UC Davis

Title

Upper bounds on mixing time of finite Markov chains

Permalink

https://escholarship.org/uc/item/3t17145f

Authors

Rhodes, John
Schilling, Anne

Publication Date

2020-10-17

Copyright Information

Peer reviewed

UPPER BOUNDS ON MIXING TIME OF FINITE MARKOV CHAINS

JOHN RHODES AND ANNE SCHILLING

Dedicated to Ron Graham and Vaughan Jones

ABSTRACT. We provide a general framework for computing upper bounds on mixing times of finite Markov chains when its minimal ideal is left zero. Our analysis is based on combining results by Brown and Diaconis with our previous work on stationary distributions of finite Markov chains. Stationary distributions can be computed from the Karnofsky–Rhodes and McCammond expansion of the right Cayley graph of the finite semigroup underlying the Markov chain. Using loop graphs, which are planar graphs consisting of a straight line with attached loops, there are rational expressions for the stationary distribution in the probabilities. From these we obtain bounds on the mixing time. In addition, we provide a new Markov chain on linear extension of a poset with n vertices, inspired by but different from the promotion Markov chain of Ayyer, Klee and the last author. The mixing time of this Markov chain is $O(n \log n)$.

1. INTRODUCTION

A *Markov chain* is a model that describes transitions between states in a state space according to certain probabilistic rules. The defining characteristic of a Markov chain is that the transition from one state to another only depends on the current state and the elapsed time, but not how the system arrived there. In other words, a Markov chain is "memoryless". Markov chains have an abundance of applications, from data analysis, population dynamics to traffic models.

For a Markov chain, the *stationary distribution* Ψ is the long-term limiting distribution. Mathematically speaking, it is the eigenvector of the transition matrix T of the Markov chain with eigenvalue one. That is

$$T\Psi = \Psi.$$

An important question is how quickly does the Markov chain converge to the stationary distribution. In Markov chain theory, distance is usually the total variation distance or half the L^1-norm in classical analysis. If Ω is the state space, the total variation distance between two probability distributions ν and μ is defined as

$$\|\nu - \mu\| = \max_{A \subseteq \Omega} |\nu(A) - \mu(A)|.$$

For a given small $\epsilon > 0$, the *mixing time* t_{mix} is the smallest t such that

$$\|T^t \nu - \Psi\| \leqslant \epsilon,$$

independent of the initial distribution ν.

In seminal work of Bidigare, Hanlon and Rockmore [8], which was continued by Diaconis, Brown, Athanasiadis, Björner, Chung and Graham, amongst others [15, 9, 14, 10, 11, 3, 18, 44], the special family of semigroups, now known as *left regular bands* first studied by Schützenberger [45] in the forties, was applied to random walks or Markov chains on hyperplane arrangements. In his 1998 ICM lecture [20], Diaconis discussed these developments. In Section 4.1, entitled *What is the ultimate generalization?*, he asks how far the semigroup techniques can be taken.
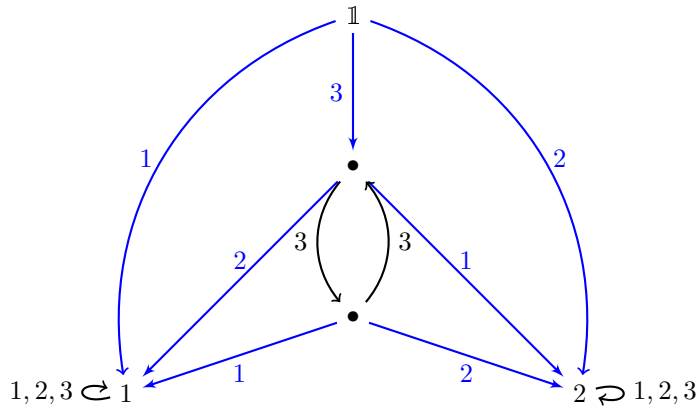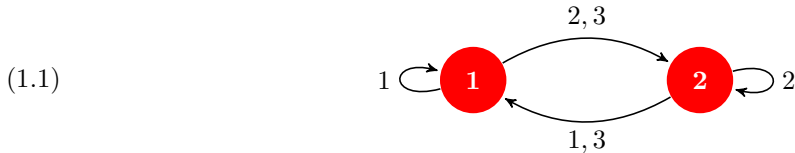
FIGURE 1.   The right Cayley graph $\mathsf{RCay}(S, A)$ of the semigroup that gives the Markov chain in (1.1) with generators $A = \{1, 2, 3\}$.

Every finite state Markov chain $\mathcal{M}$ has a random letter representation, that is, a representation of a semigroup S acting on the left on the state space Ω. See for example [31, Proposition 1.5] and [6, Theorem 2.3]. In this setting, there is a transition $s \xrightarrow{a} s'$ with probability $0 \leqslant x_a \leqslant 1$, where $s, s' \in \Omega$, $a \in S$ and $s' = a.s$ is the action of a on the state s. It is enough to consider the semigroup S generated by the elements a with $x_a > 0$, called the generating set A. For example, the Markov chain with state space $\Omega = \{1, 2\}$ and transition diagram

(1.1)

can be associated to the semigroup with right Cayley graph depicted in Figure 1. The conceptual reason why a Markov chain described using the *left* action of a semigroup can be analyzed using the *right* Cayley graph is that if time goes left (due to the left action), then coupling from the past corresponds to right multiplication. The transition matrix in this Markov chain is

$$T = \begin{pmatrix} x_1 & x_1 + x_3 \\ x_2 + x_3 & x_2 \end{pmatrix}.$$

In the pursuit of finding Diaconis' ultimate generalization [20], the arguments in Brown and Diaconis [15] were generalized to Markov chains for $\mathscr{R}$-trivial semigroups [6]. In [39, 38], the current authors developed a general theory for computing the stationary distribution for any finite Markov chain. The theory uses semigroup methods such as the Karnofsky–Rhodes and McCammond expansion of a semigroup. These expansions give rise to loop graphs which immediately yield Kleene expressions for all paths from the root of the graph to elements in the minimal ideal of the semigroup. The Kleene expressions in turn give rational expressions for the stationary distribution.

In this paper we apply the findings of [39, 38] to study upper bounds on the mixing time of the Markov chain. In particular, Theorems 2.4 and 2.5 provide upper bounds for the mixing time directly from the rational expression of the stationary distribution in the case when the minimal ideal of the semigroup is left zero. This general theory is applied to specific examples (Tsetlin library, edge flipping on a line Markov chain, and a new Markov chain on linear extensions) in Section 3.

The paper is organized as follows. In Section 2, we develop the main theory. In Section 2.6, we present our main theorems regarding the upper bounds on the mixing time (see Theorems 2.4

and 2.5). We discuss the relation to the Shannon entropy in Section 2.7. In Section 2.8, we refine bounds on the mixing time using certain statistics that were developed in [6, 5]. In Section 2.9, we consider semigroups syntactic at zero. In particular, we prove in Theorem 2.14 that the upper bounds on the mixing time do not change by replacing the semigroup by its syntactic image. In Section 2.10, we relate observations on mixing time to d-testable languages. Finally, in Section 3 we consider specific examples such as the Tsetlin library [17], edge flipping on a line [15, 18], and a new Markov chain on linear extensions of a poset with n vertices, which is inspired by but different from the promotion Markov chain [4]. This new Markov chain has a mixing time of $O(n \log n)$ as compared to the mixing time of the model of Bubley and Dyer [16] with mixing time $O(n^3 \log n)$.

2. Mixing time

Let T be the *transition matrix* of a finite Markov chain. Assuming that the Markov chain is *ergodic* (meaning that it is irreducible and aperiodic), by the Perron–Frobenius Theorem there exists a unique *stationary distribution* Ψ and $T^t \nu$ converges to Ψ as $t \to \infty$ for any initial state ν. A Markov chain is irreducible if the graph of the Markov chain is strongly connected. It is aperiodic if the gcd of the cycle lengths in the graph of the Markov chain is one. In fact, the stationary distribution is the right eigenvector of eigenvalue one of T

$$T\Psi = \Psi.$$

The *mixing time* measures how quickly the Markov chain converges to the stationary distribution. For a given small $\epsilon > 0$, t_{mix} is the smallest t such that

$$\|T^t \nu - \Psi\| \leqslant \epsilon.$$

We begin this section by reviewing methods to compute upper bounds on mixing times in Section 2.1 (see in particular Theorem 2.1), relations between ideals and semaphore codes and how this relates to mixing time in Section 2.2, and the Markov and Chernoff inequalities to bound mixing time in Sections 2.3 and 2.4. The semigroup methods of [39, 38] to compute rational expressions of the stationary distribution of a Markov chain in terms of the probabilities x_a for the generators $a \in A$ of the semigroup are reviewed in Section 2.5. Our main new results for the upper bounds of the mixing times in terms of truncations of the rational expressions of the stationary distribution (Theorem 2.4) and using a Cauchy–Euler operator (Theorem 2.5) are stated in Section 2.6. In Section 2.7 we discuss the relation between Shannon entropy and mixing time. Sections 2.8-2.10 are devoted to new results in special settings, for example for monoids which are syntactic at zero (Theorem 2.14) and d-testable languages (Remark 2.21).

2.1. **Upper bound.** Brown and Diaconis [15] [14, Theorem 0] showed, for Markov chains associated to left regular bands, that the total variational distance from stationarity after t steps is bounded above by the probability $\mathsf{Pr}(\tau > t)$, where τ is the first time that the walk hits a certain ideal. The arguments in Brown and Diaconis [15] can be generalized to arbitrary finite Markov chains (not just those related to left regular bands). To state the details, we need some more notation.

Let $\mathcal{M}(S, A)$ be a finite state Markov chain with state space Ω and transition matrix T associated to the semigroup S with generators A with probabilities $0 < x_a \leqslant 1$ for $a \in A$.

A two-sided *ideal* I (or ideal for short) is a subset $I \subseteq S$ such that $uIv \subseteq I$ for all $u, v \in S^{\mathbb{1}}$, where $S^{\mathbb{1}}$ is the semigroup S with identity $\mathbb{1}$ added (even if S already contains an identity). If I, J

are ideals of S, then $IJ \subseteq I \cap J$, so that $I \cap J \neq \emptyset$. Hence every finite semigroup has a unique nonempty minimal ideal denoted $K(S)$.

Assume that the minimal ideal $K(S)$ is *left zero*, that is, $xy = x$ for all $x, y \in K(S)$. This assumption implies that the Markov chain on the minimal ideal (given by the left action) is ergodic. Let τ be the random variable which is the time that the random walk is absorbed into the minimal ideal $K(S)$.

Theorem 2.1. [6] *Let S be a finite semigroup whose minimal ideal $K(S)$ is a left zero semigroup and let T be the transition matrix of the associated Markov chain. Then*

$$\|T^t \nu - \Psi\| \leqslant \Pr(\tau > t).$$

Proof. By [6, Corollary 3.5(3)], we have

$$\|T^t \nu - \Psi\| \leqslant P^{\star t}(S \setminus K(S)),$$

where $P^{\star n}$ denotes the n-th convolution power of P. By [6, Eq. (4.6)], the right hand side equals $\Pr(\tau > t)$. □

2.2. **Ideals and semaphore codes.** Let A be a finite alphabet, A^+ the set of all nonempty words in the alphabet A, and $A^\star$ the set of all words in the alphabet A.

As shown in [41], ideals in A^+ are in bijection with semaphore codes [7]. A *prefix code* is a subset of A^+ such that all elements are incomparable in prefix order (meaning that no element is the prefix of any other element of the code). A *semaphore code* $\mathcal{S}$ is a prefix code such that $A\mathcal{S} \subseteq \mathcal{S}A^\star$. There is a natural left action on a semaphore code. If $u \in \mathcal{S} \subseteq A^+$ and $a \in A$, then au has a prefix in $\mathcal{S}$ (and hence a unique prefix of au). The left action $a.u$ is the prefix of au that is in $\mathcal{S}$. Assigning probability $0 \leqslant x_a \leqslant 1$ to $a \in A$, the left action on a semaphore code $\mathcal{S}$ defines a Markov chain with a countable state space $\mathcal{S}$.

The bijection between ideals $I \subseteq A^+$ and semaphore codes $\mathcal{S}$ over A is given as follows (see [41, Proposition 4.3]). If $u = a_1 a_2 \ldots a_j \in I \subseteq A^+$, find the (necessarily unique) index $1 \leqslant i \leqslant j$ such that $a_1 \ldots a_{i-1} \notin I$, but $a_1 \ldots a_i \in I$. Then $a_1 \ldots a_i$ is a code word and the set of all such words forms the semaphore code $\mathcal{S}$. Conversely, given a semaphore code $\mathcal{S}$, the corresponding ideal is $\mathcal{S}A^\star$.

In this setting, τ can be interpreted as the random variable given by the length of the semaphore code words. Let $\mathcal{S}$ be a semaphore code and I the ideal under the bijection described above. A semaphore code word $s = s_1 s_2 \ldots s_\ell$ has the property that $s \in I$, but $s_1 s_2 \ldots s_{\ell-1} \notin I$. Hence τ can be interpreted as the random variable given by the length ℓ.

Next we discuss two ways to approximate $\Pr(\tau > t)$ using Markov's and Chernoff's inequality.

2.3. **Markov's inequality.** By Markov's inequality (see for example [31, 19]), we have

$$(2.1) \qquad \Pr(\tau > t) \leqslant \frac{E[\tau]}{t+1},$$

where $E[\tau]$ is the expected value for τ, the first time the walk hits the ideal. We have

$$(2.2) \qquad E[\tau] = \sum_{a=1}^{\infty} \Pr(\tau \geqslant a).$$

2.4. **Chernoff's inequality.** Chernoff's inequality uses the moment generating function combined with Markov's inequality (2.1) to give an upper bound on $\Pr(\tau \geqslant t)$. More precisely,

$$\Pr(\tau \geqslant t) = \Pr(e^{s\tau} \geqslant e^{st}) \qquad \text{for } s > 0.$$

Hence by Markov's inequality (2.1)

$$\Pr(\tau \geqslant t) \leqslant \frac{E[e^{s\tau}]}{e^{st}}$$

and since this is true for all $s > 0$

$$\Pr(\tau \geqslant t) \leqslant \min_{s>0} \left\{ \frac{E[e^{s\tau}]}{e^{st}} \right\}.$$

2.5. **Rational expressions for stationary distributions.** Let $\mathcal{M}(S, A)$ be the Markov chain associated to the finite semigroup S with generators in A. Assume that its minimal ideal $K(S)$ is left zero, so that $K(S)$ can be taken as the state space Ω of the Markov chain. Denote by $\mathcal{S}(S, A)$ the semaphore code associated to $K(S)$ (see Section 2.2). For a word $s \in A^+$, we denote by $[s]_S$ the image of the word in the alphabet A in S. The following theorem is stated in [39, Corollaries 2.23 & 2.28].

Theorem 2.2. [39] *If $K(S)$ is left zero, the stationary distribution of the Markov chain $\mathcal{M}(S, A)$ labeled by $w \in K(S)$ is given by*

$$(2.3) \qquad \Psi_w(x_1, \ldots, x_n) = \sum_{\substack{s \in \mathcal{S}(S, A) \\ [s]_S = w}} \prod_{a \in s} x_a.$$

In [39, 38], we developed a strategy using loop graphs to compute the expressions in Theorem 2.2 as rational functions in the probabilities x_a for $a \in A$. This is done in several steps:

(1) We used the McCammond and Karnofsky–Rhodes expansion $\mathsf{Mc} \circ \mathsf{KR}(S, A)$ of the right Cayley graph $\mathsf{RCay}(S, A)$ of the semigroup S with generators A. In this paper we do not require the details of these definitions, except that the right Cayley graph as well as its expansions are rooted graphs with root $\mathbb{1}$. The Karnofsky–Rhodes expansion is another right Cayley graph, whereas the McCammond expansion is only an automata. For the precise definition of the Karnofsky–Rhodes expansion, we refer the reader to [34, Definition 4.15], [33, Section 3.4], [38, Section 2.4], and [42, Section 2]. For the definition of the McCammond expansion, we refer the reader to [34, Section 2.7] and [38, Section 2.5]. The Markov chain $\mathcal{M}(S, A)$ is a *lumping* [31] of the Markov chains associated to the expansions.

(2) The stationary distributions of the Markov chains associated to the expansions can be expressed using *loop graphs* G, see [38]. A loop graph is a straight line path from $\mathbb{1}$ to an endpoint s with directed loops of any finite length attached recursively to any vertex (besides $\mathbb{1}$ and s). In this way [38, Theorem 1.4]

$$(2.4) \qquad \Psi_w(x_1, \ldots, x_n) = \sum_G \Psi_G(x_1, \ldots, x_n),$$

where the sum is over certain loop graphs G with end point s such that $[s]_S = w$. Here [38, Definition 1.3]

$$(2.5) \qquad \Psi_G(x_1, \ldots, x_n) = \sum_p \prod_{a \in p} x_a,$$

where the sum is over all paths p in G starting at $\mathbb{1}$ and ending in s.

(3) There is a *Kleene expression* for the set of all paths from $\mathbb{1}$ to s in G. The Kleene expression immediately yields a rational expression for the stationary distribution $\Psi_G(x_1, \ldots, x_n)$ and hence $\Psi_w(x_1, \ldots, x_n)$ by (2.4).

Remark 2.3. An important property of the above construction is that in the series expansion of the rational expression for $\Psi_w(x_1, \ldots, x_n)$ (resp. $\Psi_G(x_1, \ldots, x_n)$) the total degree of each term corresponds to the length of the underlying semaphore code word in (2.3) (resp. the underlying path in G in (2.5)).

2.6. **Mixing time via truncation of Kleene expressions.** As stated in Theorem 2.1, $\Pr(\tau \geqslant t)$ provides an upper bound on the mixing time in the setting that $K(S)$ is left zero. As discussed in Section 2.2, τ can be interpreted as the random variable given by the length of the semaphore code words or paths in the loop graph. To compute $\Pr(\tau \geqslant t)$, one needs to compute the sum of probabilities of all paths of length weakly greater than t. By Remark 2.3, the length of the paths is given by the total degree in the probability variables $x_1, \ldots, x_n$ for the generators $a_1, \ldots, a_n$ of the semigroup S. Hence we obtain $\Pr(\tau \geqslant t)$ by truncating the rational function for the stationary distribution to total degree weakly bigger than t.

Let $\Psi_w^{\geq t}(x_1, \ldots, x_n)$ be the truncation of the formal power series associated to the rational function $\Psi_w(x_1, \ldots, x_n)$ to terms of degree weakly bigger than t and let $\Psi_w^{<t}(x_1, \ldots, x_n)$ be the truncation of the formal power series associated to the rational function $\Psi_w(x_1, \ldots, x_n)$ to terms of degree strictly smaller than t. Note that

$$\Psi_w(x_1, \ldots, x_n) = \Psi_w^{<t}(x_1, \ldots, x_n) + \Psi_w^{\geq t}(x_1, \ldots, x_n).$$

Theorem 2.4. *Suppose the Markov chain satisfies the conditions of Theorem 2.1. If $\Psi_w(x_1, \ldots, x_n)$ is represented by a rational function such that each term of degree ℓ in its formal power sum expansion corresponds to a semaphore code word s of length ℓ with $[s]_S = w$, we have*

$$\mathsf{Pr}_w(\tau \geq t) = \frac{\Psi_w^{\geq t}(x_1, \ldots, x_n)}{\Psi_w(x_1, \ldots, x_n)} = 1 - \frac{\Psi_w^{<t}(x_1, \ldots, x_n)}{\Psi_w(x_1, \ldots, x_n)}.$$

For each $w \in K(S)$, we can also give an explicit formula for the expected number of steps $E_w[\tau]$ it takes to reach the endpoint of w using the Cauchy–Euler operator.

Theorem 2.5. *Suppose the Markov chain satisfies the conditions of Theorem 2.1. If $\Psi_w(x_1, \ldots, x_n)$ is represented by a rational function such that each term of degree ℓ in its formal power sum expansion corresponds to a semaphore code word s of length ℓ with $[s]_S = w$, we have*

$$E_w[\tau] = \left(\sum_{i=1}^{n} x_i \frac{\partial}{\partial x_i} \right) \ln \Psi_w(x_1, \ldots, x_n).$$

Remark 2.6. Note that the formal expression for $\Psi_w(x_1, \ldots, x_n)$ cannot be manipulated using that $x_1 + \cdots + x_n = 1$ when using Theorems 2.4 and 2.5.

Proof of Theorem 2.5. Let the formal power sum expression for the rational function $\Psi_w(x_1, \ldots, x_n)$ be as follows

$$\Psi_w(x_1, \ldots, x_n) = \sum_{m_1, \ldots, m_n \geq 0} c_{m_1, \ldots, m_n} x_1^{m_1} \cdots x_n^{m_n}.$$

Then formally

$$\left(\sum_{i=1}^{n} x_i \frac{\partial}{\partial x_i} \right) \ln \Psi_w(x_1, \ldots, x_n) = \frac{\left(\sum_{i=1}^{n} x_i \frac{\partial}{\partial x_i} \right) \Psi_w(x_1, \ldots, x_n)}{\Psi_w(x_1, \ldots, x_n)}$$

$$= \frac{\sum_{m_1, \ldots, m_n \geq 0} c_{m_1, \ldots, m_n} (m_1 + \cdots + m_n) x_1^{m_1} \cdots x_n^{m_n}}{\sum_{m_1, \ldots, m_n \geq 0} c_{m_1, \ldots, m_n} x_1^{m_1} \cdots x_n^{m_n}}.$$

Note that a term $x_1^{m_1} \cdots x_n^{m_n}$ of degree $m_1 + \cdots + m_n$ corresponds to a semaphore code word of length $m_1 + \cdots + m_n$. Hence $c_{m_1, \ldots, m_n} (m_1 + \cdots + m_n) x_1^{m_1} \cdots x_n^{m_n} / \Psi_w(x_1, \ldots, x_n)$ is the length of the path times the probability of having taken a path with m_i steps along the i-th generator. The sum over all such terms is precisely $E_w[\tau]$. $\square$
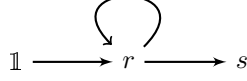
Remark 2.7. Let $\mathsf{Pr}_G(\tau \geq t)$ be the probability that the length of the paths in the loop graph G from $\mathbb{1}$ to the end point s is weakly bigger than t. Then by analogous argument as above, we also have

$$(2.6) \qquad \mathsf{Pr}_G(\tau \geq t) = \frac{\Psi_G^{\geq t}(x_1, \ldots, x_n)}{\Psi_G(x_1, \ldots, x_n)} = 1 - \frac{\Psi_G^{<t}(x_1, \ldots, x_n)}{\Psi_G(x_1, \ldots, x_n)}$$

and

$$(2.7) \qquad E_G[\tau] = \left(\sum_{i=1}^{n} x_i \frac{\partial}{\partial x_i} \right) \ln \Psi_G(x_1, \ldots, x_n).$$

Example 2.8 (Single loop). Suppose the path in the loop graph G from $\mathbb{1}$ to the ideal is a straight line with a single loop

where the loop is taken with probability p and the step to the ideal $r \to s$ with probability $1 - p$. Then the probability that one starts at $\mathbb{1}$ and hits the element s in the ideal in precisely t steps is

$$\mathsf{Pr}_G(\tau = t) = (1 - p)p^{t-2} \qquad \text{for } t \geqslant 2.$$

Hence

$$(2.8) \qquad \mathsf{Pr}_G(\tau \geqslant t) = \sum_{j=t}^{\infty} \mathsf{Pr}_G(\tau = j) = (1-p)p^{t-2} \sum_{j=0}^{\infty} p^j = (1-p)p^{t-2}\frac{1}{1-p} = p^{t-2} \quad \text{for } t \geqslant 2.$$

The expectation value is

$$(2.9) \qquad E_G[\tau] = \sum_{t=1}^{\infty} \mathsf{Pr}_G(\tau \geqslant t) = 1 + \sum_{t=2}^{\infty} p^{t-2} = 1 + \frac{1}{1-p}.$$

Indeed by Markov's inequality

$$t p^{t-2} \leqslant 1 + \frac{1}{1-p} \quad \text{for all } t \geqslant 2.$$

Now let us use (2.6) to compute $\mathsf{Pr}_G(\tau \geqslant t)$. Suppose that the step $\mathbb{1} \to r$ is labelled by the generator 1, the loop from r to r is labelled 2, and the step $r \to s$ is labelled 3. Then the Kleene expression for the paths from $\mathbb{1}$ to s is

$$12^\star 3.$$

Let the probability for generator i be x_i for $i \in \{1, 2, 3\}$. Then by [39]

$$\Psi_G(x_1, x_2, x_3) = \frac{x_1 x_3}{1 - x_2} = x_1 x_3 \sum_{j=0}^{\infty} x_2^j.$$

By (2.6), we obtain $\mathsf{Pr}_G(\tau \geqslant t) = 1$ for $t = 0, 1$ and

$$\mathsf{Pr}_G(\tau \geqslant t) = \frac{x_1 x_3 \sum_{j=t-2}^{\infty} x_2^j}{x_1 x_3 \sum_{j=0}^{\infty} x_2^j} = x_2^{t-2} \qquad \text{for } t \geqslant 2.$$

This agrees with (2.8), where $x_2 = p$.

Next let us use (2.7) to compute $E_G[\tau]$

$$E_G[\tau] = \left(x_1 \frac{\partial}{\partial x_1} + x_2 \frac{\partial}{\partial x_2} + x_3 \frac{\partial}{\partial x_3} \right) \ln \Psi_G(x_1, x_2, x_3) = 2 + \frac{x_2}{1 - x_2} = 1 + \frac{1}{1 - x_2},$$

which agrees with (2.9) when $x_2 = p$.

2.7. **Shannon entropy and exponential bounds.** It turns out that the mixing time has close ties to information theory and in particular Shannon's entropy. See [43] and [24, Chapter 3] as references on information theory.

Let X be a random variable with probability distribution $p(x)$. The amount of information of an elementary event x is $\log \frac{1}{p(x)}$. Therefore, the average amount of information about X is given by the expected value, known as *Shannon's entropy*

$$(2.10) \qquad H(X) = E[\log \frac{1}{p}] = \sum_{x \in X} p(x) \log \frac{1}{p(x)}.$$

Shannon's entropy features in the *asymptotic equipartition property* or *entropy ergodic theorem*, which can be stated as follows [46] (see also [43]). Let $\mathbf{x} = (x_1, \ldots, x_t)$ be a long sequence of independent and identically distributed outcomes with probability distribution $p(x)$. By the independence,

$p(\mathbf{x})$ is given by the product

$$p(\mathbf{x}) = p(x_1)p(x_2)\cdots p(x_t) = \prod_{x \in X} p(x)^{t(x)},$$

where $t(x)$ is the number of x_i equal to x. Since t is large, by the law of large numbers

$$\frac{t(x)}{t} \approx p(x),$$

which implies

(2.11) $$p(\mathbf{x}) \approx \Big(\prod_{x \in X} p(x)^{p(x)} \Big)^t = e^{-tH(X)}.$$

In other words, for very large (but fixed) t, the value of the probability of a given "typical" sequence $\mathbf{x} = (x_1, x_2, \ldots, x_t)$ is likely to be close to the constant $e^{-tH(X)}$.

The precise formulation of the asymptotic equipartition property is the *Shannon–McMillan–Breiman Theorem* [46, 35, 12] (see also [24, Chapter 4]). Applied to $P^{\star t}(S \setminus K(S))$ in Theorem 2.1, this gives an exponential bound on $\|T^t \nu - \Psi\|$. In probability, this is also known as the *Convergence Theorem* (see [31, Theorem 4.9]).

Theorem 2.9 (Convergence Theorem). *Suppose T is the transition matrix of an ergodic Markov chain with stationary distribution Ψ. Then there exist constants $\alpha \in (0,1)$ and $C > 0$ such that*

$$\|T^t \nu - \Psi\| \leqslant C\alpha^t.$$

A concept related to entropy is the *entropy rate*. It is defined as the rate of information innovation

$$H' = \lim_{t \to \infty} H(X_t \mid X_{t-1}, \ldots, X_1).$$

When X_i is stationary, the entropy rate is equal to the average entropy per symbol

$$\overline{H} = \lim_{t \to \infty} \frac{H(X_1, \ldots, X_t)}{t},$$

that is $H' = \overline{H}$.

Since an ergodic Markov chain has a unique stationary distribution Ψ, the entropy rate is independent of the initial distribution. If the Markov chain is defined on the finite (or countable) state space Ω, then

$$H' = - \sum_{s,s' \in \Omega} T_{s,s'} \Psi_{s'} \log(T_{s,s'}).$$

A simple consequence of this definition is that indeed a stochastic process with independent and identically distributed random variables has an entropy rate that is the same as the entropy of any individual member of the process.

2.8. **Mixing time via decreasing statistics.** In [6, 5], a technique was developed for an upper bound on the mixing time using a decreasing statistics on the semigroup underlying the Markov chain.

Lemma 2.10. [6, Lemma 3.6] *Let $\mathcal{M}$ be an irreducible Markov chain associated to the semigroup S and probability distribution $0 \leqslant p(s) \leqslant 1$ for $s \in S$. We assume that $\{s \in S \mid p(s) > 0\}$ generates S. Let Ψ be the stationary distribution and $f \colon S \to \mathbb{N}$ be a function, called a* statistic, *such that:*

    (1) $f(ss') \leqslant f(s)$ for all $s, s' \in S$;
    (2) if $f(s) > 0$, then there exists $s' \in S$ with $p(s') > 0$ such that $f(ss') < f(s)$;
    (3) $f(s) = 0$ if and only if $s \in K(S)$.

Then if $p = \min\{p(s) \mid s \in S, p(s) > 0\}$ and $L = f(\mathbb{1})$, we have that

$$\|T^t \nu - \Psi\|_{TV} \leqslant \sum_{i=0}^{L-1} \binom{t}{i} p^i (1-p)^{t-i} \leqslant \exp\left(-\frac{(tp - (L-1))^2}{2tp}\right),$$

for any probability distribution ν on S, where the last inequality holds as long as $t \geqslant (L-1)/p$.

The bound

$$\sum_{i=0}^{L-1} \binom{t}{i} p^i (1-p)^{t-i} \leqslant \exp\left(-\frac{(tp - (L-1))^2}{2tp}\right)$$

works well for p close to $\frac{1}{2}$. A better bound for $0 < \frac{L-1}{t} < p$ is given by [2]

$$\sum_{i=0}^{L-1} \binom{t}{i} p^i (1-p)^{t-i} \leqslant \exp\left(-t\, D\!\left(\frac{L-1}{t} \,\middle\|\, p\right)\right),$$

where

$$D(a \parallel p) = a \log \frac{a}{p} + (1-a) \log \frac{1-a}{1-p}.$$

This can be rewritten as

$$\sum_{i=0}^{L-1} \binom{t}{i} p^i (1-p)^{t-i} \leqslant \left(\frac{p}{a}\right)^{ta} \left(\frac{1-p}{1-a}\right)^{t(1-a)},$$

where $a = \frac{L-1}{t}$.

2.9. Syntactic at 0.

Syntactic monoids were introduced in mathematics and computer science as the smallest monoid that recognizes a given formal language, see for example [48]. Here we develop this idea in the context of the mixing time.

Recall that for a semigroup S, denote by $S^{\mathbb{1}}$ the semigroup S with a new added identity $\mathbb{1}$ (even if a one already exists).

Definition 2.11. Let S be a semigroup with zero 0. Define the congruence on $s_1, s_2 \in S$ by

(2.12) $\qquad s_1 \equiv s_2 \qquad$ if and only if $\qquad \left(\text{for any } x, y \in S^{\mathbb{1}} \quad x s_1 y = 0 \Longleftrightarrow x s_2 y = 0\right).$

Then S is called *syntactic at zero* if the congruence (2.12) has singleton classes, that is,

$$S/\equiv \quad \cong \quad S.$$

We call $S/\equiv$ the *syntactic image* of S, which is syntactic at zero. In other words, the syntactic semigroup associated to S is the smallest image under the homomorphism $f \colon S \to S/\equiv$ such that $f^{-1}(0) = 0$.

Example 2.12. Consider the semigroup $S = \{0, 1, 2, \ldots, n\}$, where multiplication is taking the minimum. The $\equiv$-classes are given by $\{1, 2, \ldots, n\}$ and $\{0\}$. Hence, the syntactic semigroup $S/\equiv$ associated to S is isomorphic to $\{0, 1\}$ with multiplication being minimum.

Example 2.13. The *Rees matrix semigroup* $(S; I, I'; P)$ is indexed by a semigroup S, two non-empty sets I and I', and a matrix P indexed by I and I' with entries $p_{i',i} \in S$ (see for example [39, Section 3.4]). It is the set $I \times S \times I'$ with multiplication

$$(i, s, i')(j, t, j') = (i, s p_{i',j} t, j').$$

The *Rees matrix semigroup with zero* $(S; I, I'; P)^{\square}$ is the set $I \times S \times I' \cup \{\square\}$, where the entries in P are in $S \cup \{\square\}$, with multiplication

$$(i, s, i')(j, t, j') = \begin{cases} (i, s p_{i',j} t, j') & \text{if } p_{i',j} \neq \square, \\ \square & \text{otherwise.} \end{cases}$$

Then the syntactic image of $(S; I, I'; P)^\square$ is isomorphic to $(\{1\}; \tilde{I}, \tilde{I}'; \tilde{P})$, where $\tilde{P}$ is a matrix of 0 and 1 without equal rows or columns.

It turns out that we can replace a semigroup with zero with its syntactic image without changing the upper bound on the mixing time of the underlying Markov chain, but the stationary distribution can change.

Theorem 2.14. *Let (S, A) be a finite semigroup S with zero and generators A, whose minimal ideal $K(S)$ is a left zero semigroup. Then the Markov chains associated to (S, A) and $(S/\equiv, f(A))$ have the same upper bound $\mathsf{Pr}(\tau > t)$ on the mixing time.*

Remark 2.15.
  (1) If the probability associated to the generator $a \in A$ is x_a, then the probability associated to the generator $b \in f(A)$ is $\sum_{a \in f^{-1}(b)} x_a$.
  (2) Note that the stationary distributions of the Markov chains associated to (S, A) and $(S/\equiv, f(A))$ may differ.

Proof of Theorem 2.14. Let $\mathcal{S}$ be the semaphore code corresponding to the ideal $K(S)$. Then for a codeword $s \in \mathcal{S}$, $f(s)$ is a codeword in the semaphore code corresponding to $K(S/\equiv)$. If the probabilities match up as in Remark 2.15, the random variable τ matches and hence the upper bound on the mixing time determined from $\mathsf{Pr}(\tau > t)$ matches. $\qquad\square$

Theorem 2.14 is powerful in the sense that the upper bound on the mixing time for Markov chains with potentially complicated stationary distributions can be deduced from those for small semigroups which are syntactic at zero.

Example 2.16. Let us continue with Example 2.12. The semigroup (S, A) with $S = \{0, 1\}$, $A = \{a, b\}$ and $a = 0, b = 1$ is syntactic. The minimal ideal $K(S)$ is $A^\star a A^\star$ and the semaphore code is $\mathcal{S} = b^\star a = \{b^j a \mid j \geqslant 0\}$. The left action on $\mathcal{S}$ is given by

$$a \cdot b^j a = a \qquad\qquad\qquad (\text{reset to } a),$$
$$b \cdot b^j a = b^{j+1} a \qquad\qquad\quad (\text{free}),$$

with stationary distribution

$$\Psi_{b^j a} = x_b^j x_a \qquad \text{for } j \geqslant 0.$$

Note that

$$E[\tau] = \sum_{j=0}^{\infty} (j+1) x_b^j x_a = x_a \frac{\partial}{\partial x_b} \left(\sum_{j=0}^{\infty} x_b^{j+1} \right) = x_a \frac{\partial}{\partial x_b} \frac{x_b}{1 - x_b} = \frac{x_a}{(1 - x_b)^2} = \frac{1}{x_a}.$$

In contrast, let us compute

$$\mathsf{Pr}(\tau > t) = \sum_{j=t}^{\infty} x_b^j x_a = \frac{x_a x_b^t}{1 - x_b} = x_b^t.$$

Indeed $\mathsf{Pr}(\tau > t) \leqslant \frac{E[\tau]}{t+1}$ as in Example 2.8.

Example 2.17. We can amend Example 2.16 by making the semigroup finite and aperiodic by imposing $b^w = b^{w+1}$. Using the methods in [39] (or comparing the in-flow with the out-flow), the stationary distribution can be derived to be

$$\Psi_{b^j a} = x_b^j x_a \qquad \text{for } 0 \leqslant j < w,$$
$$\Psi_{b^w a} = \frac{x_a x_b^w}{1 - x_b}.$$

The associated syntactic semigroup is $(\{0, 1\}, A)$, which means by Theorem 2.14 that the upper bound on the mixing time is unchanged, even though the stationary distribution is different.

Example 2.18. Let (S, A) be an arbitrary finite semigroup with generators $A = \{a_1, \dots, a_k\}$ (with or without zero). Let $S^\square$ be the semigroup with a zero $\square$ adjoined. Then

$$\left(S^\square/\equiv\right) = (\{\square, 1\}, A \cup \{\square\}).$$

In this setting the stationary distribution can be complicated, however the upper bound on the mixing time is trivial by Theorem 2.14

$$\mathsf{Pr}(\tau > t) = (1 - x_\square)^t.$$

Example 2.19. Consider the Rees matrix semigroup $S = B(2)$ of [39, Example 3.3] with generators $A = \{a, b\}$, where $a = (1, 2)$ and $b = (2, 1)$. The minimal ideal $K(S)$ is $A^\star \{aa, bb\} A^\star$ with semaphore code

$$\mathcal{S} = \{(ab)^\star aa, (ba)^\star bb, b(ab)^\star aa, a(ba)^\star bb\}.$$

The left action on $\mathcal{S}$ is given by

$$
\begin{aligned}
a \cdot (ab)^j aa &= aa & \text{(reset)}, \\
a \cdot (ba)^j bb &= a(ba)^j bb & \text{(free)}, \\
a \cdot b(ab)^j aa &= (ab)^{j+1} aa & \text{(free)}, \\
a \cdot a(ba)^j bb &= aa & \text{(reset)},
\end{aligned}
$$

and similarly with a and b interchanged. Note that

$$\mathsf{Pr}(\tau > 2k) = \sum_{j=k}^{\infty} (x_a^2 + x_b^2 + x_a + x_b)(x_a x_b)^j = \frac{(x_a x_b)^k (x_a^2 + x_b^2 + 1)}{1 - x_a x_b} = 2(x_a x_b)^k,$$

$$\mathsf{Pr}(\tau > 2k + 1) = \sum_{j=k}^{\infty} (x_a^2 + x_b^2 + x_a^2 x_b + x_b^2 x_a)(x_a x_b)^j = \frac{(x_a x_b)^k (x_a^2 + x_b^2 + x_a x_b)}{1 - x_a x_b} = (x_a x_b)^k,$$

which by Theorem 2.1 gives an upper bound on the mixing time.

Example 2.20. Consider the Rees matrix semigroup (see Example 2.13) with $I = I' = \{1, 2\}$, $S = \{0, 1\}$,

$$P = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

and generators $A = \{a, b\}$ with $a = (1, 1, 2)$ and $b = (2, 1, 1)$. The minimal ideal is $A^\star aa A^\star$ with semaphore code $\mathcal{S} = b^\star (abb^\star)^\star aa$. The left action on $\mathcal{S}$ is given by

$$a \cdot b^j \left(\prod_{k=1}^{\ell} abb^{e_k} \right) aa = \begin{cases} ab^j \left(\prod_{k=1}^{\ell} abb^{e_k} \right) aa & \text{if } j > 0 \text{ (free)}, \\ aa & \text{if } j = 0 \text{ (reset)}, \end{cases}$$

$$b \cdot b^j \left(\prod_{k=1}^{\ell} abb^{e_k} \right) aa = b^{j+1} \left(\prod_{k=1}^{\ell} abb^{e_k} \right) aa \qquad \text{(free)}.$$

In this case, the bound on the mixing time is given by

$$\mathsf{Pr}(\tau > k) = x_a^2 \sum_{j \geqslant k-1} \sum_{i=0}^{\lfloor \frac{j}{2} \rfloor} \binom{j-i}{i} x_a^i x_b^{j-i}.$$

2.10. **Ideals and d-testable languages.** As we have seen, ideals are important in the study of Markov chains in the context of semigroups. In addition, ideals are closely related to semaphore codes.

Let (S_j, A) be two semigroups with zero for $j = 1, 2$ with the same generating set A and I_j the ideal of strings in A^+ that is zero in (S_j, A) for $j = 1, 2$. Let $\mathcal{S}_j$ for $j = 1, 2$ be the semaphore code associated to the ideal I_j. Recall that through the left action of A^+ on $\mathcal{S}_j$ we have two Markov chains.

Remark 2.21 (*Ideal principle*). If $I_1 \subseteq I_2$, the upper bound on the mixing time of the Markov chain associated to $\mathcal{S}_2$ is smaller or equal to the upper bound on the mixing time of the Markov chain associated to $\mathcal{S}_1$.

Remark 2.21 is true since by [6, Corollary 3.5(3)] the mixing time is bounded above by $P^{\star t}(S \setminus K(S))$ (see Theorem 2.1). If $I_1 \subseteq I_2$, we hence have

$$P^{\star t}(S_2 \setminus I_2) \leqslant P^{\star t}(S_1 \setminus I_1),$$

since I_j consists of all words in A^+ which are zero in S_j.

By Remark 2.21 we want to study Markov chains with the smallest ideals as they have the worst mixing time. To this end, we will study the *complete lattice of ideals* of $A^\star$. All ideals (including $\emptyset$) of $A^\star$ form a complete lattice under union and intersection.

Lemma 2.22. *Every nonempty ideal I has a* descending chain

$$I \supset A^\star t_1 A^\star \supset A^\star t_2 A^\star \supset \cdots \supset A^\star t_k A^\star \supset \cdots.$$

Proof. Since $I \neq \emptyset$, there exists an element $t_1 \in I$. The unique smallest length element in $A^\star t_1 A^\star$ is of length $|t_1|$. Choose $t_2 \in A^\star t_1 A^\star$ with $|t_2| > |t_1|$. Then $A^\star t_1 A^\star \supset A^\star t_2 A^\star$ and repeat. $\qquad\square$

Some ideals I have an infinite *ascending chain*

$$I \subset I_1 \subset I_2 \subset \cdots$$

and some do not. Let $A = \{a, b\}$. Then $A^\star \setminus \{a\}$, for example, does not have an infinite ascending chain. On the other hand (compare also with Example 2.20)

$$A^\star aa A^\star \subset A^\star aa A^\star \cup A^\star aba A^\star \subset \cdots \subset \bigcup_{j=0}^{k} A^\star ab^j a A^\star \subset \cdots$$

does.

Every ideal $I \subseteq A^+$ has a unique set of minimal generators, namely all $t = a_1 a_2 \cdots a_{\ell-1} a_\ell \in I$ such that $a_1 \cdots a_{\ell-1} \notin I$ and $a_2 \cdots a_\ell \notin I$. Hence by Lemma 2.22, the smallest ideals are of the form $A^\star t A^\star$, where $|t|$ is big. Since by Remark 2.21 smaller ideals have worse upper bounds on the mixing times, we would like to analyze ideals of the form $A^\star t A^\star$, where $|t|$ is large. This is related to d-testable languages, which are finite ideals generated by $\bigcup_{i=1}^n A^\star t_i A^\star$, see [49].

Let $t \in A^+$. The minimal automata $\mathsf{Test}(t)$ accepting the language $A^\star t A^\star$ for $t = a_1 a_2 \ldots a_\ell$ is given as follows. There are $\ell + 1$ states: $\mathbb{1}, a_1, a_1 a_2, \ldots, a_1 \ldots a_{\ell-1}, a_1 \ldots a_\ell \equiv 0$. We have $q \xrightarrow{a} qa$ if both q and qa are prefixes of t and otherwise $q \xrightarrow{a} \mathbb{1}$.

Using [38, Definition 3.5], $\mathsf{Test}(t)$ can be transformed into a loop graph with loops labeled by words $w \in A^+$ such that $|w| \leqslant \ell$, $w = w_1 \cdots w_k$ is not a prefix of t, but $w_1 \cdots w_{k-1}$ is a prefix of t. Let us denote the set of all such words W_t. Hence the Kleene expression for the paths in $\mathsf{Test}(t)$ is $(\cup_{w \in W_t} \{w\})^\star t$ and hence the stationary distribution is

$$\Psi_t = \frac{x_{a_1} \cdots x_{a_\ell}}{1 - \sum_{w \in W_t} \prod_{a \in w} x_a}.$$

By Theorem 2.5 we obtain

$$E_t[\tau] = \ell + \frac{\sum_{w \in W_t} |w| \prod_{a \in w} x_a}{1 - \sum_{w \in W_t} \prod_{a \in w} x_a},$$

which gives an upper bound on the mixing time using the Markov inequality (2.1). Theorem 2.4 can also be used to obtain an upper bound on the mixing time using the series expansion of Ψ_t.

Remark 2.23. The loop graphs in [38] are not allowed to have loops at vertex $\mathbb{1}$. Here we do allow loops at $\mathbb{1}$. To remedy the situation, one could rename $\mathbb{1}$ by 1 and have an edge with probability 1 from $\mathbb{1}$ to 1.

Example 2.24. Let $A = \{a, b\}$ and $t = aba$. Then $\mathsf{Test}(t)$ can be depicted by
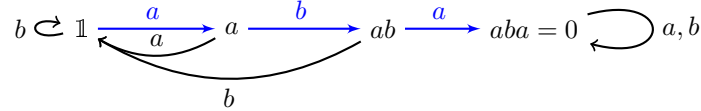
The corresponding loop graph is
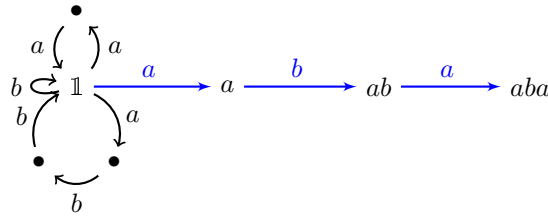
Hence the stationary distribution is

$$\Psi_{aba} = \frac{x_a^2 x_b}{1 - x_b - x_a^2 - x_a x_b^2}.$$

By Theorem 2.5, this hence gives

$$E_t[\tau] = 3 + \frac{x_b + 2x_a^2 + 3x_a x_b^2}{1 - x_b - x_a^2 - x_a x_b^2}.$$

Example 2.25. Now let us take $A = \{a, b\}$ and $t = a^\ell$. In this case $W_t = \{a^k b \mid 0 \leqslant k < \ell\}$ and hence

$$\Psi_t = \frac{x_a^\ell}{1 - \sum_{k=0}^{\ell-1} x_a^k x_b}$$

with an upper bound for the mixing time given by

$$E_t[\tau] = \ell + \frac{\sum_{k=0}^{\ell-1}(k+1)x_a^k x_b}{1 - \sum_{k=0}^{\ell-1} x_a^k x_b}$$

using (2.1).

3. Examples

In this section, we analyze the mixing time of several examples using the methods developed in Section 2. In Section 3.1 we derive upper bounds for the mixing time of the famous Tsetlin library [17] and in Section 3.2 for edge flipping on a line [18]. In Section 3.3, we provide a new Markov chain on linear extension of a poset with n vertices, inspired by but different from the promotion Markov chain of Ayyer, Klee and the last author. The mixing time of this Markov chain is $O(n \log n)$ (Theorem 3.11).
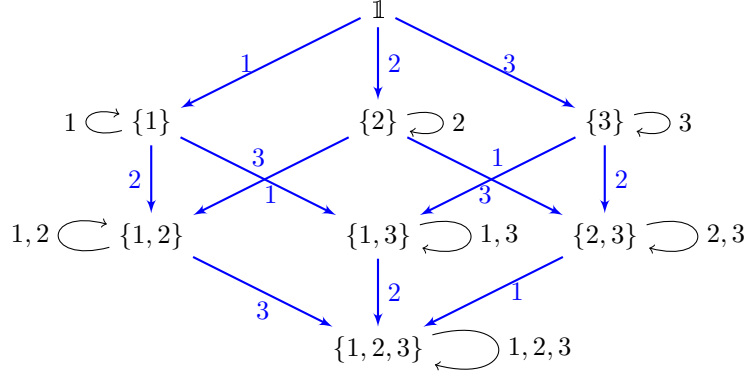
FIGURE 2. The right Cayley graph $\mathsf{RCay}(S, A)$ with $S = P(3)$ and $A = \{1, 2, 3\}$.
Transition edges are drawn in blue.

3.1. **The Tsetlin library.** The Tsetlin library [17] is a Markov chain whose states are all permutations S_n of n books (on a shelf). Given $\pi \in S_n$, construct $\pi' \in S_n$ from π by removing book a from the shelf and inserting it to the front. In this case write $\pi \xrightarrow{a} \pi'$. Let $0 < x_a \leqslant 1$ be probabilities for each $1 \leqslant a \leqslant n$ such that $\sum_{a=1}^{n} x_a = 1$. In the Tsetlin library Markov chain, we transition $\pi \xrightarrow{a} \pi'$ with probability x_a. The stationary distribution for the Tsetlin library was derived by Hendricks [26, 27] and Fill [22]

$$(3.1) \qquad \Psi_\pi = \prod_{i=1}^{n} \frac{x_{\pi_i}}{1 - \sum_{j=1}^{i-1} x_{\pi_j}} \qquad \text{for all } \pi \in S_n.$$

The stationary distribution was derived using right Cayley graphs and their Karnofsky–Rhodes and McCammond expansions in [39, Section 3.1].

Consider the semigroup $P(n)$, which consists of the set of all non-empty subsets of $\{1, 2, \ldots, n\}$. Multiplication in $P(n)$ is union of sets. We pick as generators $A = [n] := \{1, 2, \ldots, n\}$. Then the right Cayley graph $\mathsf{RCay}(P(n), [n])$ is the Boolean poset with $\mathbb{1}$ as root. The right Cayley graph for $P(3)$ is depicted in Figure 2. Except for the loops at a given vertex, all edges are transitional. Hence $\mathsf{Mc} \circ \mathsf{KR}(P(n), [n]) = \mathsf{KR}(P(n), [n])$ is a tree with leaves given by the permutations S_n of $[n]$. The case $n = 3$ is depicted in Figure 3.

To obtain an upper bound on the mixing time, we compute $E[\tau]$ from the Karnofsky–Rhodes expansion of the right Cayley graph. The ideal consists of the leaves of the tree $\mathsf{KR}(P(n), [n])$, which are labeled by permutations in S_n. Recall that $E[\tau]$ can be computed via (2.2). Any path from $\mathbb{1}$ to the ideal is of length at least n. Hence $\mathsf{Pr}(\tau \geqslant t) = 1$ for $1 \leqslant t \leqslant n$.

Now for concreteness consider the loop graph G associated to the path from $\mathbb{1}$ to $12 \ldots n$ in $\mathsf{Mc} \circ \mathsf{KR}(P(n), [n])$. The contributions of the loops can be treated in a similar fashion to Example 2.8. The Kleene expression for all paths from $\mathbb{1}$ to $12 \ldots n$ is given by

$$11^\star 2\{1, 2\}^\star 3\{1, 2, 3\}^\star \ldots \{1, 2, \ldots, n - 1\}^\star n.$$

Hence we obtain (compare with (3.1))

$$\Psi_G(x_1, \ldots, x_n) = \frac{x_1 \cdots x_n}{(1 - x_1)(1 - x_1 - x_2) \cdots (1 - x_1 - \cdots - x_{n-1})}$$

and by Theorem 2.5

$$(3.2) \qquad E_G[\tau] = n + \frac{x_1}{1 - x_1} + \frac{x_1 + x_2}{1 - x_1 - x_2} + \cdots + \frac{x_1 + \cdots + x_{n-1}}{1 - x_1 - \cdots - x_{n-1}},$$
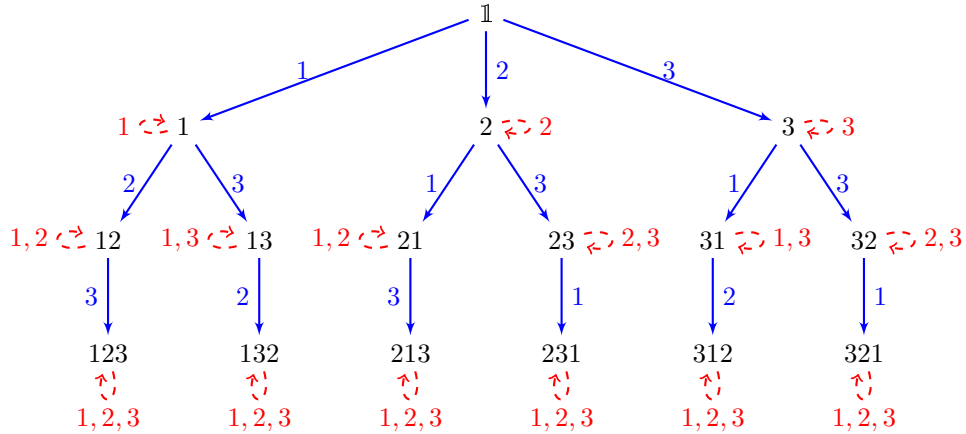
FIGURE 3. $\mathsf{Mc} \circ \mathsf{KR}(P(3), [3]) = \mathsf{KR}(P(3), [3])$, which is the Karnofsky–Rhodes expansion of the right Cayley graph of Figure 2.

which can also be checked directly. If $x_i = \frac{1}{n}$ for all $1 \leqslant i \leqslant n$, we hence have

$$(3.3) \qquad E_G[\tau] = n + \frac{1}{n-1} + \frac{2}{n-2} + \cdots + \frac{n-1}{1} = n \left(\sum_{i=1}^{n} \frac{1}{i} \right).$$

The last equality can be proved by induction on n. It is well-known that the sequence $t_n = \sum_{i=1}^{n} \frac{1}{i} - \ln(n)$ approaches the Euler–Mascheroni constant γ as $n \to \infty$. Therefore

$$E[\tau] = E_G[\tau] \leqslant n \ln(n) + n\gamma$$

and by (2.1)

$$\|T^t \nu - \pi\| \leqslant \frac{n \ln(n) + n\gamma}{t+1}.$$

Nestoridi [36] has proven upper/lower bounds for the mixing time of the separation distance. Pike [37] has discussed the eigenfunctions of the transition matrix. Note that, given the rational expression of the stationary distribution (3.1), our methods work for general weights x_i. Truncating the degree of the expansion of the stationary distribution (3.1) gives a precise expression for an upper bound of the mixing time by Theorem 2.4.

3.2. **Edge flipping on a line.** In [39, Section 3.2], we treated the Markov chain obtained by edge flipping on a line using the semigroup methods of [39]. Take a line with $n+1$ vertices. Each vertex can either be 0 or 1. So the state space is $\Omega = \{0,1\}^{n+1}$ of size 2^{n+1}. Pick edge i for $1 \leqslant i \leqslant n$ (between vertices i and $i+1$) with probability x_i. Then with probability $\frac{1}{2}$ make the adjacent vertices both 0 (respectively both 1). Let us call this Markov chain $\mathcal{M}$. This Markov chain is a Boolean arrangement [8] for which the stationary distribution was derived in [15] and which was also analyzed in [18].

In [39, Section 3.2], we analyzed the stationary distribution in a similar fashion to the Tsetlin library by considering the semigroup $P^{\pm}(n)$, which is the set of signed subsets of $[n]$. That is, take a subset of $[n]$ and in addition associate to each letter a sign $+$ or $-$. Right multiplication of such a subset X by a generator $x \in [\pm n] := \{\pm 1, \ldots, \pm n\}$ is addition of x to X if neither x nor $-x$ are in X and otherwise return X. The minimal ideal in the Karnofsky–Rhodes expansion of this monoid is the set of signed permutations $S_n^{\pm}$. In the Markov chain on the minimal ideal, we transition from $\pi \xrightarrow{a} \pi'$ with probability y_a for $a \in [\pm n]$, where π' is obtained from π by prepending a to π and removing the letter a or $-a$ from π. The stationary distribution associated to $\pi \in S_n^{\pm}$ was computed

to be

$$(3.4) \qquad \Psi_\pi^{\mathsf{KR}(P^\pm(n),[\pm n])} = \prod_{i=1}^{n} \frac{y_{\pi_i}}{1 - \sum_{j=1}^{i-1}(y_{\pi_j} + y_{-\pi_j})}.$$

The stationary distribution for a word $s \in \Omega$ for the Markov chain $\mathcal{M}$ is a lumping (or sum) of the $\Psi_\pi^{\mathsf{KR}(P^\pm(n),[\pm n])}$ in (3.4). By the same analysis as in Section 3.1 the mixing time for $\mathcal{M}$ is of order $O(n \ln(n))$.

3.3. **Promotion Markov chain.** Let P be a *partially ordered set*, also known as a *poset*, on n elements with partial order $\preccurlyeq$. A partial order must be reflexive ($a \preccurlyeq a$ for all $a \in P$), antisymmetric ($a \preccurlyeq b$ and $b \preccurlyeq a$ implies $a = b$ for $a, b \in P$), and transitive ($a \preccurlyeq b$ and $b \preccurlyeq c$ implies $a \preccurlyeq c$ for $a, b, c \in P$). We assume that the elements of P are labeled by integers in $[n] := \{1, 2, \ldots, n\}$ such that if $i, j \in P$ with $i \preccurlyeq j$ then $i \leqslant j$ as integers. Let $\mathcal{L} := \mathcal{L}(P)$ be the set of *linear extensions* of P defined as

$$\mathcal{L}(P) = \{\pi \in S_n \mid i \prec j \text{ in } P \implies \pi_i^{-1} < \pi_j^{-1} \text{ as integers}\}.$$

In computer science, linear extensions are also known as *topological sortings* [29, 30]. Computing the number of linear extensions is an important problem for real world applications [28]. For example, it relates to sorting algorithms. Suppose one wants to schedule a sequence of tasks based on their dependencies. Specifying that a certain task has to come before another task gives rise to a partial order. A linear extension gives a total order in which to perform the jobs. In social sciences, linear extensions are used in voting procedures [23, 1], where voters rank the candidates according specified traits (view on foreign policies, view on domestic policies etc). A recursive formula for the number of linear extensions for a given poset P was given in [21]. Brightwell and Winkler [13] showed that counting the number of linear extensions is $\#P$-complete. Bubley and Dyer [16] provided an algorithm to (almost) uniformly sample the set of linear extensions of a finite poset of size n with mixing time $O(n^3 \log n)$. In [4], the promotion Markov chain was introduced, which is a random walk on the linear extensions of a finite poset P. Here we discuss a variant of the promotion Markov chain which has mixing time of order $O(n \log n)$.

3.3.1. *The model.* We now explain the promotion Markov chain introduced in [4]. For a given poset P with n vertices, the state space of the *promotion Markov chain* is the set of linear extensions $\mathcal{L}(P)$. For $\pi, \pi' \in \mathcal{L}(P)$, we transition $\pi \xrightarrow{\partial_j} \pi'$ with probability x_{π_j} if $\pi' = \partial_j \pi$, where ∂_j is the promotion operator. The promotion operator is defined in terms of more elementary operators τ_i ($1 \leqslant i < n$) which appeared in [25, 32, 47] and was used explicitly to count linear extensions in [21]. Let $\pi = \pi_1 \ldots \pi_n \in \mathcal{L}(P)$ be a linear extension of P in one-line notation. Then

$$(3.5) \qquad \tau_i \pi = \begin{cases} \pi_1 \ldots \pi_{i-1} \pi_{i+1} \pi_i \ldots \pi_n & \text{if } \pi_i \text{ and } \pi_{i+1} \text{ are not comparable in } P, \\ \pi_1 \ldots \pi_n & \text{otherwise.} \end{cases}$$

In other words, τ_i acts non-trivially on a linear extension if interchanging entries π_i and π_{i+1} yields another linear extension. Then the *promotion operator* on $\mathcal{L}(P)$ is defined as

$$(3.6) \qquad \partial_j = \tau_1 \tau_2 \cdots \tau_{j-1}.$$

Note that we use a different convention here to [4], where $\partial_j = \tau_j \tau_{j+1} \cdots \tau_{n-1}$. Our convention here is compatible with the conventions for the Tsetlin library as in Section 3.1, where we moved letters to the front of the word rather than the end of the word.

Example 3.1. Let P be the poset on four vertices defined by its covering relations $\{(1,4), (2,4), (2,3)\}$. Then its Hasse diagram is the following:
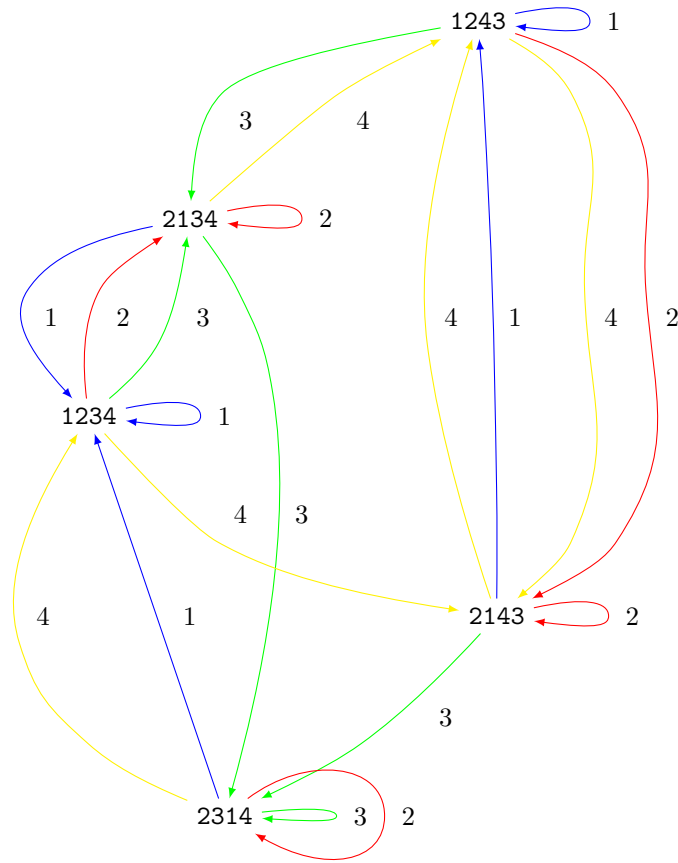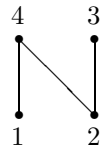
FIGURE 4. The promotion Markov chain digraph for the poset in Example 3.1.

This poset has five linear extensions

(3.7) $$\mathcal{L}(P) = \{1234, 1243, 2134, 2143, 2314\}.$$

The promotion Markov chain for P is depicted in Figure 4, where the vertices are the linear extensions and an arrow labelled by i from π to π' indicates that $\pi' = \partial_i \pi$.

We may represent the promotion operator ∂_i by a $|\mathcal{L}(P)| \times |\mathcal{L}(P)|$-dimensional matrix, where row k and column j contains 1 if the j-th linear extension in (3.7) is mapped to the k-th linear extension in (3.7) under ∂_i; the rest of the entries are zero. For example, ∂_1 is represented by the matrix

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

The right Cayley graph of the monoid generated by the matrices for the promotion operators $\partial_1, \partial_2, \partial_3, \partial_4$ is depicted in Figure 5. The vertices in the right Cayley graph are labeled by reduced words in the generators. For example $[1, 4, 1]$ stands for the element $\partial_1 \partial_4 \partial_1$.

We prove some useful properties of the right Cayley graph of the semigroup S generated by ∂_i for $1 \leqslant i \leqslant n$.

Proposition 3.2. *Any element in $K(S)$ can be written as $\partial_{w_1} \cdots \partial_{w_{n-1}}$, where $w_1, \ldots, w_{n-1} \in \{1, 2, \ldots, n\}$ are distinct. In particular, the length of any reduced word for the elements in $K(S)$ is less than n.*

Proof. Each element in $K(S)$ corresponds to a linear extension in $\mathcal{L}(P)$. For a given $\pi \in \mathcal{L}(P)$, we now construct a word $w_1 \ldots w_{n-1}$ with distinct letters such that $\pi = \partial_{w_1} \cdots \partial_{w_{n-1}} \pi'$ for all $\pi' \in \mathcal{L}(P)$. In particular, this means that $\partial_{w_1} \cdots \partial_{w_{n-1}} \in K(S)$.

Write $\pi = \pi_1 \ldots \pi_n$ in one-line notation and set $\pi^{(1)} = \pi$. Construct $\pi^{(m+1)}$ from $\pi^{(m)}$ for $1 \leqslant m < n$ as follows. Set $i_1^{(m)} = 1$ and then recursively find the smallest $i_{j+1}^{(m)} > i_j^{(m)}$ such that $\pi_{i_j^{(m)}}^{(m)} \prec \pi_{i_{j+1}^{(m)}}^{(m)}$ if possible. If there is no such $i_{j+1}^{(m)}$, set $k^{(m)} = j$. Define $w_m = \pi_{i_{k^{(m)}}^{(m)}}^{(m)}$. Next construct $\pi^{(m+1)}$ from $\pi^{(m)}$ by removing $\pi_1^{(m)}$ and replacing $\pi_{i_j^{(m)}}^{(m)}$ by $\pi_{i_{j-1}^{(m)}}^{(m)}$ for $2 \leqslant j \leqslant k^{(m)}$.

Next we show that $\pi = \partial_{w_1} \cdots \partial_{w_{n-1}} \pi'$ for any $\pi' \in \mathcal{L}(P)$, proving that $\partial_{w_1} \cdots \partial_{w_{n-1}} \in K(S)$ corresponding to the linear extension π. We will do so by induction on n. For $n = 2$, P is either the antichain with vertex 1 incomparable to vertex 2 or 2 is bigger than 1. In the first case, there are two linear extension $\pi = 12$ or 21. The algorithm determines $w = \pi_1$ and indeed $\partial_{\pi_1}(12) = \partial_{\pi_1}(21) = \pi_1 \pi_2 = \pi$. In the second case, there is only one linear extension $\pi = 12$ and the algorithm determines $w = 2$. Indeed $\partial_2(12) = 12$.

Now assume by induction that the algorithm works for posets with strictly less than n vertices. In particular, for $\pi^{(2)}$ from the algorithm $\pi^{(2)} = \partial_{w_2} \cdots \partial_{w_{n-1}} \pi'$ for any linear extension π' of the poset P' obtained from P by deleting the vertex w_1. Also, by induction $w_2, \ldots, w_{n-1}$ are distinct and different from w_1. Note that w_1 is a maximal element in P. Hence for any linear extension π' of P, we have that $\partial_{w_2} \cdots \partial_{w_{n-1}} \pi'$ is a linear extension of P such that removing the letter w_1 results in $\pi^{(2)}$. Let $\sigma \in \mathcal{L}(P)$ be such a linear extension, that is, $\sigma \setminus w_1 = \pi^{(2)}$. Consider the saturated chain $\pi_1 = a_1 \prec a_2 \prec \cdots \prec a_k = w_1$ in P from π_1 to w_1. Such a chain exists by the definition of w_1. In $\pi^{(2)}$ and hence also in σ the letter a_{k-1} is the rightmost letter that is covered in P by w_1. This is since by the algorithm to construct $\pi^{(2)}$, the letter a_{k-1} replaced the letter $a_k = w_1$ in π. In σ, the letter w_1 must sit to the right of the letter a_{k-1} since $a_{k-1} \prec w_1$. Hence, when acting with ∂_{w_1} on σ, the letter w_1 interchanges with all letters to its left until it reaches the letter a_{k-1}. By the action of τ_i as in (3.5), the letter w_1 will stay in the position where a_{k-1} was in σ and then the letter a_{k-1} starts moving left. The letter a_{k-2} is the rightmost letter in σ that is covered by a_{k-1} in P, again by the definition of the algorithm. The letter a_{k-1} replaces the letter a_{k-2} and a_{k-2} starts moving left and so on. Finally, the letter $a_1 = \pi_1$ moves into first position. Hence $\partial_{w_1} \sigma = \pi$. This proves the claim. $\square$

Example 3.3. Take the poset from Example 3.1 and the linear extension $\pi = 1243$. Set $\pi^{(1)} = \pi$. The first sequence of increasing entries in $\pi^{(1)}$ is given by the underlined entries

$$\underline{1}\underline{2}\underline{4}3.$$

Hence $w_1 = 4$ and $\pi^{(2)} = 213$. The next sequence of increasing entries is given by

$$\underline{2}1\underline{3}.$$

Hence $w_2 = 3$ and $\pi^{(3)} = 12$. Next we find the increasing sequence $\underline{1}2$, so that $w_3 = 1$. Indeed, comparing with Figure 5, we see that

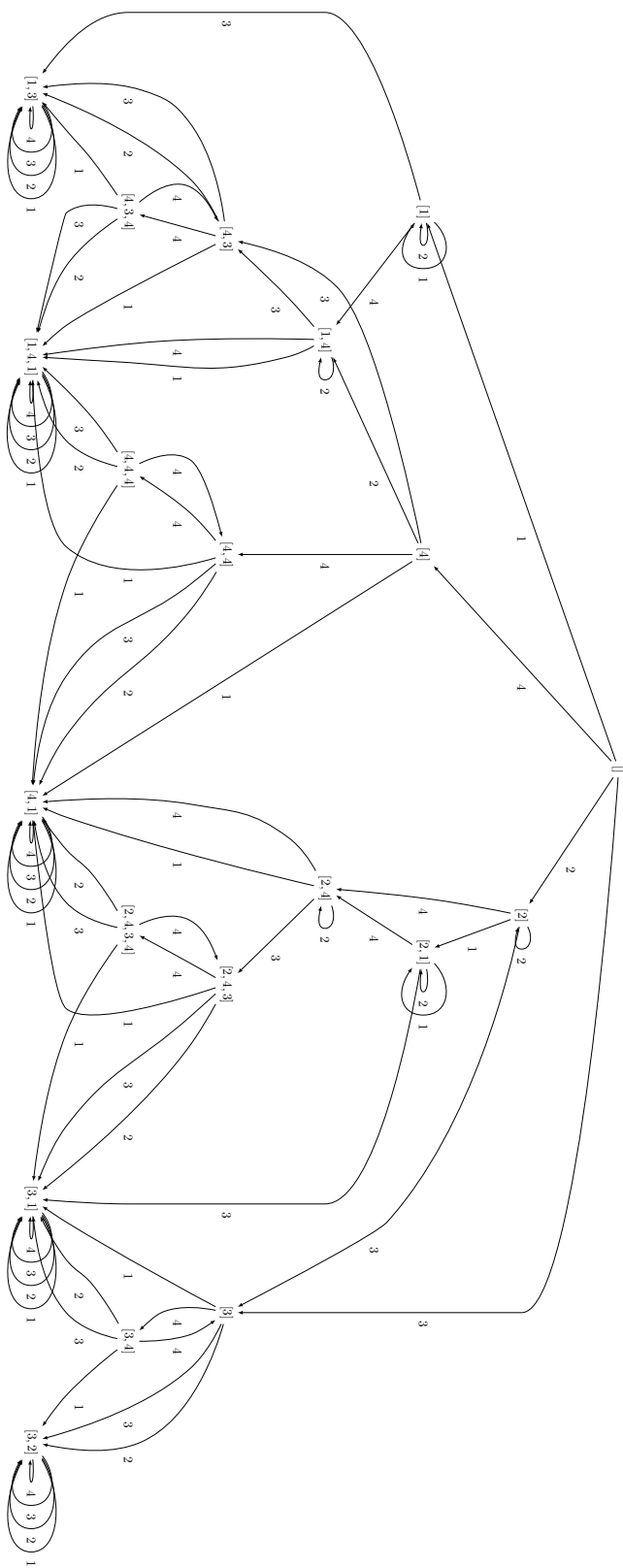$$\partial_4 \partial_3 \partial_1 = \partial_1 \partial_4 \partial_1$$

FIGURE 5. Right Cayley graph for the promotion Markov chain of Example 3.1.

is in $K(S)$.

Note that the above algorithm does not always give a shortest path to the ideal in the right Cayley graph. For example, if $\pi = 2143$ the algorithm gives

$$\underline{2}1\underline{4}3 \to \underline{1}23 \to \underline{23} \to 2,$$

so that $w_1 w_2 w_3 = 413$. From Figure 5, we see that $\partial_4 \partial_1 \partial_3 = \partial_4 \partial_1$ is in $K(S)$.

The (unnormalized) stationary distribution of the promotion Markov chain was computed in [4, Theorem 4.5]. Recall that our conventions are different from [4].

Theorem 3.4. [4, Theorem 4.5] *The (unnormalized) stationary distribution for the promotion Markov chain* Ψ_π *for* $\pi \in \mathcal{L}(P)$ *for a finite poset* P *with* $n = |P|$ *is given by*

$$(3.8) \qquad\qquad \Psi_\pi = \prod_{i=1}^{n} \frac{1}{1 - (x_{\pi_1} + \cdots + x_{\pi_{i-1}})}.$$

Despite the fact that by Proposition 3.2 the right Cayley graph is shallow in the sense that each vertex is at most $n-1$ steps away from the minimal ideal and the existence of an explicit formula for the stationary distribution, this is not enough to give a tight bound on the mixing time. The reason is that the expression for Ψ_π does not have the property required in Theorems 2.4 and 2.5 that each term of degree ℓ in its formal power sum expansion corresponds to a semaphore code word s of length ℓ. Furthermore, the $\mathscr{R}$-classes (or strongly connected components) in the right Cayley graph can become very big, especially when P has a maximal element. This makes it hard to analyze the mixing time for the promotion Markov chain in general. Here we propose a new Markov chain on linear extensions of a poset which gives rise to an $\mathscr{R}$-trivial semigroup (where all strongly connected components have size one).

3.3.2. *A variant of the promotion Markov chain.* As before let P be a poset with n elements and $\mathcal{L}(P)$ the set of linear extensions of P. Denote by $\mathcal{W}(P)$ the set of subwords of linear extensions in $\mathcal{L}(P)$ and set $A = [n]$. We define a semigroup on $\mathcal{W}(P)$ as follows. Let $w \in \mathcal{W}(P)$ and $a \in A$. Then define

$$(3.9) \qquad\qquad wa = \begin{cases} w & \text{if } a \in w, \\ \mathsf{straight}(wa) & \text{if } a \notin w. \end{cases}$$

Here $\mathsf{straight}(wa)$ is defined as follows. If wa is a subword of a linear extension of P, then $\mathsf{straight}(wa) = wa$. If not, write $w = w_1 \ldots w_k$ and find the largest $1 \leqslant j_1 \leqslant k$ such that $a \prec w_{j_1}$ in P. Interchange w_{j_1} and a. Repeat by finding the largest $1 \leqslant j_2 < j_1$ such that $a \prec w_{j_2}$. Interchange w_{j_2} and a. Repeat until no further element bigger than a exists to the left. The result is $\mathsf{straight}(wa)$.

Example 3.5. Take the poset P of Example 3.1, $w = 234 \in \mathcal{W}(P)$, and $a = 1$. We have $1 \prec 4$, so $j_1 = 3$. Both 2 and 3 are incomparable to 1, so we find $\mathsf{straight}(wa) = 2314 \in \mathcal{L}(P)$.

Lemma 3.6. *Let* $a \in A$ *and* $w \in \mathcal{W}(P)$ *such that* $a \notin w$. *Then* $\mathsf{straight}(wa) \in \mathcal{W}(P)$.

Proof. Since j_1 is largest such that $a \prec w_{j_1}$, either $w_j \prec a$ or w_j and a are incomparable for $j_1 < j \leqslant k$. If $w_j \prec a$ by transitivity we find that $w_j \prec w_{j_1}$ which contradicts the fact that $w \in \mathcal{W}(P)$. Hence a is incomparable with w_j for all $j_1 < j \leqslant k$. Suppose $w_{j_1} \prec w_j$ for some $j_1 < j \leqslant k$. Then again by transitivity, we have $a \prec w_j$. This contradicts the maximality of j_1. Hence w_{j_1} is incomparable to w_j for all $j_1 < j \leqslant k$. Therefore $a w_{j_1+1} \cdots w_k w_{j_1} \in \mathcal{W}(P)$. Repeating similar arguments for the next segments (interchanging a with w_{j_2} etc), we find $\mathsf{straight}(wa) \in \mathcal{W}(P)$. □

Proposition 3.7. *The set* $\mathcal{W}(P)$ *together with the product defined in* (3.9) *forms a semigroup.*

Proof. Note that by the proof of Lemma 3.6, the letters inbetween any letters that are interchanged by the product are incomparable to the interchanged letters. By transitivity, if there are three letters that are interchanged, say $w_i \ldots w_j \ldots w_k$ with $w_k \prec w_j \prec w_i$, it does not matter in which order
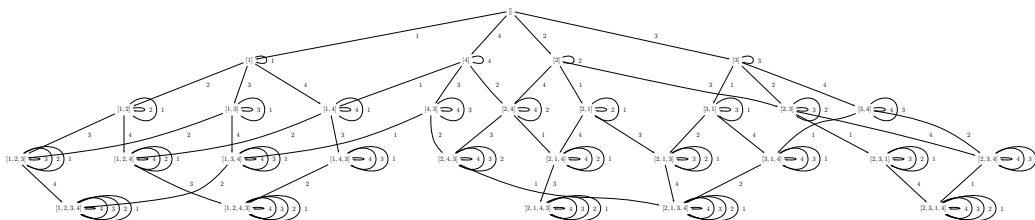
FIGURE 6. The right Cayley graph of $(\mathcal{W}(P), A)$ for the poset of Example 3.1.

this is done, the end results is $w_k \ldots w_j \ldots w_i$. This proves that the product is associative and hence $\mathcal{W}(P)$ is a semigroup with the product in (3.9). $\qquad\square$

Let us now define $(\mathcal{W}(P), A)$ to be the semigroup with product (3.9) and generators $A = [n]$.

Theorem 3.8. *The semigroup $(\mathcal{W}(P), A)$ is $\mathcal{R}$-trivial.*

Proof. In the product, the length of the word can either stay the same or increase. When the length stays the same, the word does not change. This proves that $(\mathcal{W}(P), A)$ is $\mathcal{R}$-trivial. $\qquad\square$

Example 3.9. The right Cayley graph of $(\mathcal{W}(P), A)$ for the poset of Example 3.1 is given in Figure 6.

Note that the minimal ideal of $(\mathcal{W}(P), A)$ is the set of linear extensions $\mathcal{L}(P)$ of the poset P. Let $\mathcal{M}(\mathcal{W}(P), A)$ be the Markov chain on $\mathcal{L}(P)$ induced by the semigroup $(\mathcal{W}(P), A)$. More precisely, we transition from $\pi \in \mathcal{L}(P)$ to $a\pi \in \mathcal{L}(P)$ with probability x_a.

Proposition 3.10. $\mathcal{M}(\mathcal{W}(P), A)$ *is ergodic.*

Proof. Note that $\pi\pi' = \pi$ for all $\pi, \pi' \in \mathcal{L}(P)$. Hence the graph of the Markov chain is strongly connected and hence it is irreducible. Furthermore, if $\pi = \pi_1 \ldots \pi_n \in \mathcal{L}(P)$, then $\pi_1 \pi = \pi$, which means the Markov chain is aperiodic. $\qquad\square$

The stationary distribution for $\mathcal{M}(\mathcal{W}(P), A)$ is given by

$$\Psi_\pi = \sum_{\substack{\sigma \in S_n \\ [\sigma]_{\mathcal{W}(P)} = \pi}} \left(\prod_{i=1}^n \frac{x_{\sigma_i}}{1 - \sum_{j=1}^{i-1} x_{\sigma_j}} \right) \qquad \text{for all } \pi \in \mathcal{L}(P).$$

Theorem 3.11. *The expected value $E[\tau]$ for $\mathcal{M}(\mathcal{W}(P), A)$ is bounded above by $n \ln(n) + n\gamma$.*

Proof. For a word $w \in \mathcal{W}(P)$, its length $|w| = k$ is bounded by $0 \leqslant k \leqslant n$. For a word of length $|w| = k$, there are $n - k$ transition arrows in $\mathsf{RCay}(\mathcal{W}(P), A)$ originating at w, given by all the letters that do not appear in w. Hence by the same arguments as for the Tsetlin library $E[\tau] \leqslant n \ln(n) + n\gamma$. $\qquad\square$

Remark 3.12. Note that the Markov chain $\mathcal{M}(\mathcal{W}(P), A)$ is not identical to the promotion Markov chain. For example, left multiplication by 4 on 2143 in $(\mathcal{W}(P), \{1, 2, 3, 4\})$ for the poset in Example 3.1 yields 2143, whereas we see from Figure 4 that in the promotion Markov chain 2143 goes to 1243 under ∂_4. The full Markov chain transition diagram is given in Figure 7.

Theorem 3.11 shows that the mixing time for $\mathcal{M}(\mathcal{W}(P), [n])$ is of order $O(n \log n)$. Of course, this does not take the computational complexity of computing the product (3.9) into account. For a word of length k, this involves up to k swaps.
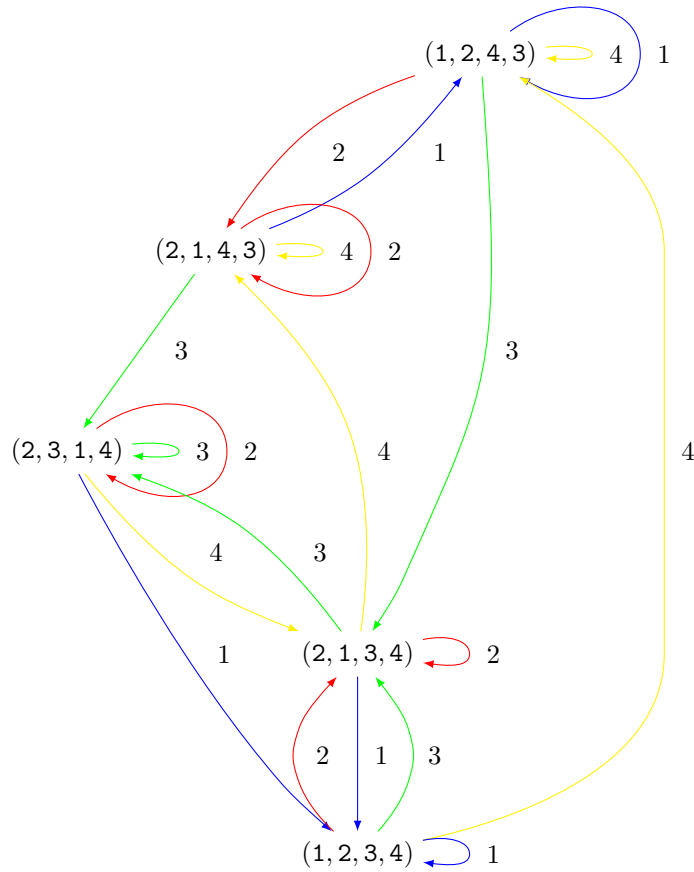
FIGURE 7. The Markov chain $\mathcal{M}(\mathcal{W}(P), [4])$ for the poset of Example 3.1.

REFERENCES

[1] Michael Ackerman, Sul-Young Choi, Peter Coughlin, Eric Gottlieb, and Japheth Wood. Elections with partially ordered preferences. *Public Choice*, 157(1/2):145–168, 2013.

[2] R. Arratia and L. Gordon. Tutorial on large deviations for the binomial distribution. *Bull. Math. Biol.*, 51(1):125–131, 1989.

[3] Christos A. Athanasiadis and Persi Diaconis. Functions of random walks on hyperplane arrangements. *Adv. in Appl. Math.*, 45(3):410–437, 2010.

[4] Arvind Ayyer, Steven Klee, and Anne Schilling. Combinatorial Markov chains on linear extensions. *J. Algebraic Combin.*, 39(4):853–881, 2014.

[5] Arvind Ayyer, Anne Schilling, Benjamin Steinberg, and Nicolas M. Thiéry. Directed nonabelian sandpile models on trees. *Comm. Math. Phys.*, 335(3):1065–1098, 2015.

[6] Arvind Ayyer, Anne Schilling, Benjamin Steinberg, and Nicolas M. Thiéry. Markov chains, $\mathscr{R}$-trivial monoids and representation theory. *Internat. J. Algebra Comput.*, 25(1-2):169–231, 2015.

[7] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and automata*, volume 129 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2010.

[8] Pat Bidigare, Phil Hanlon, and Dan Rockmore. A combinatorial description of the spectrum for the Tsetlin library and its generalization to hyperplane arrangements. *Duke Math. J.*, 99(1):135–174, 1999.

[9] Louis J. Billera, Kenneth S. Brown, and Persi Diaconis. Random walks and plane arrangements in three dimensions. *Amer. Math. Monthly*, 106(6):502–524, 1999.

[10] Anders Björner. Random walks, arrangements, cell complexes, greedoids, and self-organizing libraries. In *Building bridges*, volume 19 of *Bolyai Soc. Math. Stud.*, pages 165–203. Springer, Berlin, 2008.

[11] Anders Björner. Note: Random-to-front shuffles on trees. *Electron. Commun. Probab.*, 14:36–41, 2009.

[12] Leo Breiman. The individual ergodic theorem of information theory. *Ann. Math. Statist.*, 28:809–811, 1957.

[13] Graham Brightwell and Peter Winkler. Counting linear extensions. *Order*, 8(3):225–242, 1991.

[14] Kenneth S. Brown. Semigroups, rings, and Markov chains. *J. Theoret. Probab.*, 13(3):871–938, 2000.

[15] Kenneth S. Brown and Persi Diaconis. Random walks and hyperplane arrangements. *Ann. Probab.*, 26(4):1813–1854, 1998.

[16] Russ Bubley and Martin Dyer. Faster random generation of linear extensions. *Discrete Math.*, 201(1-3):81–88, 1999.

[17] M. L. Cetlin. Finite automata and the simulation of the simplest forms of behavior. *Uspehi Mat. Nauk*, 18(4 (112)):3–28, 1963.

[18] Fan Chung and Ron Graham. Edge flipping in graphs. *Adv. in Appl. Math.*, 48(1):37–63, 2012.

[19] Luc Devroye and Gábor Lugosi. *Combinatorial methods in density estimation*. Springer Series in Statistics. Springer-Verlag, New York, 2001.

[20] Persi Diaconis. From shuffling cards to walking around the building: an introduction to modern Markov chain theory. In *Proceedings of the International Congress of Mathematicians, Vol. I (Berlin, 1998)*, number Extra Vol. I, pages 187–204, 1998.

[21] Paul Edelman, Takayuki Hibi, and Richard P. Stanley. A recurrence for linear extensions. *Order*, 6(1):15–18, 1989.

[22] James Allen Fill. An exact formula for the move-to-front rule for self-organizing lists. *J. Theoret. Probab.*, 9(1):113–160, 1996.

[23] Peter C. Fishburn and William V. Gehrlein. A comparative analysis of methods for constructing weak orders from partial orders. *J. Mathematical Sociology*, 4(1):93–102, 1975.

[24] Robert M. Gray. *Entropy and information theory*. Springer, New York, second edition, 2011.

[25] Mark D. Haiman. Dual equivalence with applications, including a conjecture of Proctor. *Discrete Math.*, 99(1-3):79–113, 1992.

[26] W. J. Hendricks. The stationary distribution of an interesting Markov chain. *J. Appl. Probability*, 9:231–233, 1972.

[27] W. J. Hendricks. An extension of a theorem concerning an interesting Markov chain. *J. Appl. Probability*, 10:886–890, 1973.

[28] Alexander Karzanov and Leonid Khachiyan. On the conductance of order Markov chains. *Order*, 8(1):7–15, 1991.

[29] Donald E. Knuth. *The art of computer programming. Vol. 1*. Addison-Wesley, Reading, MA, 1997. Fundamental algorithms, Third edition [of MR0286317].

[30] Donald E. Knuth. *The art of computer programming. Vol. 3*. Addison-Wesley, Reading, MA, 1998. Sorting and searching, Second edition [of MR0445948].

[31] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. Second edition of [MR2466937], With contributions by Elizabeth L. Wilmer, With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson.

[32] Claudia Malvenuto and Christophe Reutenauer. Evacuation of labelled graphs. *Discrete Math.*, 132(1-3):137–143, 1994.

[33] Stuart Margolis, Franco Saliola, and Benjamin Steinberg. Combinatorial topology and the global dimension of algebras arising in combinatorics. *J. Eur. Math. Soc. (JEMS)*, 17(12):3037–3080, 2015.

[34] Jon McCammond, John Rhodes, and Benjamin Steinberg. Geometric semigroup theory. preprint, arXiv:1104.2301, 2011.

[35] Brockway McMillan. The basic theorems of information theory. *Ann. Math. Statistics*, 24:196–219, 1953.

[36] Evita Nestoridi. Optimal strong stationary times for random walks on the chambers of a hyperplane arrangement. *Probab. Theory Related Fields*, 174(3-4):929–943, 2019.

[37] John Pike. *Eigenfunctions for Random Walks on Hyperplane Arrangements*. ProQuest LLC, Ann Arbor, MI, 2013. Thesis (Ph.D.)–University of Southern California.

[38] John Rhodes and Anne Schilling. Normal distributions of finite Markov chains. *Internat. J. Algebra Comput.*, 29(8):1431–1449, 2019.

[39] John Rhodes and Anne Schilling. Unified theory for finite Markov chains. *Adv. Math.*, 347:739–779, 2019.

[40] John Rhodes and Anne Schilling. Mixing time for Markov chain on linear extensions. *Sém. Lothar. Combin.*, 85B:Art. 27, 12, 2021.

[41] John Rhodes, Anne Schilling, and Pedro V. Silva. Random walks on semaphore codes and delay de Bruijn semigroups. *Internat. J. Algebra Comput.*, 26(4):635–673, 2016.

[42] John Rhodes, Anne Schilling, and Pedro V. Silva. Holonomy theorem for finite semigroups. *Internat. J. Algebra Comput.*, 32(3):443–460, 2022.

[43] Olivier Rioul. This is IT: A primer on Shannon's entropy and information. *L'Information, Séminaire Poincaré*, XXIII:43–77, 2018.

[44] Franco Saliola. Eigenvectors for a random walk on a left-regular band. *Adv. in Appl. Math.*, 48(2):306–311, 2012.

[45] Maurice-Paul Schützenberger. Sur certains treillis gauches. *C. R. Acad. Sci. Paris*, 224:776–778, 1947.

[46] Claude Elwood Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, 1948.

[47] Richard P. Stanley. Promotion and evacuation. *Electron. J. Combin.*, 16(2, Special volume in honor of Anders Björner):Research Paper 9, 24, 2009.

[48] Howard Straubing. *Finite automata, formal logic, and circuit complexity.* Progress in Theoretical Computer Science. Birkhäuser Boston, Inc., Boston, MA, 1994.

[49] Yechezkel Zalcstein. Locally testable languages. *J. Comput. System Sci.*, 6:151–167, 1972.

(J. Rhodes) Department of Mathematics, University of California, Berkeley, CA 94720, U.S.A.
Email address: rhodes@math.berkeley.edu, blvdbastille@gmail.com

(A. Schilling) Department of Mathematics, UC Davis, One Shields Ave., Davis, CA 95616-8633, U.S.A.
Email address: anne@math.ucdavis.edu