$$

(see e.g. [12]). Using Gronwall's lemma, the statement is proved. □

Lemma 2.3. *Let π and $\mu = h\pi$ be two probability measures on a finite set $\mathscr{X}$. Then*

$$\|\mu - \pi\|_{\mathrm{TV}}^2 = \frac{1}{4}\|h-1\|_{L^1(\pi)}^2 \leqslant \frac{1}{2}\,\mathrm{Ent}_\pi(h).$$

Corollary 2.1. *Let $H_t^x(y) = H_t(x,y)$ where $H_t(x,y)$ is the kernel of the continuous time semigroup H_t associated with the reversible chain (K,π). If β is the log Sobolev constant for (K,π) then*

$$\|H_t^x - \pi\|_{\mathrm{TV}}^2 \leqslant \frac{1}{2}\log\frac{1}{\pi(x)}\cdot \mathrm{e}^{-4\beta t}.$$

In particular, if $\pi_* = \min_x \pi(x)$ *then the mixing time satisfies*

$$\tau \leqslant \frac{1}{4\beta}\left(\log\log\frac{1}{\pi_*}+2\right).$$

2.3. A modified logarithmic Sobolev inequality

Examining the proof of Lemma 2.2, we see that we use two key inequalities: The first, (2.1), follows from (2.2), while the second results from the log Sobolev inequality

$$\beta\mathscr{L}(f) \leqslant \mathscr{E}(f,f).$$

This observation motivates the definition of a modified log Sobolev inequality (see e.g. [7, 24, 30]).

Definition 2.7. For a reversible Markov chain (K,π) with Dirichlet form $\mathscr{E}$, the modified logarithmic Sobolev constant α is defined by

$$\alpha = \min\left\{\frac{\mathscr{E}(f^2,\log f^2)}{\mathscr{L}_\pi(f)};\, \mathscr{L}_\pi(f)\neq 0\right\}.$$

Modified log Sobolev inequalities have recently been studied in several settings: In [30], modified log Sobolev inequalities were found for Poisson measures on $\mathbb{N}$; and [24] derives them for birth and death process on $\mathbb{Z}$. For a discussion of several different discrete modifications of the log Sobolev inequality, see e.g. [2, 3, 5–7].

We have the following well known result relating the log Sobolev constant, the modified log Sobolev constant and the spectral gap.

Lemma 2.4. *For a reversible chain* (K,π) *the log Sobolev constant* β, *the modified log Sobolev constant* α *and the spectral gap* λ *satisfy*

$$4\beta \leqslant \alpha \leqslant 2\lambda.$$

The first inequality follows from (2.2); for a proof of the second see [24]. From the definition we see that α is the largest constant c such that the modified log Sobolev inequality

$$c\mathscr{L}(f) \leqslant \mathscr{E}(f^2,\log f^2)$$

holds for all functions f. Consequently, as in the case of the log Sobolev inequality, α controls entropy, and in turn mixing time.

Corollary 2.2. *Let* $H_t^x(y) = H_t(x,y)$ *where* $H_t(x,y)$ *is the kernel of the continuous time semigroup* H_t *associated with the reversible chain* (K,π). *If* α *is the modified log Sobolev constant for* (K,π) *then*

$$\|H_t^x-\pi\|_{\mathrm{TV}}^2 \leqslant \frac{1}{2}\log\frac{1}{\pi(x)}\cdot \mathrm{e}^{-\alpha t}.$$

In particular, if $\pi_* = \min_x \pi(x)$ *then the mixing time satisfies*

$$\tau \leqslant \frac{1}{\alpha}\left(\log\log\frac{1}{\pi_*} + 2\right).$$

The modified log Sobolev constant is a phenomenon of the discrete state space. Let $\mathrm{d}\mu(x) = w(x)\mathrm{d}x$ be a probability measure on $\mathbb{R}^n$ with a smooth strictly positive density w. Then the continuous analog of the previously defined discrete Dirichlet form is,

$$\mathscr{E}(f,g) = \int_{\mathbb{R}^n} \nabla f(x)\cdot\nabla g(x)\mathrm{d}\mu(x),$$

where ∇ is the usual gradient. In this setting, since we have a chain rule,

$$\mathscr{E}(f^2, \log f^2) = 4\mathscr{E}(f,f).$$

On discrete state spaces, (2.2) shows that we have only inequality, suggesting that in this setting α and β may differ. However, given that they are indistinguishable on $\mathbb{R}^n$, it is surprising that we do in fact find examples where $\alpha \gg \beta$.

2.3.1. Elementary properties

The modified log Sobolev and log Sobolev inequalities share several properties, two of which we state here. The first shows that the modified log Sobolev inequality behaves well under products, and the second shows that solutions to the modified log Sobolev inequality satisfy a certain difference equation. Proofs of Lemma 2.5 and Theorem 2.1 are omitted since they are analogous to the proofs for the corresponding statements for the log Sobolev inequality given in [12, 28].

Lemma 2.5 (Product chains). *Let* (K_i, π_i), $i = 1,\dots,d$, *be reversible Markov chains on finite sets* $\mathscr{X}_i$ *with modified log Sobolev constants* α_i. *Fix* $\mu = (\mu_i)_i^d$ *such that* $\mu_i > 0$ *and* $\sum \mu_i = 1$. *Then the product chain* (K,π) *on* $\mathscr{X} = \prod_i^d \mathscr{X}_i$ *with kernel*

$$K(x,y) = \sum_{i=1}^{d} \mu_i \delta(x_1,y_1)\dots\delta(x_{i-1},y_{i-1})K_i(x_i,y_i)\delta(x_{i+1},y_{i+1})\dots\delta(x_d,y_d)$$

(*where* $\delta(x,x) = 1$ *and* $\delta(x,y) = 0$ *for* $x \neq y$) *and stationary measure* $\pi = \otimes\pi_i$ *satisfies*

$$\alpha = \min_i \mu_i\alpha_i.$$

Theorem 2.1. *Let* (K,π) *be a reversible Markov chain with modified log Sobolev constant* α *and spectral gap* λ. *Then either* $\alpha = 2\lambda$ *or there exists a positive, non-constant function* u *which is a solution of*

$$u^2\log u^2 - u^2\log\|u\|_2^2 - \frac{1}{\alpha}u^2(I-K)\log u^2 - \frac{1}{\alpha}(I-K)u^2 = 0 \tag{2.3}$$

and satisfies

$$\alpha\mathscr{L}(u) = \mathscr{E}(u^2, \log u^2).$$

In particular, if K *is irreducible, then* $\alpha > 0$.

2.3.2. First examples: The 2-point space and related spaces

The symmetric walk on the 2-point space $\mathscr{X} = \{x_1, x_2\}$ is perhaps the simplest of all Markov chains. The kernel K is given by

$$K(x_1,x_2) = K(x_2,x_1) = 1,$$

$$K(x_1,x_1) = K(x_2,x_2) = 0$$

and the stationary measure π is uniform. In [18], it is shown that the log Sobolev inequality satisfies $\beta = 1$. A trivial computation shows that the spectral gap $\lambda = 2$. Consequently, by Lemma 2.4, the modified log Sobolev constant satisfies $\alpha = 4$. By Lemma 2.5 and the fact that both the spectral gap and the log Sobolev constant are also well-behaved under products (see e.g. [28]), the walk on the n-dimensional hypercube has $4\beta = \alpha = 2\lambda = 4/n$.

A generalization of the walk on the 2-point space is the complete walk on n-points, addressed in the following lemma:

Lemma 2.6 (The complete walk). *Consider the Markov chain on the n point space $X = \{x_1, \ldots, x_n\}$ with uniform kernel $U(x_i, x_j) = 1/(n-1)$ for $x_i \neq x_j$ and $U(x_i, x_i) = 0$. For $n \geqslant 2$, the modified log Sobolev constant α_n satisfies*

$$\frac{n}{n-1} \leqslant \alpha_n \leqslant \left(1 + \frac{4}{\log(n+1)}\right) \frac{n}{n-1}.$$

Proof. Since the chain has the uniform stationary distribution $\pi(x_i) = 1/n$, we have

$$\begin{aligned}
\mathscr{E}(f^2, \log f^2) &= \frac{1}{2n(n-1)} \sum_{i,j=1}^{n} [f^2(x_i) - f^2(x_j)][\log f^2(x_i) - \log f^2(x_j)] \\
&= \frac{n}{n-1} (E[f^2 \log f^2] - Ef^2 \cdot E \log f^2) \\
&= \frac{n}{n-1} \left(E\left[f^2 \log \frac{f^2}{Ef^2} \right] - Ef^2 \cdot E \log \frac{f^2}{Ef^2} \right).
\end{aligned}$$

By Jensen's Inequality $E \log(\frac{f^2}{Ef^2}) \leqslant 0$ and the lower bound is established. For the upper bound, take f with $f^2(x_1) = n+1$ and $f^2(x_i) = 1$ for $2 \leqslant x_i \leqslant n$. $\square$

Diaconis and Saloff-Coste [12] prove that for the complete walk the log Sobolev constant satisfies

$$\beta = \frac{1 - 2/n}{\log(n-1)}$$

showing that for this example, $\alpha \gg \beta$.

An alternative generalization of the symmetric walk on the 2-point space is the asymmetric walk of Corollary 2.3.

Corollary 2.3 (Weighted 2-point space). *Consider the Markov chain on the two point space* $\mathcal{X}_2 = \{x_1, x_2\}$ *with kernel* $K(x_i, x_1) = \rho$ *and* $K(x_i, x_2) = 1 - \rho$ *with* $i = 1, 2$ *and* $0 < \rho \leqslant \frac{1}{2}$. *Then the modified log Sobolev constant satisfies* $1 \leqslant \alpha \leqslant 2$.

Proof. (K, π) is a reversible chain with stationary distribution $\pi(x_1) = \rho$, $\pi(x_2) = 1 - \rho$. First we establish the lower bound, observing that it is sufficient to restrict our attention to functions with $Ef = 1$. Consider rational ρ and write $\rho = p/q$ for integer p, q. Since we can identify functions f on $\mathcal{X}_2$ with functions $\tilde{f}$ on $\mathcal{X}_q = \{x_1, \ldots, x_q\}$ that are constant on the subsets $\{x_1, \ldots, x_p\}$ and $\{x_{p+1}, \ldots, x_q\}$, Lemma 2.6 shows that

$$\rho f^2(x_1) \log f^2(x_1) + (1 - \rho) f^2(x_2) \log f^2(x_2)$$
$$\leqslant \rho(1 - \rho)[f^2(x_1) - f^2(x_2)][\log f^2(x_1) - \log f^2(x_2)].$$

The result for irrational ρ follows by holding f fixed and taking the limit as $\rho_n \to \rho$ for rational $\{\rho_n\}$. The upper bound follows from the fact that the spectral gap $\lambda = 1$, and Lemma 2.4. □

For the asymmetric walk, the log Sobolev constant was calculated in [12] (and also independently in [20]) and shown to satisfy

$$\beta = \frac{1 - 2\rho}{\log[(1 - \rho)/\rho]},$$

again exemplifying the difference between α and β. For a further discussion of the asymmetric walk, see [7].

3. Examples of modified logarithmic Sobolev inequalities

In this section we derive modified log Sobolev inequalities for some models of random walk, including the random transposition shuffle and the top-random transposition shuffle on the symmetric group S_n, and the walk generated by 3-cycles on the alternating group A_n. These results are used to deduce sharp bounds on mixing times. We also show that a generic r-regular graph has modified log Sobolev constant much smaller than its spectral gap.

3.1. Random transposition and related walks

The random transposition walk on the symmetric group S_n is a shuffle on a deck of n cards where we uniformly at random select and swap pairs of cards. The log Sobolev constant for this walk was determined in [23] to satisfy $\beta \asymp \frac{1}{n \log n}$, and with respect to Corollary 2.1 is inadequate to sharply bound the mixing time [13]. Using the method of [23], Theorem 3.1 bounds the modified log Sobolev constant for walks including and related to random transposition. In contrast to the log Sobolev constant, our estimate of the modified log Sobolev constant for random transposition is sufficiently strong to yield the correct mixing time. After this work was completed, it came to our attention that Gao and Quastel [17] had proven Theorem 3.1 for the case of random transposition.

Let $G_n \subset S_n$ be subgroups of the symmetric group, generated by the symmetric sets $\mathscr{C}_n \subset G_n$. Then we have associated random walks given by the kernel,

$$K_n(\tau,\tau') = \begin{cases} \frac{1}{|\mathscr{C}_n|} \ \tau' = \tau \cdot \sigma & \text{for some } \sigma \in \mathscr{C}_n, \\ 0 & \text{otherwise.} \end{cases}$$

The stationary distribution π is uniform on G_n and the Dirichlet form for (K_n,π) is given explicitly by

$$\mathscr{E}_n(f,g) = \frac{1}{2|\mathscr{C}_n|} E\left[\sum_{\tau' \in \mathscr{C}_n} [f(\tau) - f(\tau\cdot\tau')][g(\tau) - g(\tau\cdot\tau')]\right].$$

For $\tau \in S_n$, we let τ_i denote the particle in position i, and so $\tau \cdot \sigma$ denotes the configuration after we permute the positions according to σ. If this Markov chain has enough symmetry, Theorem 3.1 gives a bound on the modified log Sobolev constant α.

Definition 3.1 (Self-similarity). A sequence of groups $G_n \subset S_n$ with symmetric generating sets $\mathscr{C}_n$ is called self-similar if:

(1) For $1 \leqslant s \leqslant n$, there exist isomorphisms $g_s^{n-1} : G_{n-1} \to \{\sigma \in G_n | \sigma_s = s\}$, with $g_s^{n-1}(\mathscr{C}_{n-1}) = \{\sigma \in \mathscr{C}_n | s \notin \text{supp}(\sigma)\}$.
(2) G_n acts transitively on the set $\{1,\dots,n\}$.
(3) There exists k, such that for all n and $\sigma \in \mathscr{C}_n$, $|\text{supp}(\sigma)| = k$, where $\text{supp}(\sigma) = \{i | \sigma_i \neq i\}$.

Definition 3.1 encompasses a collection of random walks including random transposition on S_n and the walk generated by 3-cycles on the alternating group A_n. In general consider a sequence of random walks generated by conjugacy classes of S_n. Recall, that for $n \neq 4$, a non-trivial conjugacy class $\mathscr{C}_n$ generates either the alternating group A_n or S_n. For a permutation $\tau \in S_n$, let $c(\tau) = (c_1,\dots,c_n)$ denote the cycle structure of τ. That is, c_i is the number of cycles of length i in the disjoint cycle decomposition of τ. Then, two permutations are conjugate if and only if their cycle structure is the same. Now, for a conjugacy class $\mathscr{C}_{n_0}$ of S_{n_0} (respectively A_{n_0}) with corresponding cycle structure $c^{n_0} = (c_1^{n_0},\dots,c_{n_0}^{n_0})$, define a sequence of conjugacy classes $\mathscr{C}_n$ for $n > n_0$ corresponding to the cycle structures $c^n = (c_1^n, c_2^{n_0},\dots,c_{n_0}^{n_0},0,\dots,0)$ where $c_1^n = n - \sum_{i=2}^{n_0} i c_i^{n_0}$. Then this sequence of walks is self-similar.

For $1 \leqslant s \leqslant n$, let $\sigma_s : S_n \to \{x_1,\dots,x_n\}$ be the random variable that takes $\tau \to \tau_s$. The idea behind the proof of Theorem 3.1 is to first condition on each σ_s. Then we break up $\mathscr{L}(f)$ into two parts: The first we bound by the Dirichlet form on S_{n-1} where we have fixed the sth position to hold particle σ_s. The second we bound by looking at the complete walk described in Lemma 2.6 with stationary measure corresponding to the distribution of σ_s (i.e. the uniform distribution on $\{1,\dots,n\}$). By averaging over s, we can pass from the Dirichlet forms on S_{n-1} to S_n. This then gives us a recurrence relation between the modified log Sobolev constants, yielding the result.

Theorem 3.1. *Let $\mathscr{C}_n \subset G_n$ be self-similar for $n \geqslant n_0$, and consider the sequence of walks generated as above. Then, if a_n denotes the reciprocal of the modified log Sobolev constant for these chains,*

$$a_n \leqslant a_{n_0} + (n - n_0).$$

Proof. To begin we fix a function $f : S_n \to \mathscr{R}$, with $f > 0$. By homogeneity, it is sufficient to establish the modified log Sobolev inequality for f with $\pi(f^2) = 1$. Define

$$f_s(x) = E[f^2(\cdot)|\sigma_s = x]^{1/2},$$

$$f_s(\tau|x) = \frac{f(\tau)\delta_x(\tau_s)}{f_s(x)},$$

where δ_x is the dirac point mass at x. Let

$$I_{1,s} = E\left\{ f_s^2(\sigma_s) E\left[\sum_{\substack{\tau' \in \mathscr{C}_n \\ \operatorname{supp}(\tau') \not\ni s}} [f_s^2(\tau|\sigma_s) - f_s^2(\tau \cdot \tau'|\sigma_s)] \right.\right.$$

$$\left.\left. \times [\log f_s^2(\tau|\sigma_s) - \log f_s^2(\tau \cdot \tau'|\sigma_s)] \Big| \, \sigma_s \right] \right\}$$

$$= E\left\{ f_s^2(\sigma_s) E\left[\sum_{\substack{\tau' \in \mathscr{C}_n \\ \operatorname{supp}(\tau') \not\ni s}} \left[\frac{f^2(\tau)}{f_s^2(\sigma_s)} - \frac{f^2(\tau \cdot \tau')}{f_s^2(\sigma_s)} \right] \right.\right.$$

$$\left.\left. \times \left[\log \frac{f^2(\tau)}{f_s^2(\sigma_s)} - \log \frac{f^2(\tau \cdot \tau')}{f_s^2(\sigma_s)} \right] \Bigg| \, \sigma_s \right] \right\}$$

$$= E\left(E\left[\sum_{\substack{\tau' \in \mathscr{C}_n \\ \operatorname{supp}(\tau') \not\ni s}} [f^2(\tau) - f^2(\tau \cdot \tau')] \, [\log f^2(\tau) - \log f^2(\tau \cdot \tau')] \right| \sigma_s \right] \right)$$

$$= E\left[\sum_{\substack{\tau' \in \mathscr{C}_n \\ \operatorname{supp}(\tau') \not\ni s}} [f^2(\tau) - f^2(\tau \cdot \tau')] \, [\log f^2(\tau) - \log f^2(\tau \cdot \tau')] \right].$$

And define

$$I_{2,s} = E[f_s^2(\sigma_s)\log f_s^2(\sigma_s)],$$

$$I_i = \frac{1}{n}\sum_{s=1}^{n} I_{i,s} \quad i = 1, 2.$$

To express $\mathscr{L}(f)$ in terms of the above definitions note that

$$\begin{aligned} f_s^2(\sigma_s)E[f_s^2(\tau|\sigma_s)\log f_s^2(\tau|\sigma_s)|\sigma_s] &= f_s^2(\sigma_s)E\left[\frac{f^2(\tau)}{f_s^2(\sigma_s)}\log\frac{f^2(\tau)}{f_s^2(\sigma_s)}\middle|\,\sigma_s\right] \\ &= E[f^2(\tau)\log f^2(\tau)|\sigma_s] - f_s^2(\sigma_s)\log f_s^2(\sigma_s). \end{aligned}$$

Taking expectations

$$\mathscr{L}(f) = E\{f_s^2(\sigma_s)E[f_s^2(\tau|\sigma_s)\log f_s^2(\tau|\sigma_s)|\sigma_s]\} + I_{2,s}. \tag{3.1}$$

Since $f_s^2(\tau|\sigma_s) = 0$ for $\tau_s \neq \sigma_s$, we can naturally consider $f_s(\cdot|\sigma_s)$ as a function on S_{n-1} (where we fix position s to hold particle σ_s). Specifically, let $h_{s\to\sigma_s} \in G_n$ be such that $h_{s\to\sigma_s}(s) = \sigma_s$, and define $\tilde{f}_s^2$ on G_{n-1} by

$$\tilde{f}_s^2(\tau) = f_s^2(h_{s\to\sigma_s} g_s^{n-1}(\tau)|\sigma_s).$$

Since $E_{n-1}\tilde{f}_s^2(\cdot) = 1$,

$$\begin{aligned} &E[f_s^2(\tau|\sigma_s)\log f_s^2(\tau|\sigma_s)|\sigma_s] \\ &\quad = E[\tilde{f}_s^2\log \tilde{f}_s^2] \\ &\quad \leqslant \frac{a_{n-1}}{2|\mathscr{C}_{n-1}|}E\left[\sum_{\tau'\in\mathscr{C}_{n-1}}[\tilde{f}_s^2(\tau) - \tilde{f}_s^2(\tau\cdot\tau')][\log\tilde{f}_s^2(\tau) - \log\tilde{f}_s^2(\tau\cdot\tau')]\right] \\ &\quad = \frac{a_{n-1}}{2|\mathscr{C}_{n-1}|}E\left[\sum_{\substack{\tau'\in\mathscr{C}_n \\ \mathrm{supp}(\tau')\not\ni s}}[f_s^2(\tau|\sigma_s) - f_s^2(\tau\cdot\tau'|\sigma_s)]\right. \\ &\qquad\qquad \left.\times[\log f_s^2(\tau|\sigma_s) - \log f_s^2(\tau\cdot\tau'|\sigma_s)]\middle|\,\sigma_s\right]. \end{aligned}$$

Applying this to (3.1) gives,

$$\mathscr{L}(f) \leqslant \frac{a_{n-1}}{2|\mathscr{C}_{n-1}|}I_{1,s} + I_{2,s}.$$

Averaging over s, we have

$$\mathscr{L}(f) \leqslant \frac{a_{n-1}}{2|\mathscr{C}_{n-1}|} I_1 + I_2. \tag{3.2}$$

Let $k = \text{supp}(\sigma)$ for $\sigma \in \mathscr{C}_n$. Then note that each term $[f^2(\tau) - f^2(\tau \cdot \tau')][\log f^2(\tau) - \log f^2(\tau \cdot \tau')]$ appears in I_1 exactly $n-k$ times. So we have

$$\begin{aligned} I_1 &= \frac{n-k}{n} E\left[\sum_{\tau' \in \mathscr{C}_n} [f^2(\tau) - f^2(\tau \cdot \tau')][\log f^2(\tau) - \log f^2(\tau \cdot \tau')]\right] \\ &= \frac{2(n-k)|\mathscr{C}_n|}{n} \mathscr{E}(f^2, \log f^2). \end{aligned}$$

Note that

$$\begin{aligned} |\mathscr{C}_n|(n-k) &= \sum_{\sigma \in \mathscr{C}_n} \sum_{i=1}^{n} 1_{\{i | i \notin \text{supp}(\sigma)\}} \\ &= \sum_{i=1}^{n} \sum_{\sigma \in \mathscr{C}_n} 1_{\{i | i \notin \text{supp}(\sigma)\}} \\ &= n|\mathscr{C}_{n-1}|. \end{aligned}$$

Substituting this into (3.2), we have

$$\mathscr{L}(f) \leqslant a_{n-1} \mathscr{E}(f^2, \log f^2) + I_2. \tag{3.3}$$

To bound I_2 we consider the Markov chain on state space $\{x_1, \ldots, x_n\}$ with uniform kernel $K(x_i, x_j) = 1/(n-1)$ for $i \neq j$. First note that since $|\{\tau | \tau_s = i\}| = |\{h_{s \to i}\{\tau | \tau_s = s\}| = |G_{n-1}|$ for all i, σ_s is uniformly distributed on $\{x_1, \ldots, x_n\}$. Furthermore,

$$\begin{aligned} E[f_\cdot^2(x_m)] &= \frac{1}{n} \sum_{s=1}^{n} E[f^2 | \sigma_s = x_m] \\ &= \sum_{s=1}^{n} E[f^2; \sigma_s = x_m] \\ &= 1. \end{aligned}$$

Consequently,

$$\begin{aligned} I_2 &= \frac{1}{n} \sum_{s=1}^{n} E[f_s^2(\cdot) \log f_s^2(\cdot)] \\ &= \frac{1}{n} \sum_{m=1}^{n} \mathscr{L}_U(f_\cdot(x_m)) \end{aligned}$$

$$\leqslant \frac{n-1}{n^2}\sum_{m=1}^{n}\mathscr{E}(f_\cdot^2(x_m), \log f_\cdot^2(x_m)) \quad \text{by Lemma 2.6}$$

$$= \frac{1}{2n^3}\sum_{m=1}^{n}\sum_{i\neq j}[f_i^2(x_m)-f_j^2(x_m)][\log f_i^2(x_m)-\log f_j^2(x_m)].$$

Now, for $\tau'\in\mathscr{C}_n$ such that $\tau'_i=j$

$$\begin{aligned}
&[f_i^2(x_m)-f_j^2(x_m)][\log f_i^2(x_m)-\log f_j^2(x_m)]\\
&\quad = [E[f^2(\tau)|\sigma_i=x_m]-E[f^2(\tau)|\sigma_j=x_m]]\\
&\qquad \times[\log E[f^2(\tau)|\sigma_i=x_m]-\log E[f^2(\tau)|\sigma_j=x_m]]\\
&\quad = [E[f^2(\tau\cdot\tau')|\sigma_j=x_m]-E[f^2(\tau)|\sigma_j=x_m]]\\
&\qquad \times[\log E[f^2(\tau\cdot\tau')|\sigma_j=x_m]-\log E[f^2(\tau)|\sigma_j=x_m]]\\
&\quad \leqslant E[[f^2(\tau)-f^2(\tau\cdot\tau')][\log f^2(\tau)-\log f^2(\tau\cdot\tau')]|\sigma_j=x_m]\\
&\quad = nE[f^2(\tau)-f^2(\tau\cdot\tau')][\log f^2(\tau)-\log f^2(\tau\cdot\tau');\sigma_j=x_m].
\end{aligned}$$

Above we have used Jensen's Inequality with $g(x,y)=(x-y)(\log x-\log y)$ (which is convex for $x,y>0$). Let $\mathscr{C}_{i\to j}\subset\mathscr{C}_n$ consist of those τ' with $\tau'_i=j$. Then averaging over $\tau'\in\mathscr{C}_{i\to j}$, we get

$$\begin{aligned}
&[f_i^2(x_m)-f_j^2(x_m)][\log f_i^2(x_m)-\log f_j^2(x_m)]\\
&\quad \leqslant \frac{n}{|\mathscr{C}_{i\to j}|}\sum_{\tau'\in\mathscr{C}_{i\to j}}E[f^2(\tau)-f^2(\tau\cdot\tau')][\log f^2(\tau)-\log f^2(\tau\cdot\tau');\sigma_j=x_m].
\end{aligned}$$

And

$$\begin{aligned}
n(n-1)|\mathscr{C}_{i\to j}| &= \sum_{i\neq j}\sum_{\sigma\in\mathscr{C}_n}1_{\{\sigma_i=j\}}\\
&= \sum_{\sigma\in\mathscr{C}_n}\sum_{i\neq j}1_{\{\sigma_i=j\}}\\
&= |\mathscr{C}_n|k
\end{aligned}$$

yields,

$$\begin{aligned}
I_2 &\leqslant \frac{n-1}{2nk|\mathscr{C}_n|}\sum_{m=1}^{n}\sum_{i\neq j}\sum_{\tau'\in\mathscr{C}_{i\to j}}E[f^2(\tau)-f^2(\tau\cdot\tau')]\\
&\qquad \times[\log f^2(\tau)-\log f^2(\tau\cdot\tau');\sigma_j=x_m]\\
&= \frac{n-1}{2nk|\mathscr{C}_n|}\sum_{i\neq j}\sum_{\tau'\in\mathscr{C}_{i\to j}}E[f^2(\tau)-f^2(\tau\cdot\tau')][\log f^2(\tau)-\log f^2(\tau\cdot\tau')]
\end{aligned}$$

$$= \frac{n-1}{2n|\mathscr{C}_n|} \sum_{\tau' \in \mathscr{C}_n} E[f^2(\tau) - f^2(\tau \cdot \tau')][\log f^2(\tau) - \log f^2(\tau \cdot \tau')]$$

$$= \frac{n-1}{n} \mathscr{E}(f^2, \log f^2).$$

Using (3.3), the result follows from the recurrence,

$$a_n \leqslant a_{n-1} + \frac{n-1}{n}. \qquad \square$$

Corollary 3.1 (Random transposition). *Consider the random transposition walk on S_n, i.e. the walk generated by $\mathscr{C}_n = \{(i,j) | 1 \leqslant i < j \leqslant n\}$ for $n \geqslant 2$. Then the modified log Sobolev constant α satisfies*

$$\frac{4}{n-1} \geqslant \alpha \geqslant \frac{1}{n-1}.$$

In particular, the mixing time satisfies $\tau \leqslant n(\log\log n! + 2)$.

Remark 3.1. In [27] it is shown that for $t = \frac{n-1}{2}(\log(n-1) - c)$

$$\|H_t^x - \pi\|_{\mathrm{TV}} \geqslant 1 - 8\mathrm{e}^{-2c} - 4\mathrm{e}^{-c} - \frac{4\log(n-1)}{n-1}\mathrm{e}^{(2\log n)/n}$$

and results in [13] show that this bound is sharp. Consequently, the mixing time bound of Corollary 3.1 is within a factor of 2 of the critical time.

Proof. Since the walk on S_2 is the symmetric walk on the 2-point space, the discussion in Section 2.3 shows that $a_2 = \frac{1}{4}$, yielding the lower bound. This chain is studied in detail in [9], where it is shown that the spectral gap satisfies $\lambda = 2/(n-1)$. The upper bound is then a consequence of Lemma 2.4. The mixing time follows from Corollary 2.2. $\square$

Corollary 3.2 (3-cycles). *For the random walk on A_n generated by* 3*-cycles, the modified log Sobolev constant satisfies*

$$\frac{6}{n-1} \geqslant \alpha \geqslant \frac{1}{n-2}.$$

In particular, the mixing time satisfies $\tau \leqslant n(\log\log n! + 2)$.

Remark 3.2. In [25, 26] it is shown that the above walk has cutoff with critical time $t_n = (n/3)\log n$.

Proof. The walk on A_3 is the uniform walk on the 3-point space. Consequently, by Lemma 2.6, $a_3 \leqslant 1$, yielding the lower bound. The upper bound follows from results in [25, 26] that $\lambda = 3/(n-1)$, and Lemma 2.4. $\square$

Example 3.1 (Bernoulli–Laplace). Informally, the Bernoulli–Laplace (BL) model is a random transposition walk on S_n with n distinct sites and $1 \leqslant r \leqslant n-1$ identical particles, with each site occupied by at most one particle. The state space $C_{n,r}$ is the set of r-subsets of $\{1,\ldots,n\}$, and accordingly is of order $\binom{n}{r}$. For $\eta \in C_{n,r}$ let η_i denote the number of particles in site i, so η_i is either 0 or 1. Let η^{ij} denote the configuration in which we have swapped the particles in positions i and j. Then the kernel for this chain is given by

$$K(\eta,\eta') = \begin{cases} \frac{1}{r(n-r)} & \eta' = \eta^{ij}, \qquad \eta_i = (1-\eta_j) = 1. \\ 0 & \text{otherwise,} \end{cases}$$

The log Sobolev constant for this walk was found in [23] to satisfy

$$\beta_{n,r} \asymp \frac{n}{r(n-r)} \log \frac{n^2}{r(n-r)}$$

and in [17] the authors used the method of Theorem 3.1 to directly show that modified log Sobolev constant for BL satisfies $\alpha_{n,r} \asymp \frac{n}{r(n-r)}$. We can find this same bound on the modified log Sobolev constant by relying on our analysis of the random transposition walk.

To analyze this chain, map functions f on $C_{n,r}$ to functions $\tilde{f}$ on S_n by letting $\tilde{f}(\sigma) = f(\{\sigma_1,\ldots,\sigma_r\})$. For $\eta \in C_{n,r}$ let $\bar{\eta} \in S_n$ be any permutation such that $\eta = \{\bar{\eta}_1,\ldots,\bar{\eta}_r\}$. Note that there are $r!(n-r)!$ such permutations and that $f(\eta^{ij}) = \tilde{f}(\bar{\eta}^{ij})$. Therefore,

$$\begin{aligned} \mathscr{E}(f,g) &= \frac{1}{2r(n-r)} E_\pi \left[\sum_{\substack{i,j \\ \eta_i=(1-\eta_j)=1}} [f(\eta)-f(\eta^{ij})][g(\eta)-g(\eta^{ij})] \right] \\ &= \frac{1}{4r(n-r)} \frac{r!(n-r)!}{n!} \left[\sum_{i,j} [f(\eta)-f(\eta^{ij})][g(\eta)-g(\eta^{ij})] \right] \\ &= \frac{1}{4r(n-r)} \frac{1}{n!} \left[\sum_{i,j} [\tilde{f}(\sigma)-\tilde{f}(\sigma^{ij})][\tilde{g}(\sigma)-\tilde{g}(\sigma^{ij})] \right] \\ &= \frac{n(n-1)}{2r(n-r)} \mathscr{E}'(\tilde{f},\tilde{g}), \end{aligned}$$

where $\mathscr{E}'(f,g)$ is the Dirichlet form associated with the random transposition model (K',π') of Corollary 3.1. Furthermore, since $\mathscr{L}(f) = \mathscr{L}'(\tilde{f})$, $\alpha_{n,r} \geqslant \frac{n(n-1)}{2r(n-r)} \alpha'_n$. By Corollary 3.1 and the fact that the spectral gap for Bernoulli–Laplace is given by $\lambda_{n,r} = \frac{n}{r(n-r)}$ [14], the modified log Sobolev constant for the BL model satisfies

$$\frac{2n}{r(n-r)} \geqslant \alpha_{n,r} \geqslant \frac{n}{2r(n-r)}.$$

By Corollary 2.2, the mixing time $\tau_{n,r}$ for the BL model satisfies

$$\tau_{n,r} \leqslant \frac{2r(n-r)}{n}\left(\log\log\binom{n}{r}+2\right).$$

3.2. Top-random transposition walk

A walk similar to those considered in Section 3.1 is the complete (k,l)-bipartite shuffle. We can visualize this walk on a deck of cards by first splitting the deck into two pieces—of size k and of size l—and then uniformly swapping pairs of cards between the piles. In the case $k=1$, we have the top-random transposition shuffle. That is, at each step we swap the top card and another chosen uniformly at random.

Using the same notation as above, for $\tau \in S_n$, we let τ_i denote the particle in position i, and let τ^{ij} denote the configuration after we swap the particles in positions i and j. The kernel for the (k,l)-complete bipartite shuffle is given by,

$$K(\tau,\tau') = \begin{cases} \frac{1}{kl}\ \tau' = \tau^{ij} & \text{for any } 1 \leqslant i \leqslant k < j \leqslant k+l, \\ 0 & \text{otherwise.} \end{cases}$$

The stationary distribution π is uniform and the Dirichlet form for (K,π) is given explicitly by

$$\mathscr{E}(f,g) = \frac{1}{2kl}E\left[\sum_{1\leqslant i\leqslant k<j\leqslant k+l}[f(\tau)-f(\tau^{ij})][g(\tau)-g(\tau^{ij})]\right].$$

As above, before computing the modified log Sobolev constant for the walk on S_n, we restrict our attention to the movement of one particle. In this case we have the walk on the complete (k,l)-bipartite graph: Our state space is the $k+l$ point space $\{x_1,\ldots,x_k,y_1,\ldots y_l\}$; the kernel is given by

$$\tilde{K}(x_i,y_j) = \frac{1}{k+l}, \quad 1 \leqslant i \leqslant k,\ 1 \leqslant j \leqslant l,$$

$$\tilde{K}(y_j,x_i) = \frac{1}{k+l}, \quad 1 \leqslant i \leqslant k,\ 1 \leqslant j \leqslant l,$$

$$\tilde{K}(x_i,x_i) = \frac{k}{l+k}, \quad 1 \leqslant i \leqslant k,$$

$$\tilde{K}(y_j,y_j) = \frac{l}{l+k}, \quad 1 \leqslant j \leqslant l$$

and zero otherwise. Then $(\tilde{K},\pi)$ is reversible with respect to the uniform stationary measure π.

Lemma 3.1. *For the random walk on the complete* (k,l)*-bipartite graph with* $l \geqslant 2$, *the modified log Sobolev constant satisfies* $\alpha \leqslant 2k/(k+l)$. *In the case of the star, i.e. the complete* $(1,l)$*-bipartite graph, we have the lower bound* $\alpha \geqslant 1/(l+1)$.

Proof. By explicitly computing the eigenvalues of $\tilde{K}$ we find the spectral gap $\lambda = k/(k+l)$. The upper bound for α then follows from Lemma 2.4.

To lower bound α for the star observe that

$$\mathscr{E}(f^2, \log f^2) = \frac{1}{(l+1)^2}\left[\sum_i [f^2(x_1) - f^2(y_i)][\log f^2(x_1) - \log f^2(y_i)]\right]$$
$$= \frac{1}{l+1} E[[f^2(x_1) - f^2(\cdot)][\log f^2(x_1) - \log f^2(\cdot)]].$$

By homogeneity, we only need to show the modified log Sobolev inequality for functions f with $f^2(x_1) = 1$. And in this case, the above simplifies to

$$\mathscr{E}(f^2, \log f^2) = \frac{1}{l+1}[-E \log f^2 + E f^2 \log f^2].$$

By Jensen's Inequality,

$$E \log f^2 \leqslant \log E f^2 \leqslant E f^2 \cdot \log E f^2.$$

Since $\mathscr{L}(f) = E f^2 \log f^2 - E f^2 \cdot \log E f^2$,

$$\mathscr{L}(f) \leqslant E f^2 \log f^2 - E \log f^2$$

and the lower bound is established. □

The proof of the following theorem is analogous to the proof of Theorem 3.1, the primary difference being that here we bound I_2 using the Markov chain on the star described in Lemma 3.1.

Theorem 3.2. *For $l \geqslant 2$, let $a_{k,l}$ denote the reciprocal of the modified log Sobolev constant for the complete (k,l)-bipartite walk on S_{k+l}, and let $\tilde{a}_{k,l}$ denote the reciprocal of the modified log Sobolev constant for the complete (k,l)-bipartite walk on $\{x_1, \ldots, x_k, y_1, \ldots, y_l\}$. Then*

$$a_{k,l} \leqslant a_{k,l-1} + \frac{2k}{k+l}\tilde{a}_{k,l}.$$

Proof. For $1 \leqslant s \leqslant k+l$, let $\sigma_s : S_n \to \{x_1, \ldots, x_n\}$ be the random variable that takes $\tau \to \tau_s$. To begin we fix a function $f : S_n \to \mathfrak{R}$, with $f > 0$. By homogeneity, it is sufficient to establish the modified log Sobolev inequality for f with $\pi(f^2) = 1$. Let

$$f_s(x) = E[f^2(\cdot)|\sigma_s = x]^{1/2},$$

$$f_s(\tau|x) = \frac{f(\tau)\delta_x(\tau_s)}{f_s(x)}.$$

And for $k < s \leqslant k+l$, define

$$I_{1,s} = E\left\{ f_s^2(\sigma_s) E\left[\sum_{\substack{1 \leqslant i \leqslant k < j \leqslant k+l \\ j \neq s}} [f_s^2(\tau|\sigma_s) - f_s^2(\tau^{ij}|\sigma_s)] \times [\log f_s^2(\tau|\sigma_s) - \log f_s^2(\tau^{ij}|\sigma_s)] \,\Big|\, \sigma_s \right] \right\}$$

$$= E\left[\sum_{\substack{1 \leqslant i \leqslant k < j \leqslant k+l \\ j \neq s}} [f^2(\tau) - f^2(\tau^{ij})]\,[\log f^2(\tau) - \log f^2(\tau^{ij})] \right]$$

$$I_{2,s} = E[f_s^2(\sigma_s) \log f_s^2(\sigma_s)]$$

$$I_i = \frac{1}{l} \sum_{s=k+1}^{k+l} I_{i,s}, \quad i = 1, 2.$$

As before,

$$\mathscr{L}(f) = E\{f_s^2(\sigma_s) E[f_s^2(\tau|\sigma_s) \log f_s^2(\tau|\sigma_s)|\sigma_s]\} + I_{2,s}$$

and, considering $f_s(\cdot|\sigma_s)$ as a function on S_{n-1},

$$\mathscr{L}_{k,l-1}(f_s(\cdot|\sigma_s)) \leqslant a_{k,l-1} \mathscr{E}_{k,l-1}(f_s^2(\cdot|\sigma_s), \log f_s^2(\cdot|\sigma_s)).$$

Consequently,

$$E[f_s^2(\tau|\sigma_s) \log f_s^2(\tau|\sigma_s)|\sigma_s]$$

$$\leqslant \frac{a_{k,l-1}}{2k(l-1)} E\left[\sum_{\substack{1 \leqslant i \leqslant k < j \leqslant k+l \\ j \neq s}} (f_s^2(\tau|\sigma_s) - f_s^2(\tau^{ij}|\sigma_s)) \times (\log f_s^2(\tau|\sigma_s) - \log f_s^2(\tau^{ij}|\sigma_s)) \,\Big|\, \sigma_s \right]$$

and

$$\mathscr{L}(f) \leqslant \frac{a_{k,l-1}}{2k(l-1)} I_{1,s} + I_{2,s}.$$

Averaging over s, we have

$$\mathscr{L}(f) \leqslant \frac{a_{k,l-1}}{2k(l-1)} I_1 + I_2. \tag{3.4}$$

Since each term $[f^2(\tau) - f^2(\tau^{ij})][\log f^2(\tau) - \log f^2(\tau^{ij})]$ appears in I_1 exactly $l-1$ times, we have

$$\begin{aligned} I_1 &= \frac{l-1}{l} E\left[\sum_{1\leqslant i\leqslant k<j\leqslant k+l} [f^2(\tau) - f^2(\tau^{ij})][\log f^2(\tau) - \log f^2(\tau^{ij})]\right] \\ &= 2k(l-1)\mathscr{E}(f^2, \log f^2). \end{aligned}$$

Substituting this into (3.4), we have

$$\mathscr{L}(f) \leqslant a_{k,l-1}\mathscr{E}(f^2, \log f^2) + I_2. \tag{3.5}$$

To bound I_2 we consider the Markov chain on the state spaces $\{x_1, \ldots, x_{k+l}\}$ with kernel given by the complete (k,l)-bipartite graph. Since σ_s is uniformly distributed on $\{1, \ldots, k+l\}$,

$$\begin{aligned} I_2 &= \frac{1}{l} \sum_{s=k+1}^{k+l} E[f_s^2(\cdot)\log f_s^2(\cdot)] \\ &\leqslant \frac{1}{l} \sum_{s=1}^{k+l} E[f_s^2(\cdot)\log f_s^2(\cdot)] \quad \text{since entropy is non-negative} \\ &= \frac{1}{l} \sum_{m=1}^{k+l} \mathscr{L}_U(f_{\cdot}(x_m)) \\ &\leqslant \frac{\tilde{a}_{k,l}}{l} \sum_{m=1}^{k+l} \mathscr{E}(f_{\cdot}^2(x_m), \log f_{\cdot}^2(x_m)) \\ &= \frac{\tilde{a}_{k,l}}{l(k+l)^2} \sum_{m=1}^{k+l} \sum_{1\leqslant i\leqslant k<j\leqslant k+l} [f_i^2(x_m) - f_j^2(x_m)][\log f_i^2(x_m) - \log f_j^2(x_m)]. \end{aligned}$$

Since

$$\begin{aligned} &[f_i^2(x_m) - f_j^2(x_m)][\log f_i^2(x_m) - \log f_j^2(x_m)] \\ &\qquad \leqslant (k+l)E[f^2(\tau) - f^2(\tau^{ij})][\log f^2(\tau) - \log f^2(\tau^{ij}); \sigma_j = x_m], \end{aligned}$$

we have

$$I_2 \leqslant \frac{\tilde{a}_{k,l}}{l(k+l)} \sum_{1\leqslant i\leqslant k<j\leqslant k+l} [f_i^2(x_m) - f_j^2(x_m)][\log f_i^2(x_m) - \log f_j^2(x_m)]$$

$$= \frac{2k}{k+l}\tilde{a}_{k,l}\mathscr{E}(f^2, \log f^2)$$

and the corresponding recurrence

$$a_{k,l} \leqslant a_{k,l-1} + \frac{2k}{k+l}\tilde{a}_{k,l}. \qquad \square$$

Corollary 3.3 (Top-random transposition). *For the top-random transposition walk on* S_n, *i.e. the complete* $(1, n-1)$*-bipartite walk, the modified log Sobolev constant satisfies*

$$\frac{2}{n-1} \geqslant \alpha \geqslant \frac{1}{2(n-1)}.$$

In particular, the mixing time satisfies $\tau \leqslant 2(n-1)[\log\log n! + 2]$.

Remark 3.3. Diaconis [10] outlines a proof that the top-random transposition walk exhibits cutoff at critical time $t_n = n\log n$.

Proof. By Lemma 3.1, $\tilde{a}_{1,l} \leqslant l+1$ and the recurrence reduces to $a_{1,l} \leqslant a_{1,l-1} + 2$. Since the top-random transposition walk on S_2 is the symmetric walk on the 2-point space, $a_{1,1} = \frac{1}{4}$, yielding the lower bound. The upper bound follows from Lemma 2.4 and the fact that $\lambda = 1/(n-1)$ [16]. $\square$

3.3. Random regular graphs

In the examples that we have examined thus, the modified log Sobolev constant α is of approximately the same magnitude as the spectral gap λ. This is not however always the case.

For fixed $r > 3$, [8] introduced a model for random r-regular graphs on n vertices. For this model, [2] shows that as n tends to infinity, a random r-regular graph $\mathscr{G}$ has spectral gap $\lambda(\mathscr{G}) \geqslant \varepsilon(r) > 0$ with probability $1 - o(1)$. Using this example, [12] shows that a generic r-regular graph has log Sobolev constant $\beta \ll \lambda$. The following lemma shows that for this family of random graphs we also have $\alpha \ll \lambda$.

Lemma 3.2. *Let* $G = (\mathscr{X}, E)$ *be a connected* r*-regular graph with* $|\mathscr{X}| \geqslant 8r$ *and let* (K, π) *be the canonical walk on* G *with uniform stationary distribution* π. *Then the modified log Sobolev constant satisfies*

$$\alpha \leqslant 4\log r \frac{2 + \log\log|\mathscr{X}|}{\log[|\mathscr{X}|/8]}.$$

In particular, for fixed r

$$\alpha \overset{|\mathscr{X}|\to\infty}{\to} 0.$$

Proof. For $x, y \in \mathscr{X}$, let $d(x, y)$ be the natural graph distance. Let $A_x^n = \{y \in \mathscr{X} \mid d(x, y) > n\}$. Then $K_x^n(A_x^n) = 0$, and

$$\pi(A_x^n) \geqslant 1 - \frac{1 + r + r^2 + \cdots + r^n}{|\mathscr{X}|}$$

$$\geqslant 1 - \frac{2r^n}{|\mathscr{X}|}$$

since by assumption we must have $r > 1$. Let $\mathrm{Po}(\lambda)$ denote a Poisson random variable with mean λ. Then,

$$H_t^x(A_x^n) \leqslant \mathrm{Prob}(\mathrm{Po}(t) \geqslant n) \leqslant \frac{t + t^2}{n^2}.$$

Furthermore for $n_0 = \log(|\mathscr{X}|/8)/\log r$ and $t_0 = n_0/4$, $\pi(A_x^{n_0}) \geqslant \frac{3}{4}$ and $H_{t_0}^x(A_x^{n_0}) \leqslant \frac{1}{4} + \frac{1}{16}$. Since

$$\|H_x^{t_0} - \pi\|_{TV} = \max_{A \subset \mathscr{X}} |H_x^{t_0}(A) - \pi(A)| \geqslant \frac{7}{16}$$

the mixing time satisfies $\tau > t_0$. But from Lemma 2.2

$$\tau \leqslant \frac{1}{\alpha}(\log\log|\mathscr{X}| + 2)$$

and the result follows. □

4. Comparison techniques

A perturbed chain often has log Sobolev and modified log Sobolev constants similar to the original. Lemma 4.1 illustrates this phenomenon; the proof is similar to the log Sobolev case presented in [12] and is omitted here.

Lemma 4.1. *Let* (K, π) *and* (K', π') *be two finite, reversible Markov chains defined on* $\mathscr{X}$ *and* $\mathscr{X}'$, *respectively, with modified log Sobolev constants* α *and* α'. *Assume there exists a map*

$$l^2(\mathscr{X}, \pi) \to l^2(\mathscr{X}', \pi') : f \to \tilde{f}$$

and constants $A, B, a > 0$ *such that for all* $f \in l^2(\mathscr{X}, \pi)$

$$\mathscr{E}'(\tilde{f}^2, \log \tilde{f}^2) \leqslant A\mathscr{E}(f^2, \log f^2) \text{ and } a\mathscr{L}_\pi(f) \leqslant \mathscr{L}_{\pi'}(\tilde{f}) + B\mathscr{E}(f^2, \log f^2),$$

then

$$\frac{a\alpha'}{A + B\alpha'} \leqslant \alpha.$$

In the case $\mathscr{X} = \mathscr{X}', \mathscr{E}'(f^2, \log f^2) \leqslant A\mathscr{E}(f^2, \log f^2)$ *and a* $\pi \leqslant \pi'$, *we have*

$$\frac{a\alpha'}{A} \leqslant \alpha.$$

Saloff-Coste [28] shows that for any two finite irreducible Markov chains K' and K on the same state space, there exists a constant A such that for all functions f, $\mathscr{E}'(f,f) \leqslant A\mathscr{E}(f,f)$. Consequently, the analog of Lemma 4.1 proven in [12] shows that we can always use the log Sobolev constant of one chain to estimate the constant for any other chain on the same space—although in practice this estimate may be quite bad.

There does not in general, however, exist a constant A such that for all f, $\mathscr{E}'(f^2, \log f^2) \leqslant A\mathscr{E}(f^2, \log f^2)$. Given the numerous similarities between the log Sobolev and modified log Sobolev constant, this fact is quite surprising. Consider the three-point space $\mathscr{X} = \{x_1, x_2, x_3\}$. Let K' be the complete graph on $\mathscr{X}$, and let K be the line graph with holding probability $\frac{1}{2}$ at the endpoints. Then both chains have uniform stationary distribution. Let

$$\Delta_{x_i x_j}(f) = [f^2(x_i) - f^2(x_j)][\log f^2(x_i) - \log f^2(x_j)].$$

Then

$$\frac{\mathscr{E}'(f^2, \log f^2)}{\mathscr{E}(f^2, \log f^2)} = 1 + \frac{\Delta_{x_1 x_3}}{\Delta_{x_1 x_2} + \Delta_{x_2 x_3}}.$$

Fix $A > 1$ and for $b > A$ let $f^2(x_1) = 1$, $f^2(x_2) = b/A$, and $f^2(x_3) = b$. Then

$$\begin{aligned}
\frac{\Delta_{x_1 x_3}}{\Delta_{x_1 x_2} + \Delta_{x_2 x_3}} &= \frac{(b-1)\log b}{(\frac{b}{a} - 1)\log \frac{b}{a} + \frac{A-1}{A}\, b \log A} \\
&\geqslant \frac{(b-1)\log b}{\frac{b}{a}\log b + b\log A} \\
&\overset{b\to\infty}{\to} A.
\end{aligned}$$

So for every A, there exists an f with $\mathscr{E}'(f^2, \log f^2) > A\mathscr{E}(f^2, \log f^2)$. While this shows that we cannot always compare chains, several interesting examples are amenable to comparison.

Example 4.1 (The Heavy Ace). Recall our informal description of the top-random transposition walk on the permutation group S_n: Uniformly at random pick a position i from 2 to n, and then swap the top card and the card in position i. More formally, this is the group walk on S_n with generating set $\{(1,i) | 2 \leqslant i \leqslant n\}$. Consider the following variant of this walk: Uniformly at random pick a position i from 2 to n; if either the top card or the card in position i is the 'Ace of Spades' do nothing with probability $\frac{1}{2}$ and swap the cards with probability $\frac{1}{2}$; otherwise, swap the cards as usual. Like the top-random transposition walk, this modified walk is reversible with respect to the uniform stationary distribution. However, unlike the former walk, the latter cannot be realized as a walk on a group. While intuitively this small perturbation should not dramatically affect mixing time, the comparison techniques of [11] to obtain precise results crucially rely on group structure.

The kernel of the Heavy Ace walk is given explicitly by

$$\tau_1 \neq 1 \quad K(\tau,\tau') = \begin{cases} \frac{1}{(n-1)}, & \tau' = \tau^{1i}, \quad 1 < i \leqslant n,\ \tau_i \neq 1, \\ \frac{1}{2(n-1)}, & \tau' = \tau^{1i}, \quad 1 < i \leqslant n,\ \tau_i = 1, \\ \frac{1}{2(n-1)}, & \tau' = \tau, \quad \tau_1 = 1, \\ 0, & \text{otherwise}, \end{cases}$$

$$\tau_1 = 1 \quad K(\tau,\tau') = \begin{cases} \frac{1}{2(n-1)}, & \tau' = \tau^{1i}, \quad 1 < i \leqslant n, \\ \frac{1}{2}, & \tau' = \tau, \quad \tau_1 = 1, \\ 0, & \text{otherwise}. \end{cases}$$

Letting $\mathscr{E}'$ be the Dirichlet form of the top-random transposition walk, we see that $\mathscr{E}'(f^2, \log f^2) \leqslant 2\mathscr{E}(f^2, \log f^2)$. By Lemma 4.1, $\alpha \geqslant \alpha'/2$. By Corollary 3.3, $\alpha \geqslant 1/4(n-1)$, and consequently by Corollary 2.2 the mixing time for the Heavy Ace walk satisfies

$$\tau \leqslant 4(n-1)[\log\log n! + 2].$$

Using the method detailed in [9], we can find a corresponding lower bound. For simplicity we will examine the discrete time chain K^n (the argument for H_t is analogous). Let A be the subset of permutations with at least one fixed point. That is $A = \{\sigma \in S_n \mid \sigma_i = i \text{ for some } 1 \leqslant i \leqslant n\}$. Under the uniform measure π, this is the matching problem, and arguments in [15] show that

$$\pi(A) = 1 - \frac{1}{\mathrm{e}} + \mathrm{O}\left(\frac{1}{n!}\right).$$

Let $\{(1,X_1),(1,X_2),\ldots,(1,X_k)\}$ denote the transpositions that we considered making up to step k. That is, at step i, $1 \leqslant i \leqslant k$, we choose cards 1 and X_i, checked if either was an 'Ace of Spades', and continued accordingly. Then to bound $K^k(A)$, observe that $K^k(A) \geqslant K^k(B)$ where $B = \{(\bigcup_{1\leqslant i\leqslant k} X_i) \neq \{2,\ldots,n\}\}$, i.e. B is the event that by step k we had not even chosen all of the positions. Arguments in [15] show that

$$K^k(B) = 1 - \mathrm{e}^{-n\mathrm{e}^{-k/n}} + \mathrm{o}(1) \quad \text{uniformly in } k \text{ as } n \to \infty.$$

For $k = n\log n + cn$, $K^k(A) \geqslant 1 - \mathrm{e}^{-\mathrm{e}^{-c}} + \mathrm{o}(1)$, and consequently

$$\begin{aligned} \|K^k - \pi\|_{\mathrm{TV}} &\geqslant |K^k(A) - \pi(A)| \\ &\geqslant \frac{1}{\mathrm{e}} - \mathrm{e}^{-\mathrm{e}^{-c}} + \mathrm{o}(1). \end{aligned}$$

5. Concentration of measure

In this section we present the well known connection between log Sobolev and concentration inequalities (see e.g. [7,21,22], and present some examples based on the inequalities derived in Section 3.

First we review the key definitions and results. Let (X, d, μ) denote a metric space (X, d) equipped with a probability measure μ on its Borel sets.

Definition 5.1. The concentration function on (X, d, μ) is given by

$$\alpha_{(X,d,\mu)}(r) = \sup\left\{1 - \mu(A_r) : A \subset X,\ \mu(A) \geqslant \frac{1}{2}\right\} \quad r > 0,$$

where $A_r = \{x \in X : d(x, A) < r\}$ is the open r-neighborhood of A with respect to the metric d.

In Theorem 5.1 we show that modified log Sobolev inequalities imply deviation inequalities for Lipschitz functions. Lemma 5.1 shows that these deviation bounds in turn yield concentration inequalities; for a proof of Lemma 5.1, see [22].

Definition 5.2. A real-valued function F on (X, d) is said to be Lipschitz if

$$\|F\|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|F(x) - F(y)|}{d(x, y)} < \infty.$$

We say that F is 1-Lipschitz if $\|F\|_{\mathrm{Lip}} \leqslant 1$.

Lemma 5.1. *Let μ be a Borel probability measure on a metric space (X, d). Assume that for some non-negative, decreasing function α on $\mathbb{R}_+$ and any bounded 1-Lipschitz function F on (X, d)*

$$\mu(\{F \geqslant EF + r\}) \leqslant \alpha(r)$$

for $r > 0$. Then

$$1 - \mu(A_r) \leqslant \alpha(\mu(A) r)$$

for every Borel set A with $\mu(A) > 0$ and every $r > 0$. In particular

$$\alpha_{(X,d,\mu)}(r) \leqslant \alpha\left(\frac{r}{2}\right).$$

For a reversible Markov chain (K, π) on state space $\mathscr{X}$, consider the graph $G = (\mathscr{X}, E)$ with symmetric edge set $E = \{(x, y) \mid \pi(x)K(x, y) > 0\}$. Then using the natural graph distance d, we can define the metric probability space $(\mathscr{X}, d, \pi)$. Theorem 5.1 follows the Herbst argument to relate the modified log Sobolev constant to a deviation inequality on this graph. For a discussion of this method and more examples, see [22].

Theorem 5.1. *Let α denote the modified log Sobolev constant for the reversible Markov chain (K,π) on $\mathscr{X}$. For any 1-Lipschitz function F on $(\mathscr{X},d)$, and $r>0$*

$$\mu(\{F \geqslant EF + r\}) \leqslant \mathrm{e}^{-\alpha r^2/2}.$$

Proof. Since by definition

$$\mathscr{L}(f) \leqslant \frac{1}{2\alpha} E\left[\sum_y [f^2(x) - f^2(y)]\cdot[\log f^2(x) - \log f^2(y)]K(x,y)\right], \tag{5.1}$$

letting $f^2 = \mathrm{e}^{\lambda F - \lambda^2/2\alpha}$ in (5.1), we get

$$E\left[\left(\lambda F - \frac{1}{2\alpha}\lambda^2\right) e^{\lambda F - (1/2\alpha)\lambda^2}\right] - \Lambda(\lambda)\log\Lambda(\lambda) \leqslant \frac{1}{2\alpha}\lambda^2\Lambda(\lambda),$$

where $\Lambda(\lambda) = Ef^2$. So,

$$\lambda\Lambda'(\lambda) \leqslant \Lambda(\lambda)\log\Lambda(\lambda).$$

Now let $H(\lambda) = \frac{\log\Lambda(\lambda)}{\lambda}$, with $H(0) = \frac{\Lambda'(0)}{\Lambda(0)} = EF$. Then

$$\begin{aligned} H'(\lambda) &= -\frac{\log\Lambda(\lambda)}{\lambda^2} + \frac{\Lambda'(\lambda)}{\lambda\Lambda(\lambda)} \\ &\leqslant 0. \end{aligned}$$

Consequently, $H(\lambda) \leqslant H(0)$. That is,

$$\Lambda(\lambda) \leqslant \mathrm{e}^{\lambda EF}$$

and so

$$E\mathrm{e}^{\lambda F} \leqslant \mathrm{e}^{\lambda EF + \lambda^2/2\alpha}.$$

Finally, for $r>0$

$$\begin{aligned} \mu(\{F \geqslant EF + r\}) &= \mu(\{\mathrm{e}^{\lambda F} \geqslant \mathrm{e}^{\lambda(EF+r)}\}) \\ &\leqslant \mathrm{e}^{-\lambda(EF+r)} E\mathrm{e}^{\lambda F} \\ &\leqslant \mathrm{e}^{\lambda^2/2\alpha - \lambda r}. \end{aligned}$$

Taking $\lambda = r\alpha$ yields the result. □

Using the modified log Sobolev inequalities derived in Section 3, we can obtain corresponding concentration inequalities. Here we consider two examples: random-transposition and the top-random transposition shuffle.

Consider the metric probability space on the symmetric group S_n given by the random transposition metric and the uniform probability distribution π. Since by Corollary 3.1 the modified log Sobolev constant for the associated walk satisfies $\alpha \geqslant 1/(n-1)$. Theorem 5.1 shows that for any 1-Lipschitz function on S_n, and $r>0$

$$\pi(\{F \geqslant EF + r\}) \leqslant \mathrm{e}^{-r^2/2(n-1)}.$$

Accordingly the concentration function satisfies

$$\alpha(r) \leqslant e^{-r^2/8(n-1)}.$$

This random transposition graph was also studied in [4] via the subgaussian constant, which implies a concentration inequality. In particular, the authors show that

$$\alpha(r) \leqslant e^{-(2r-\sqrt{n-1})^2/2(n-1)}.$$

As a second example, consider the graph associated with the top-random transposition shuffle and let $\tilde{d}$ be the associated metric on S_n. Then, since $(x, y) = (1, x)(1, y)(1, x)$, $\tilde{d}(\sigma_1, \sigma_2) \leqslant 3d(\sigma_1, \sigma_2)$. Consequently,

$$\alpha_{\tilde{d}}(r) \leqslant \alpha_d\left(\frac{r}{3}\right) \leqslant e^{-r^2/72(n-1)}.$$

However, since Corollary 3.3 shows that for this chain $\alpha \geqslant 1/2(n-1)$, for any 1-Lipschitz function F on $(S_n, \tilde{d})$ and $r > 0$

$$\pi(\{F \geqslant EF + r\}) \leqslant e^{-r^2/4(n-1)}.$$

Accordingly, we have the slightly tighter bound

$$\alpha_{\tilde{d}}(r) \leqslant e^{-r^2/16(n-1)}.$$

Acknowledgements

I thank my Ph.D. advisor Laurent Saloff-Coste for starting me on this project by suggesting the method of [23] to prove the results of Section 3, and for his continual guidance throughout this work.

References

[1] D. Aldous, J. Fill, Reversible Markov chains and random walks on graphs, Monograph in preparation available on the web at http://stat-www.berkeley.edu/users/aldous/RWG/book.html.

[2] C. Ané, S. Blachére, D. Chafaï, P. Fourgéres, I. Gentil, F. Malrieu, C. Roberto, G. Scheffer, Sur les inégalités de Sobolev logarithmiques, Société Mathématique de France, 2000.

[3] C. Ané, M. Ledoux, On logarithmic Sobolev inequalities for continuous time random walks on graphs, Probab. Theory Related Fields 116 (2000) 573–602.

[4] S.G. Bobkov, C. Houdré, P. Tetali, The subgaussian constant and concentration inequalities, Preprint available on the web at http://www.math.gatech.edu/~houdre/.

[5] S.G. Bobkov, M. Ledoux, On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures, J. Funct. Anal. 156 (1998) 347–365.

[6] S.G. Bobkov, M. Ledoux, From Brunn–Minkowski to Brascamp–Lieb and to logarithmic Sobolev inequalities, Geom. Funct. Anal. 10 (2000) 1028–1052.

[7] S.G. Bobkov, P. Tetali, Modified logarithmic Sobolov inequalities in discrete settings, Preprint available on the web at http://www.math.gatech.edu/~tetali/.

[8] B. Bollobas, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, European J. Combin. 1 (1980) 311–316.

[9] P. Diaconis, Group representations in probability and statistics, Institute of Mathematical Statistics, 1988.

[10] P. Diaconis, Finite fourier methods: access to tools, in: B. Bollobás (Ed.), Probabilistic Combinatorics and Its Application, Proceedings of Symposia in Applied Mathematics, Vol. 44, American Mathematical Society, Providence, RI, 1992.

[11] P. Diaconis, L. Saloff-Coste, Comparison techniques for random walk on finite groups, Ann. Probab. 21 (1993) 2131–2156.
[12] P. Diaconis, L. Saloff-Coste, Logarithmic Sobolev inequalities for finite Markov chains, Ann. Appl. Probab. 6 (1996) 695–750.
[13] P. Diaconis, M. Shahshahani, Generating a random permutation with random transposition, Z. Wahr. Verw. Geb. 57 (1981) 159–179.
[14] P. Diaconis, M. Shahshahani, Time to reach stationarity in the Bernoulli–Laplace diffusion model, SIAM J. Math. Anal. 18 (1987) 208–218.
[15] W. Feller, An Introduction to Probability Theory and Its Applications, Vol. I, Wiley, New York, 1968.
[16] L. Flatto, A.M. Odlyzko, D.B. Wales, Random shuffles and group representations, Ann. Probab. 13 (1985) 154–178.
[17] F. Gao, J. Quastel, Exponential decay of entropy in the random transposition and Bernoulli–Laplace models, Ann. Appl. Probab. 4 (2003) 1591–1600.
[18] L. Gross, Logarithmic Sobolev inequalities, Amer. J. Math. 97 (1975) 1061–1083.
[19] L. Gross, Logarithmic Sobolev Inequalities and Contractivity Properties of Semigroups, Lecture Notes in Mathematics, Vol. 1563, Springer, Berlin, 1993.
[20] Y. Higuchi, N. yoshida, Analytic conditions and phase transition for ising models, Unpublished lecture notes (in Japanese).
[21] M. Ledoux, Concentration of measure and logarithmic Sobolev inequalities, Séminaire de Probabilités XXXIII, Lecture Notes in Mathematics, Vol. 1709, Springer, Berlin, 1999.
[22] M. Ledoux, The concentration of measure phenomenon, American Mathematical Society, Providence, RI, 2001.
[23] T.-Y. Lee, H.-T. Yau, Logarithmic Sobolev inequalities for some models of random walks, Ann. Probab. 26 (1998) 1855–1873.
[24] C. Roberto, Inégalités de hardy et de sobolev logarithmiques, Ph.D. Thesis, Université Paul Sabatier, Toulouse III, June 2001.
[25] S. Roussel, Marches aléatories sur le groupe symétrique, Ph.D. Thesis, Université Paul Sabatier, Toulouse III, June 1999.
[26] S. Roussel, Phénoméne de cutoff pour certaines marches aléatoires sur le groupe symétrique, Colloq. Math. 86 (2000) 111–135.
[27] L. Saloff-Coste, Precise estimates on the rate at which certain diffusions tend to equlibrium, Math. Z. 217 (1994) 641–677.
[28] L. Saloff-Coste, Lectures on finite Markov chains, in: P. Bernard (Ed.), Lectures on Probability theory and Statistics, Lecture Notes in Mathematics, Vol. 1665, Ecole d'Eté de Probabiltés de Saint-Flour XXVI, Springer, Berlin, 1996.
[29] L. Saloff-Coste, Random walks on finite groups, in: H. Kesten (Ed.), Probability on Discrete Structures, Encylopaedia of Mathematical Sciences, Vol. 110, Springer, Berlin, 2004.
[30] L. Wu, A new modified logarithmic Sobolev inequality for Poisson point processes and several applications, Probab. Theory Related Fields 118 (2000) 427–438.